

## Predicting the broadly neutralizing antibody susceptibility of the HIV reservoir

Wen-Han Yu, David Su, Julia Torabi, Christine M. Fennessey, Andrea Shiakolas, Rebecca Lynch, Tae-Wook Chun, Nicole Doria-Rose, Galit Alter, Michael S. Seaman, Brandon F. Keele, Douglas A. Lauffenburger, Boris Julg

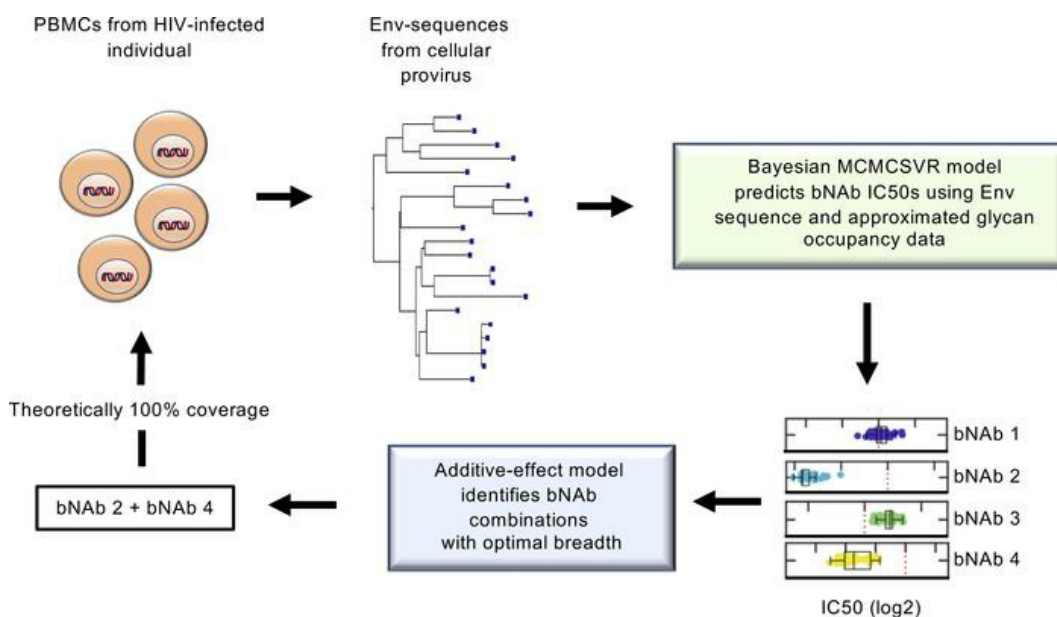
*JCI Insight*. 2019;4(17):e130153. <https://doi.org/10.1172/jci.insight.130153>.

Resource and Technical Advance

AIDS/HIV

Therapeutics

### Graphical abstract



Find the latest version:

<https://jci.me/130153/pdf>



# Predicting the broadly neutralizing antibody susceptibility of the HIV reservoir

Wen-Han Yu,<sup>1,2</sup> David Su,<sup>1</sup> Julia Torabi,<sup>1</sup> Christine M. Fennessey,<sup>3</sup> Andrea Shiakolas,<sup>4</sup> Rebecca Lynch,<sup>5</sup> Tae-Wook Chun,<sup>6</sup> Nicole Doria-Rose,<sup>4</sup> Galit Alter,<sup>1</sup> Michael S. Seaman,<sup>7</sup> Brandon F. Keele,<sup>3</sup> Douglas A. Lauffenburger,<sup>2</sup> and Boris Julg<sup>1</sup>

<sup>1</sup>Ragon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts, USA. <sup>2</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>3</sup>AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA. <sup>4</sup>Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland, USA. <sup>5</sup>Department of Microbiology, Immunology and Tropical Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, District of Columbia, USA. <sup>6</sup>Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, Maryland, USA. <sup>7</sup>Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA.

**Broadly neutralizing antibodies (bNAbs) against HIV-1 are under evaluation for both prevention and therapy. HIV-1 sequence diversity observed in most HIV-infected individuals and archived variations in critical bNAb epitopes present a major challenge for the clinical application of bNAbs, as preexistent resistant viral strains can emerge, resulting in bNAb failure to control HIV. In order to identify viral resistance in patients prior to antibody therapy and to guide the selection of effective bNAb combination regimens, we developed what we believe to be a novel Bayesian machine-learning model that uses HIV-1 envelope protein sequences and foremost approximated glycan occupancy information as variables to quantitatively predict the half-maximal inhibitory concentrations (IC<sub>50</sub>) of 126 neutralizing antibodies against a variety of cross clade viruses. We then applied this model to peripheral blood mononuclear cell-derived proviral Env sequences from 25 HIV-1-infected individuals mapping the landscape of neutralization resistance within each individual's reservoir and determined the predicted ideal bNAb combination to achieve 100% neutralization at IC<sub>50</sub> values <1 µg/ml. Furthermore, predicted cellular viral reservoir neutralization signatures of individuals before an analytical antiretroviral treatment interruption were consistent with the measured neutralization susceptibilities of the respective plasma rebound viruses, validating our model as a potentially novel tool to facilitate the advancement of bNAbs into the clinic.**

**Conflict of interest:** BJ has received research support from Gilead Sciences. GA has received research support from GlaxoSmithKline, Pfizer, Merck & Co., Janssen Pharmaceuticals, Sanofi, Gilead Sciences, and Bristol-Myers Squibb. DAL has received research support from Janssen Pharmaceuticals, Boehringer Ingelheim, and Merrimack Pharmaceuticals and is an equity shareholder of Torque Therapeutics. BJ, GA, DAL, and AS all hold patents or have filed patent applications that are outlined in the supplement.

**Copyright:** © 2019, American Society for Clinical Investigation.

**Reference information:** JCI Insight. 2019;4(17):e130153. <https://doi.org/10.1172/jci.insight.130153>.

**Submitted:** May 22, 2019

**Accepted:** July 26, 2019

**Published:** September 5, 2019.

## Introduction

HIV-1, the cause of the global AIDS epidemic, continues to be a major public health concern. While the development of antiretroviral therapy (ART) has revolutionized HIV care, it is now clear that life-long ART neither eliminates HIV, nor completely reconstitutes the immune system; in addition, it can have deleterious long-term side effects (1, 2). Thus, efforts to eradicate HIV infection, or at least induce states of viral control and drug-free remission, are therefore needed. The recent identification of highly potent and broadly neutralizing antibodies (bNAbs) against HIV-1 has led to a renaissance of antibodies as therapeutic and prophylactic tools. Initial preclinical and clinical studies with passive administration of bNAbs have shown great promise that these monoclonals could play a revolutionary role in HIV care (3–9). While these antibodies are able to temporarily reduce plasma viral loads in viremic HIV-1-infected individuals (3, 5, 10) or delay viral rebound once ART is stopped, recent studies showed that neutralization-resistant viral strains were often selected to emerge, ultimately resulting in virological failure of bNAb treatment (11–13). Single bNAb administration therefore caused strong selective pressure, without eliminating the HIV-1 latent reservoirs due to preexisting viral epitope variants with resistance to selected bNAbs. Identifying viral resistance in patients prior to antibody treatment is therefore critical to rationally design bNAb combination regimens

while taking advantage of the increasing number of available bNAbs, thereby avoiding the risk of treatment failure or development of novel resistance mutations.

So far, quantifying archived viral susceptibility to bNAbs prior to antibody treatment depends on inducing viral outgrowth from stimulated resting CD4<sup>+</sup> T cells and measuring neutralization sensitivity of these viruses in standard TZM-bl assays (12). This approach, however, is laborious, time consuming, and expensive and fails to capture the whole spectrum of minor viral variants that could emerge, as not every replication-competent provirus can and will be induced for outgrowth under in vitro conditions (14). Importantly, the median frequency of cells carrying intact proviruses is more than 60-fold higher than what is detected by the standard QVOA (15). Alternative approaches that more comprehensively characterize the landscape of neutralization-susceptible proviruses or even identify novel or unexpected mutations that could interfere with antibody recognition are therefore needed.

In the HIV treatment field, genotyping of plasma virus and cellular proviral DNA (GenoSure Archive) for the detection of antiretroviral drug resistance is routinely performed in clinical practice prior to ART initiation in order to choose optimal antiretroviral drug regimens. This viral sequence-based approach uses an extensive database of matched HIV phenotype and genotype samples to predict viral resistance to antiretroviral drugs (Monogram Biosciences, <http://www.monogrambio.com>). Mimicking this approach, we developed what we believe to be a novel computational model for the prediction of viral neutralization sensitivity based on HIV-1 envelope (Env) protein sequences. When combined with single-genome sequencing of proviral DNA from peripheral blood mononuclear cells (PBMCs), it allows the rapid and accurate mapping of the HIV-1 reservoir for preexisting resistance to current anti-HIV antibodies, with the capacity to be readily adapted to future bNAbs, therefore facilitating the selection of an effective bNAb regimen.

The HIV-1 Env trimer is coated by a glycan shield consisting of approximately 90 N-linked oligosaccharides, which represent half the mass of Env and play a crucial role in bNAb-epitope interactions by altering the antibody accessibility of the underlying protein (16, 17). High-resolution imaging approaches have revealed that bNAb and glycan interactions are common across all bNAb classes, and previous analyses to map antibody binding and antibody neutralization activity depended not only on sequence, but also on glycan occupancy (18, 19). Computational approaches for prediction of neutralization sensitivity based on Env sequences have been reported previously (20–22). Information of glycans and glycan occupancy, however, have not been included in these models, therefore excluding a critical component. Our previous work characterized site-specific N-linked glycan occupancy profiles by conducting a mass spectrometric N-linked glycoproteomic analysis from a set of 149 selected Env proteins, including monomer gp120s and trimers (19). A total of 2,891 potential N-linked glycosylation sites (PNGSs) were measured among those proteins, and distinct sequence usage signatures surrounding PNGSs were associated with different degrees of glycan occupancy (19), allowing us to estimate glycan occupancy based on protein sequence composition.

In this report, we describe the computational framework that uses both HIV Env protein sequence and glycan occupancy as variables to quantitatively predict the half-maximal inhibitory concentrations (IC<sub>50</sub>) for 126 neutralizing antibodies against a variety of cross clade viruses. We then applied this model to proviral Env sequences from 25 HIV-1-infected individuals, mapping the landscape of neutralization resistance within each individual's reservoir against bNAbs that are currently under clinical development. Using this information, we determined the predicted ideal bNAb combination to achieve 100% neutralization at IC<sub>50</sub> values <1 µg/ml for each individual. Finally, we applied the prediction model to map the viral reservoir of individuals before an analytical ART interruption (11) and correlated the reservoir signatures with the rebound plasma viruses neutralization susceptibilities.

## Results

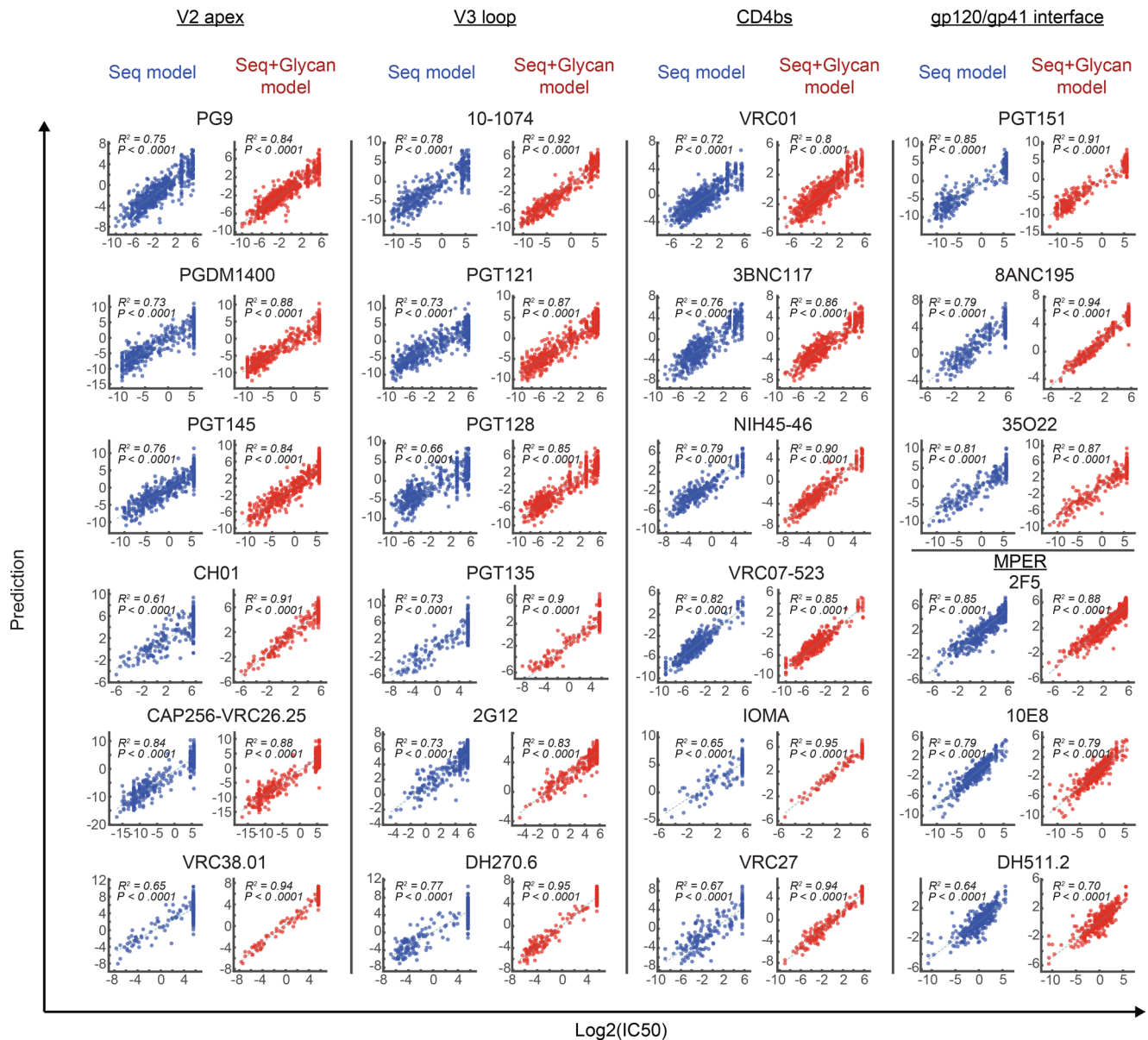
*Combining Env sequence and glycan occupancy to build the neutralization prediction models.* As the presence or absence of particular N-linked glycans on the HIV-1 Env protein have a significant effect on antibody-epitope interactions, construction of the computational framework for the predictive model was strategized as two steps: (a) estimating the glycan occupancy of PNGSs based on the neighboring sequences and its sequence-derived information and (b) predicting antibody neutralization activities combining Env sequence information and the estimated glycan occupancy. We initially collected full-length Env protein sequences from the Los Alamos HIV database and from GenBank, paired with all available mAb neutralization sensitivity values retrieved from the Los Alamos “Compile, Analyze and Tally NAb Panels” (CATNAP) database (23), a web platform that comprehensively assembles published neutralization information. To ensure sufficient statistical power,

only mAbs with available neutralization values against more than 70 viruses/Env sequences were considered for the model construction. In total, data for 126 mAbs were included in the models with 71–717 available paired Env sequence-neutralization values per mAb. These antibodies represented all 5 currently identified bNAb epitope classes, covering specificities for the V2 apex, V3 loop, CD4-binding site (CD4bs), gp120-gp41 interface, and the membrane proximal external region (MPER) (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.130153DS1>).

To deconvolute the corresponding glycan occupancy at each PNGS from the collected Env panels, we implemented a supervised machine-learning approach to build a predictive glycan regression model that defined the relationship between the glycan occupancy of a given PNGS and its neighboring sequence compositions. From the site-specific glycan occupancy profiles obtained from a set of 149 selected Env proteins, including monomer gp120s and trimers (19), a high-quality set of 766 PNGSs was selected, where glycan occupancy was widely distributed from 0% to 100 %. The sequence-derived features for individual PNGSs were generated from a sequence window between position –30 and 30, where the Asn of the PNGS was at position 0. The protein residue composition and the residue's hydrophobicity, flexibility, polarity, and  $\beta$ -turn values at each position were then extracted. The predictive glycan model was then constructed by a 2-step process: (a) feature reduction by Elastic-Net regularization and (b) model regression by a support vector regression (SVR) model. The prediction performance of glycan occupancy based on the sequence-derived features was assessed by repeated 10-fold cross-validation (10-fold CV). (The original data set is randomly partitioned into 10 sub-data sets. The 9 sub-data sets were used for model training and the remaining 1 was used for model validation. The CV process is repeated a number of times.) The predictive glycan model showed robust glycan occupancy prediction, with  $R^2 = 0.68$  ( $P < 0.001$ ) and area under the curve = 0.96 (Supplemental Figure 1), suggesting that glycan occupancy can be predicted by the sequence signatures.

With this rich protein sequence and predicted glycan occupancy data set, the computational framework Bayesian Markov Chain Monte Carlo–SVR (MCMC-SVR) model was used to build the neutralization prediction model (19) for each individual mAb. The models first defined the mAb-binding signatures by down selecting the sets of key protein residues and PNGSs associating with the mAb neutralization sensitivities, and then IC<sub>50</sub> values were estimated based on these determinant features. Overall model performance showed consistent robustness across all 126 tested mAbs from all 5 epitope-target classes with an exceptional average correlation ( $R^2 = 0.9$ ). To evaluate whether glycan information does indeed contribute to prediction accuracy, we compared a model version that solely used Env sequences with the model that used both integrated glycans and sequences. Both models were based on the exact same number of features to assure reasonable comparison. In general, the integrated model greatly improved the prediction resolution for most of mAbs, as determined by goodness of fit by  $R^2$  value (Figure 1). Particularly, the inclusion of glycan features into the model increased the neutralization prediction accuracy for certain mAbs by up to 30%, i.e., for PGDM1400, PGT121, PGT128, PGT135, CH01, VRC38.01, and IOMA. As expected, neutralization prediction for bNAbs against the V2 apex and the V3 loop benefited the most from inclusion of the glycan features, followed by CD4bs and gp120-gp41 interface mAbs. In contrast, the glycan features showed less effect on the neutralization prediction of mAbs targeting MPER (Figure 1). These observations cohesively matched the molecular understanding of mAb-epitope interaction. For instance, nearly all mAbs specific for the V2 apex and V3 loop directly target the glycans within their epitopes, whereas antibody-vulnerable glycan sites are unlikely to be recognized by the MPER mAbs, such as 2F5 and 10E8.

*Mapping the HIV reservoir for neutralization susceptibility.* There is significant diversity within the HIV-1 reservoir (24), and virologic failure of bNAb treatment due to the rapid selection of neutralization-resistant viral variants has been reported in animal models and early-phase clinical trials (3, 5–7, 10, 11). We performed bulk viral outgrowth cultures with PBMCs from 25 HIV-infected individuals, of which 23 had undetectable HIV RNA levels on ART (median 8 years on ART, range 1–18 years) and 2 had low level viremia in the absence of ART at the time of sampling (Table 1). All were US residents at the time of PBMC collection and were therefore likely to be infected with clade B viruses. The outgrown culture isolates were screened for sensitivity to PGT121, PGDM1400, 3BNC117, and VRC01 in a TZM-bl cell neutralization assay. Similar to published data (12), substantial heterogeneity in neutralization sensitivities of the outgrown viruses was observed, with some showing reduced sensitivity or resistance against multiple bNAbs (394747, 588552, 842272), while others were sensitive across all tested antibodies (190487, 237983, 387879, 486632, etc.) (Figure 2). We were therefore interested to map the landscape of neutralization sensitivity of archived viruses in the HIV-1 reservoir by applying the neutralization prediction model to proviral Env sequences derived from PBMCs of these



**Figure 1. Prediction model performance for selected mAb neutralization sensitivities ( $\log_2 \text{IC}_{50}$ ) representing relevant envelope broadly neutralizing antibody target sites (V2 apex, V3 loop, CD4-binding site, gp120/gp41 interface, and MPER).** Scatter plots illustrate the correlation between measured  $\log_2(\text{IC}_{50})$  using neutralization data from the Los Alamos “Compile, Analyze and Tally NAb Panels” (CATNAP) database and  $\log_2(\text{IC}_{50})$  values predicted by our Bayesian machine-learning model. For each antibody,  $\log_2(\text{IC}_{50})$  prediction based on HIV-1 envelope (Env) sequence alone or using sequence and approximated glycan occupancy information is shown. Overall, 71–717 available paired env sequence–neutralization values per mAb were available. The Spearman’s rho and its 2-sided  $P$  value are labeled.

individuals. In total, 727 single-genome full-length Env sequences were obtained from DNA isolated from primary PBMCs with a mean of 29 sequences per individual (range 8–83 sequences).

To calculate neutralization sensitivities based on the sequences, we focused on 10 bNAbs that are currently in clinical development, specifically the CD4bs antibodies VRC01, 3BNC117, N6, and VRC07-523; the V3 loop glycan antibodies PGT121 and 10-1074; the V2 apex glycan antibodies PGDM1400, PG9, and CAP256-VRC26.25; and the MPER antibody 10E8. As expected, we observed substantial heterogeneity of the predicted neutralization sensitivities, as determined by  $\text{IC}_{50}$  values (in  $\mu\text{g}/\text{ml}$ ), across bNAbs but also within sequences to each single bNAb (Figure 3). Results were reflective of the potency and breadth of the bNAb, e.g., we predicted that the CD4bs antibodies 3BNC117, N6, and VRC07-523 would neutralize 92%, 95%, and 95% of all Env sequences across all patients at an  $\text{IC}_{50} < 1 \mu\text{g}/\text{ml}$ , respectively, while VRC01 neutralized 86% of all Env sequences at  $\text{IC}_{50} < 2 \mu\text{g}/\text{ml}$ . The V3 mAbs 10-1074 and PGT121 were predicted

Table 1. Patient characteristics

ID	Age (yr at time of sample)	Sex	HIV diagnosis	Years on ART	HIV RNA (at time of sample)	ART
186089	46	Male	2006	N/A	165	No
190487	66	Female	1992	14	<20	Yes
237983	63	Female	1996	N/A	124	No
239667	48	Male	2004	3	<20	Yes
296260	48	Male	2001	13	<20	Yes
387879	33	Male	2006	7	<20	Yes
394747	43	Male	1986	1	<20	Yes
430062	55	Male	2000	10	<20	Yes
438531	55	Male	1991	5	<20	Yes
467037	59	Male	1991	12	<20	Yes
486632	52	Male	1995	11	<20	Yes
521183	59	Male	1987	6	<20	Yes
541683	51	Male	1990	7	<20	Yes
588552	56	Male	2001	6	<20	Yes
622054	47	Female	1994	10	<20	Yes
690419	56	Male	2003	11	<20	Yes
765757	36	Male	2000	10	<48	Yes
778292	36	Male	2008	3	<20	Yes
821034	57	Female	1995	18	<20	Yes
839338	56	Male	2001	15	<20	Yes
842272	38	Male	2011	4	<20	Yes
913820	53	Male	2002	3	<75	Yes
930213	49	Male	1993	10	<20	Yes
948970	33	Male	2011	4	<20	Yes
975240	32	Male	2008	5	<20	Yes

ART, antiretroviral therapy. HIV RNA in copies/mL.

to neutralize 63% and 70% of sequences at IC<sub>50</sub> values <1 µg/ml and 40% and 55% of sequences at IC<sub>50</sub> values <0.1 µg/ml consistent with their increased potency. The V2 apex antibodies required IC<sub>50</sub> values of ≥ 1.4 and ≥ 11.6 µg/ml (PGDM1400 and CAP256-VRC26.25) to neutralize at least 50% of the viruses, most likely reflective of their limited activity against clade B. The MPER antibody 10E8 achieved 91% Env coverage at IC<sub>50</sub> <1 µg/ml and covered 43% of sequences at IC<sub>50</sub> values of <0.1 µg/ml.

Among antibodies of the same specificity, predicted neutralization susceptibility of sequences was consistent, e.g., for the V3 glycan-dependent antibodies, predicted IC<sub>50</sub> values for PGT121 and 10-1074 correlated strongly (spearman  $r = 0.81$ ,  $P < 0.0001$ ) (Supplemental Figure 2A). For the other specificities, cross class correlation was less robust (Supplemental Figure 2B). Specifically for the CD4bs antibodies, overall correlation of predicted IC<sub>50</sub> values was more heterogeneous, e.g., VRC01, 3BNC117, and VRC07-523 correlated weakly (spearman  $r$  ranging from 0.16 to 0.43), and there was a clear disconnect between N6 and VRC07-523 susceptibilities, for which predicted IC<sub>50</sub> values did not correlate ( $R = 0.03$ ,  $P = 0.34$ ) (Supplemental Figure 2, C and D), potentially related to the specific mode of recognition of N6, permitting it to avoid steric clashes with glycans in the V5 region of Env (25). These data are consistent with the differing epitope-binding pattern of the CD4bs antibodies (26), corresponding in neutralization susceptibility signatures that are bNAb specific.

*Reservoir diversity of neutralization sensitivity.* Within each individual's reservoir, we observed a significant proviral sequence diversity (mean diversity of all 25 patients was 2.0%; range 0.1%–3.9%) (Supplemental Figure 3). As a result, predicted IC<sub>50</sub> values for any 1 tested bNAb were quite diverse, ranging from sensitive to resistant (Figure 3 and Figure 4). For the CD4bs antibodies N6 and VRC07-523, respectively, sequences with predicted IC<sub>50</sub> values <1 µg/ml were found in all patients, but at the same time, 28% and 20% of patients also harbored sequences with predicted IC<sub>50</sub> values >1 µg/ml. For 3BNC117 and VRC01, respectively, 4% and 16% of patients only had sequences with predicted IC<sub>50</sub> values >1 µg/ml, while in 24% and 72% of patients, sequences with predicted IC<sub>50</sub> values <1 µg/ml and >1 µg/ml were found. For

ID	VRC01	3BNC117	PGT121	PGDM1400
975240	14.929	1.767	0.157	1.469
948970	4.69	0.35	0.12	>25
930213	>25	1.95	0.04	1.02
913820	10.52	1.11	4.27	>25
842272	15.42	3.23	>25	>25
839338	6.36	0.55	2.32	>25
821034	12.446	3.121	0.334	>25
778292	10.34	0.78	0.21	>25
765757	>25	1.12	0.04	>25
690419	13.96	1.03	5.57	>25
622054	13.30	2.72	>25	2.72
588552	10.58	1.41	>25	>25
541683	0.189	0.018	0.022	0.069
521183	16.20	0.70	>25	4.66
486632	1.29	0.54	0.02	4.60
467037	6.54	1.22	0.02	>25
438531	6.22	0.28	0.07	>25
430062	>25	8.50	0.05	>25
394747	>25	17.67	0.16	>25
387879	7.56	0.19	0.69	3.33
296260	6.56	4.93	0.13	>25
239667	3.16	1.27	>25	>25
237983	1.08	0.11	0.01	0.44
190487	7.18	0.91	0.11	0.38
186089	2.14	0.10	>25	>25

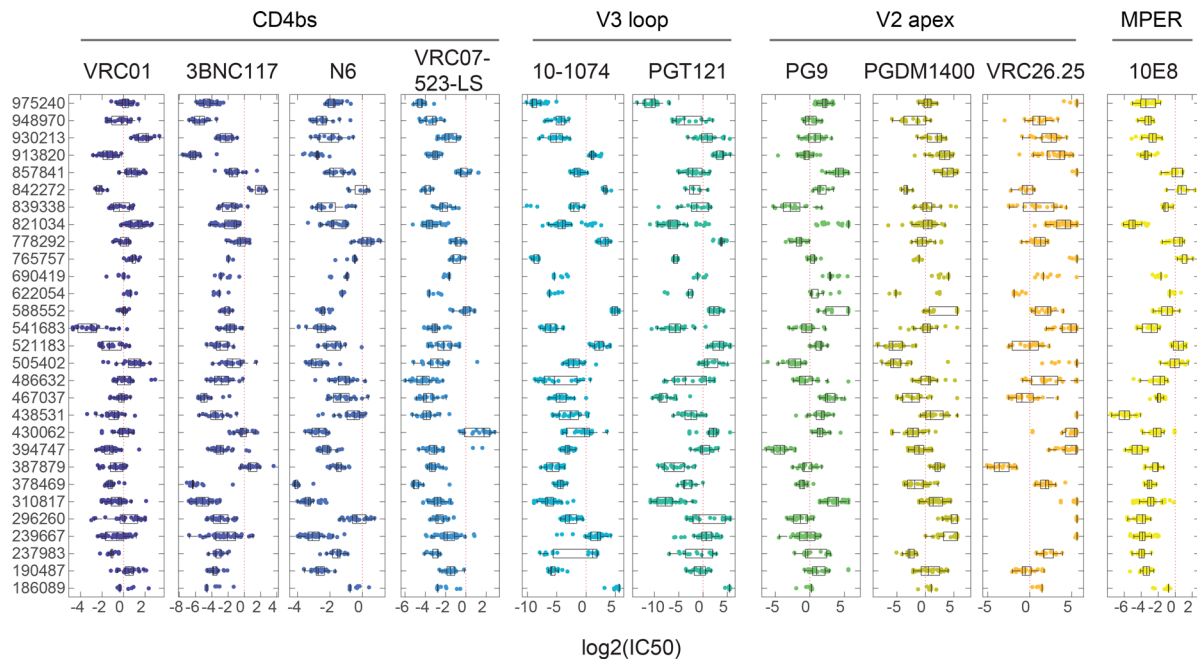
Color shading represents potency (IC<sub>50</sub> in µg/mL) as follows:

<0.1
0.1-1
1-10.0
10.0-25.0
>25.0

**Figure 2. IC<sub>50</sub> values (µg/mL) for outgrown viruses, as determined by TZM-bl neutralization assay.** PBMCs from 25 HIV-infected individuals were stimulated and infectious culture supernatants were tested for neutralization sensitivity against the broadly neutralizing antibodies (bNAb) VRC01, 3BNC117, PGT121, and PGDM1400 in a TZM-bl neutralization assay using increasing amounts of antibody. IC<sub>50</sub> values ≤0.1 µg/mL are highlighted in red, values between 0.1 and 1 µg/mL are highlighted in orange, values between 1 and 10 µg/mL are highlighted in yellow, values between 10 and 25 µg/mL are highlighted in green, and values >25 µg/mL are highlighted gray.

the V3 glycan antibodies PGT121 and 10-1074, respectively, 20% and 24% of patients had only sequences with predicted IC<sub>50</sub> values >1 µg/mL, while 60% and 28% of patients had sequences with predicted IC<sub>50</sub> values <1 µg/mL and >1 µg/mL. For the V2 apex antibodies, 20% of patients for PGDM1400 and 56% of patients for CAP256-VRC2625 had only sequences with predicted IC<sub>50</sub> values >1 µg/mL while 68% and 36% of patients were found to have sequences with IC<sub>50</sub> <1 µg/mL and >1 µg/mL. For the MPER antibody 10E8, 4% of patients carried only sequences with IC<sub>50</sub> values >1 µg/mL while sequences with both IC<sub>50</sub> <1 µg/mL and >1 µg/mL were found in 20% of patients. These data suggest that only a minority of the viral reservoir is fully sensitive to a given bNAb and that coexistence of sensitive and resistant viruses is frequently found; this is consistent with prior data demonstrating significant sequence variability in key contact site of bNAbs in plasma viruses with resistant and sensitive viruses existing in parallel (6).

We were next interested to determine if the neutralization susceptibility pattern would be consistent overtime. In 3 individuals, we determined predicted IC<sub>50</sub> values for an additional time point (i.e., T<sub>1</sub> or T<sub>2</sub>) (3 months to 1 year after the first time point) and for 2 individuals we analyzed a total of 3 time points that were up to 8 years apart (Figure 5). For patient 239667 the proviral diversity in Env was 3.5% (T<sub>1</sub>) and 3.3% (T<sub>2</sub>); for patient 387879 it was 1.8% (T<sub>1</sub>) and 2.0% (T<sub>2</sub>); for patient 913820 it was 1.3% (T<sub>1</sub>) and 2.1% (T<sub>2</sub>); for patient 821034 it was 2.9% (T<sub>1</sub>), 2.7% (T<sub>2</sub>), and 1.8% (T<sub>3</sub>); and for patient 975240 it was 1.1% (T<sub>1</sub>), 1.6% (T<sub>2</sub>), and 0.5% (T<sub>3</sub>). Overall, we observed a heterogeneous spread of predicted IC<sub>50</sub> values; however, the range of IC<sub>50</sub> values between time points did not differ significantly in the majority of cases. Only for some bNAb susceptibilities, i.e., for 3BNC117 in patient 239667 (unpaired *t* test with Welch correction, *P* < 0.05) or for VRC01 and PGT121 in patient 975240 (1-way ANOVA with Bonferroni correction, *P* value ranging from < 0.0001 to < 0.05, respectively), there were significant differences

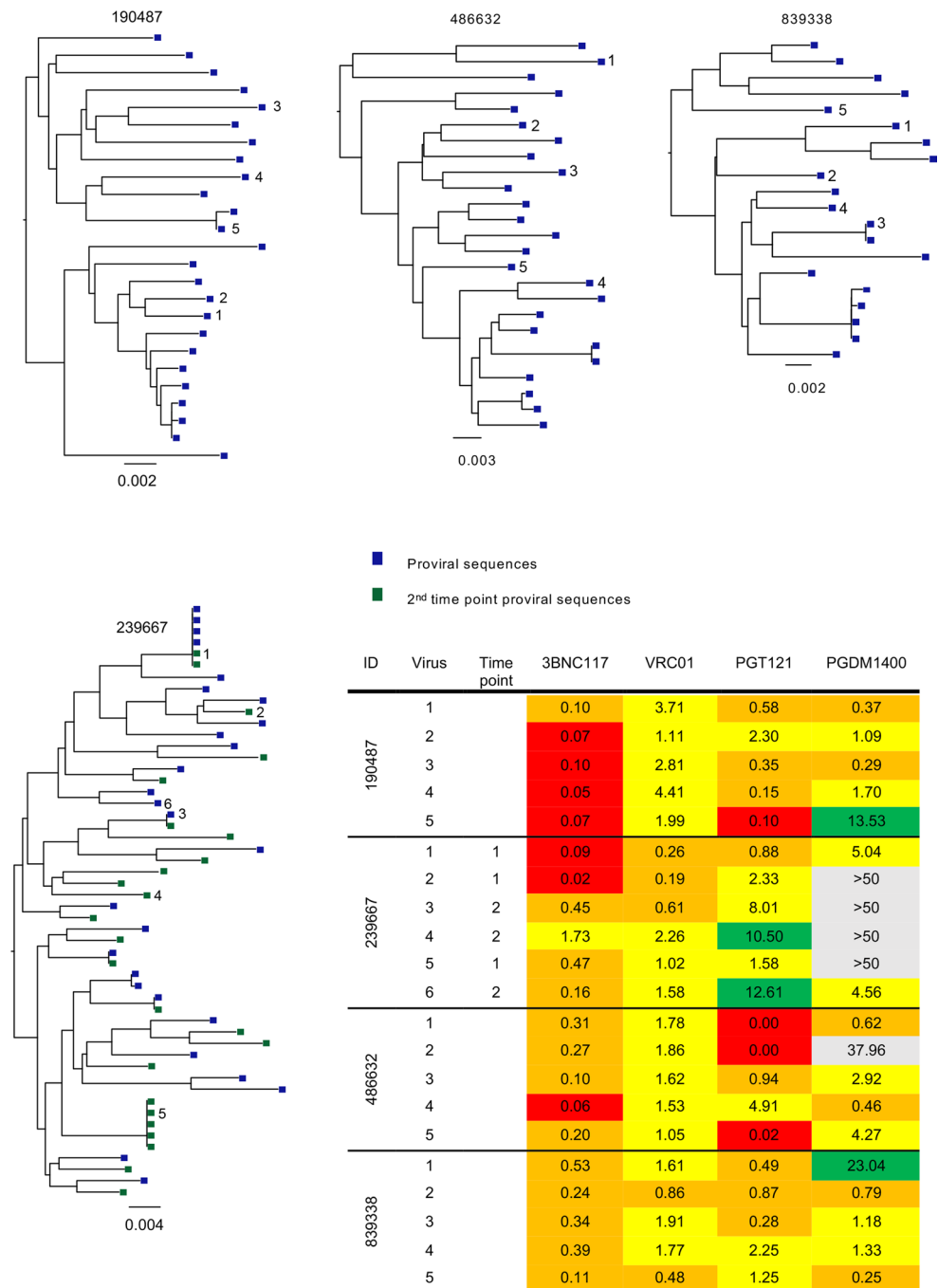


**Figure 3. Predicted  $\log_2(\text{IC}_{50})$  values for proviral HIV-1 envelope sequences from all 25 HIV-infected study participants.** In total, 727 single-genome full-length envelope (Env) sequences were obtained from DNA isolated from primary PBMCs with a mean of 29 sequences per individual (range 8–83). Shown are box-and-whisker plots demonstrating the breadth of neutralization sensitivities against multiple class broadly neutralizing antibodies (bNAb) and for each participant's proviral sequences. Each dot represents the predicted  $\text{IC}_{50}$  values for an Env sequence.

of the  $\text{IC}_{50}$  ranges between time points (Figure 5). The predicted neutralization sensitivities toward other bNAbs, however, did not seem affected. These data suggest that Env sequences and therefore respective sensitivities of reservoir viruses to neutralizing antibodies are relatively stable if plasma virus is continuously suppressed with ART as previously reported (27).

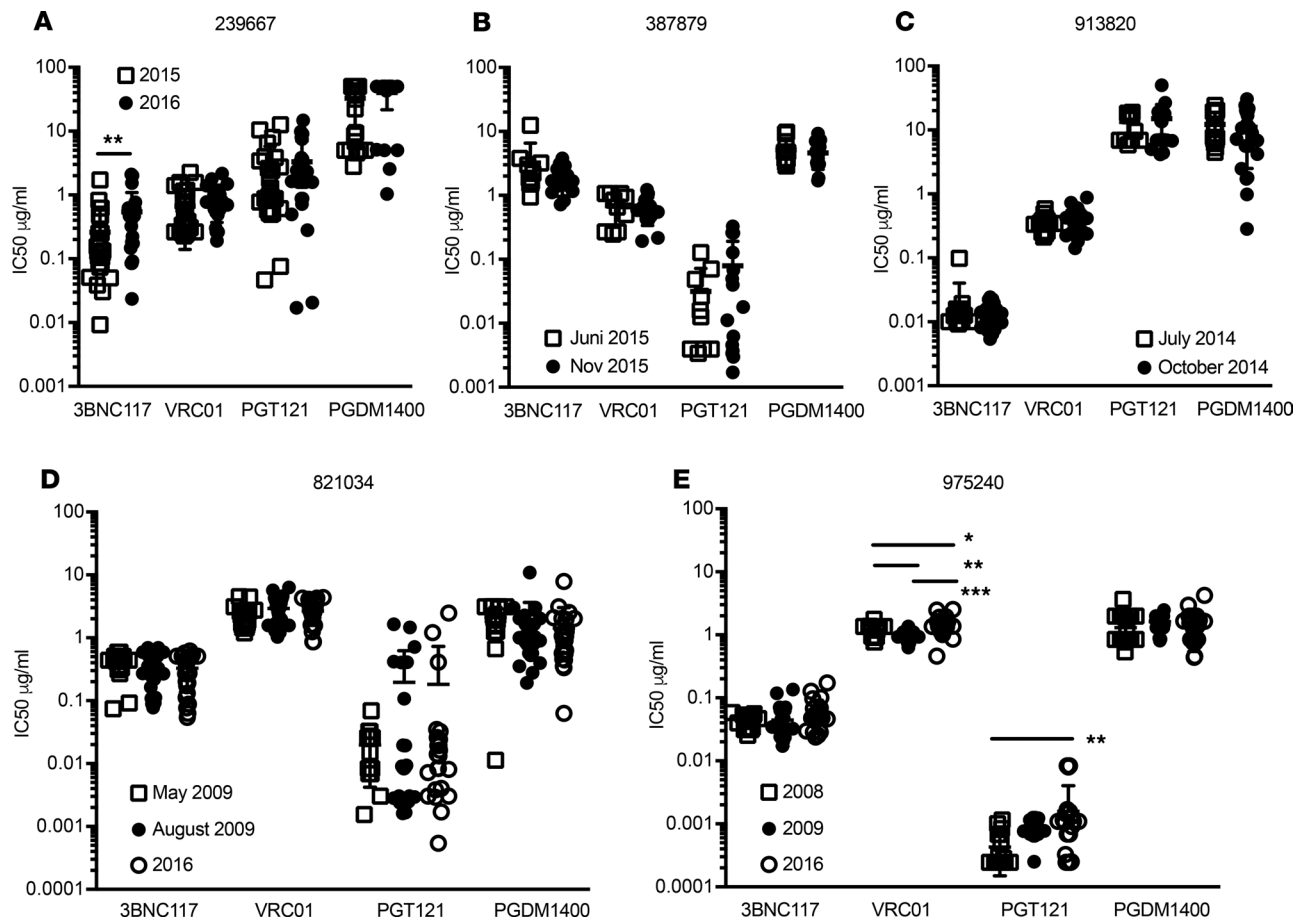
**Optimizing and validating bNAb combinations for therapy.** It has become increasingly clear that antibody combinations will be needed if bNAbs are considered for therapy or prevention to overcome preexistent resistance but also to prevent the development of novel escape mutations due to insufficient pressure on the virus. Our reservoir susceptibility data clearly demonstrate the vast heterogeneity in Env neutralization sensitivities, and we therefore determined which bNAb combination would have the highest predicted antiviral activity against all viral variants found in a given patient's reservoir. Focusing on our selected 10 bNAbs, we tested all double and triple combinations of bNAbs targeting different bNAb classes to maximally cover distinct epitopes. To estimate combined  $\text{IC}_{50}$ s, the individual bNAb  $\text{IC}_{50}$  values were first predicted. Because of the selected bNAbs that recognize distinct epitopes, the combination effect of bNAbs can be assumed to be additive (28, 29), and an additive-effect model was used to calculate the combined  $\text{IC}_{50}$  based on the following formula:  $\alpha = 1/(1/x + 1/y)$ , where  $\alpha$  denoted combined  $\text{IC}_{50}$  and  $x$  and  $y$  the  $\text{IC}_{50}$ s of single bNAbs. We again used the arbitrary cutoff ( $\text{IC}_{50} = 1 \mu\text{g}/\text{ml}$ ) to calculate total coverage of combined bNAbs. Overall, in each of the 25 study participants, a combination of 2 bNAbs was sufficient to achieve 100% neutralization coverage of all Envs that had been obtained from the given individual. Furthermore, in most cases multiple bNAb combinations achieved this goal, although with differing neutralization potencies; but for certain combinations, no or very limited viral coverage was predicted (Table 2). When focusing on the top 2 bNAb combination for each individual, based on the lowest  $\text{IC}_{50}$  concentration,  $\text{IC}_{50}$  values ranged from 0.0006 to 0.205  $\mu\text{g}/\text{ml}$ , with a mean of 0.034  $\mu\text{g}/\text{ml}$  (Table 2). The combination of CD4bs and V3 glycan antibody was the most frequently selected combination (36% of all combinations), followed by CD4bs and 10E8 (24%). While the addition of a third bNAb did result in a calculated increase of approximately 20% in potency, the nominal effect on  $\text{IC}_{50}$  values was rather moderate (mean  $\text{IC}_{50}$  reduction from 0.039 to 0.030  $\mu\text{g}/\text{ml}$ ).

Building on these in vitro data, we next analyzed a cohort of 9 HIV-infected individuals that had participated in the NIH 15-I-0140 trial (NCT02471326). In this study, participants were given multiple infusions of VRC01 before and during an analytical antiretroviral treatment interruption (11). During the study, all participants experienced viral rebound despite VRC01 plasma levels above 50  $\mu\text{g}/\text{ml}$ . Analysis of



**Figure 4. Maximum likelihood phylogenetic trees of proviral HIV-1 envelope sequences from 4 randomly selected study participants.** For participant 239667, proviral sequence from 2 time points (blue and green boxes) are shown. For some sequences (identified by numbers), predicted IC<sub>50</sub> values (in µg/mL) for the bNAbs 3BNC117, VRC01, PGT121, and PGDM1400 are shown in the chart to demonstrate the heterogeneity in neutralization susceptibility across sequences and broadly neutralizing antibodies (bNAbs). IC<sub>50</sub> values ≤0.1 µg/mL are highlighted in red, values between 0.1 and 1 µg/mL are highlighted in orange, values between 1 and 10 µg/mL are highlighted in yellow, values between 10 and 50 µg/mL are highlighted in green, and values >50 µg/mL are highlighted white. Genetic distance scale bars are shown for each tree.

neutralization sensitivity of autologous viruses before and after VRC01 infusion confirmed the emergence of HIV isolates that were less sensitive or even highly resistant to VRC01, suggesting rapid selection of archived resistant viral variants. We generated 199 single-genome Env sequences from isolated CD4<sup>+</sup> T cells obtained prior to the VRC01 infusions. Again, Env sequences were distinct and heterogeneous. The



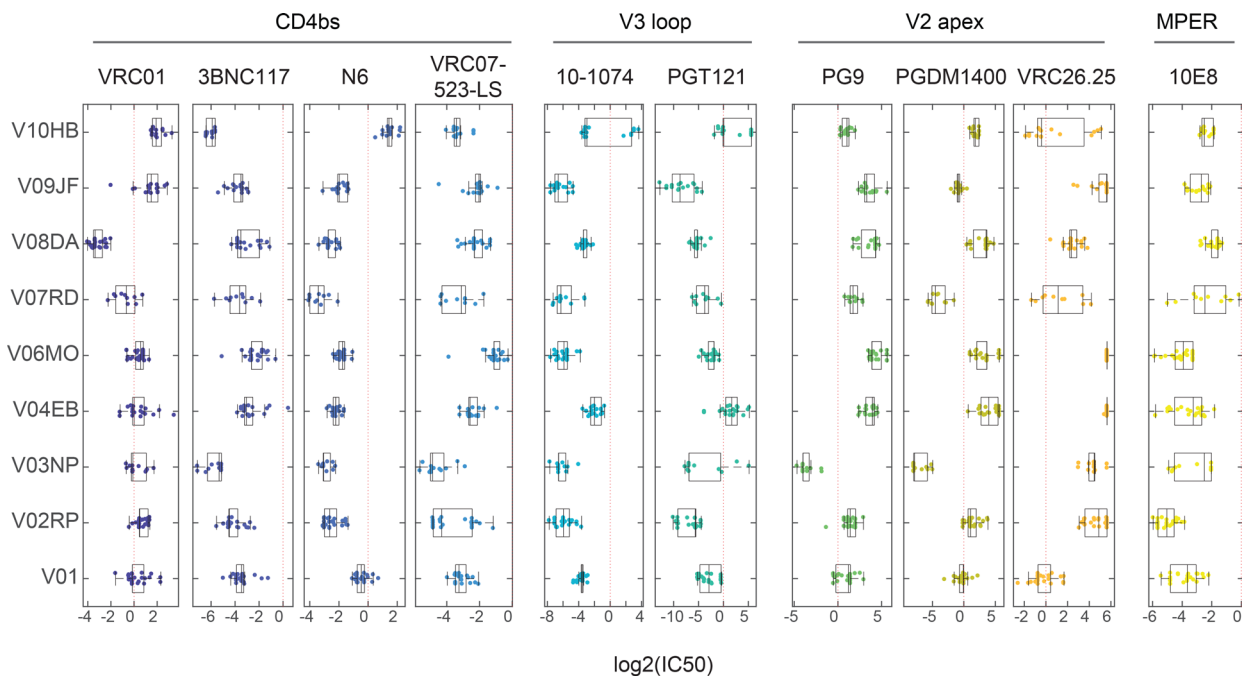
**Figure 5. Predicted IC<sub>50</sub> values (µg/mL) based on proviral HIV-1 envelope sequences from multiple time points per individual.** For 3 of the 25 study participants, proviral sequences from a second time point (3 months to 1 year after the first time point) were available and for 2 study participants 3 time points up to 8 years apart were available. Shown are predicted IC<sub>50</sub> values for the broadly neutralizing antibodies (bNAb) 3BNC117, VRC01, PGT121, and PGDM1400. Each dot represents the predicted IC<sub>50</sub> values for an envelope (Env) sequence. *P* values were calculated by 2-sided Wilcoxon rank-sum test and multiple testing corrected by the Benjamini and Hochberg method. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001.

average proviral diversity of Env for these individuals was 2.2% (range 0.8%–3.9%). In all except one participant (N08), sequences with predicted IC<sub>50</sub> values >1 µg/ml were detected with a mean IC<sub>50</sub> of 1.47 µg/ml (range from 0.06 to 11 µg/ml) across all available sequences, confirming preexistence of sequences with reduced neutralization sensitivity to VRC01 (Figure 6). Interestingly, predicted sensitivities to the other tested CD4bs antibodies were substantially higher, with mean IC<sub>50</sub>s of 0.12, 0.21, and 0.53 µg/ml for 3BNC117, VRC07-523, and N6, respectively. We then focused on the rebound viruses of 6 individuals for which sequence data was available and measured neutralization sensitivity of pseudoviruses generated from selected viral isolates to the VRC01 in a TZM-bl assay (Table 3). Overall, the predicted IC<sub>50</sub> values based on the rebound virus Env sequences against 4 bNAbs, including VRC01, 3BNC117, PGT121, and 10-1074, correlated well with the measured IC<sub>50</sub> values of the correspondent pseudoviruses, as determined in a TZM-bl assay (Figure 7), confirming the validity of our prediction model. When comparing the resistance pattern of the rebound viruses with the predicted sensitivity signatures of the reservoir viruses, we found that in 4 of the 6 study participants (N02, N03, N04, and N08), the rebound viruses had neutralization characteristics that were consistent and within the predicted sensitivity ranges determined by our proviral sequence prediction approach (Table 3). In the remaining 2 individuals (N06 and N09), rebound viruses contained highly resistant strains that had not been detected within the available proviral sequences. The latter might be due to the limited number of single virus sequences that we obtained for this study and deep-sequencing strategies to identify low-frequency variants will most likely be necessary in the future to generate a complete profile of the cellular reservoir's bNAb susceptibility. Alternatively, rebound viruses in these 2 individuals might be the result of recombination of latent viruses, as previously described (30).

**Table 2. Examples of broadly neutralizing antibody combinations and their respective predicted neutralization coverage for each patient's proviral HIV-1 envelope sequences**

ID	bNAb combination	Mean IC50 (µg/ml)	Sequence coverage (%)
186089	3BNC117:10E8 PGT121:VRC26.25	0.0437	100.00 0.00
190487	3BNC117:10-1074 VRC01:PGDM1400	0.0160	100.00 60.00
237983	10-1074:10E8 10-1074:VRC26.25	0.0319	100.00 47.37
239667	N6:10E8 10-1074:VRC26.25	0.0432	100.00 8.00
296260	3BNC117:10E8 VRC01:VRC26.25	0.0391	100.00 44.44
387879	PGT121:VRC26.25 3BNC117:PGDM1400	0.0105	100.00 30.00
394747	PG9:10E8 PGT121:VRC26.25	0.0228	100.00 60.00
430062	N6:PGDM1400 PGT121:VRC26.25	0.0727	100.00 19.23
438531	VRC07-523:10E8 N6:VRC26.25	0.0125	100.00 87.50
467037	3BNC117:PGT121 VRC01:PG9	0.0030	100.00 95.83
486632	VRC07-523:10-1074 VRC01:VRC26.25	0.0158	100.00 68.00
521183	3BNC117:PGDM1400 PGT121:PG9	0.0197	100.00 10.00
541683	VRC01:10-1074 PGT121:PGDM1400	0.0123	100.00 96.00
588552	N6:10E8 10-1074:PG9	0.1350	100.00 0.00
622054	10-1074:PGDM1400 VRC01:PG9	0.0111	100.00 86.36
690419	3BNC117:10-1074 PGT121:PG9	0.0255	100.00 87.50
765757	3BNC117:10-1074 VRC26.25:10E8	0.0026	100.00 0.00
778292	VRC07-523:PG9 PGT121:VRC26.25	0.2025	100.00 8.33
821034	PGT121:10E8 VRC01:VRC26.25	0.0069	100.00 5.88
839338	10-1074:PG9 PGT121:VRC26.25	0.0381	100.00 76.47
842272	VRC07-523:PGDM1400 3BNC117:10-1074	0.0396	100.00 0.00
913820	3BNC117:10E8 PGT121:PGDM1400	0.0111	100.00 5.41
930213	10-1074:10E8 VRC01:VRC26.25	0.0248	100.00 16.00
948970	3BNC117:10-1074 VRC01:VRC26.25	0.0135	100.00 89.47
975240	VRC07-523:PGT121 VRC01:VRC26.25	0.0006	100.00 34.12

Shown are the broadly neutralizing antibody (bNAb) combinations with the highest and the lowest sequence coverage. In cases where multiple combinations achieve 100% coverage, the one with the lowest IC50 value (µg/mL) is displayed.



**Figure 6. Predicted  $\log_2(\text{IC}_{50})$  values for proviral HIV-1 envelope sequences from 9 participants of the NIH 15-I-0140 trial.** A total of 199 single-genome envelope (Env) sequences were obtained from isolated CD4<sup>+</sup> T cells from 9 participants of the NIH 15-I-0140 trial prior to VRC01 administration and prior to analytical antiretroviral treatment interruption (ATI). Shown are box-and-whisker plots demonstrating the breadth of predicted neutralization sensitivities against multiple class broadly neutralizing antibodies (bNAbs) and for each participant's proviral sequences. Each dot represents the predicted  $\text{IC}_{50}$  values for an Env sequence.

## Discussion

In this report, we describe an analytical method, mirroring the genotyping approach used for antiretroviral drug resistance testing, to map viral susceptibility to binding and neutralization by bNAbs and guide the selection of efficacious bNAb combinations in a personalized way. Specifically, this approach combines HIV-1 single-genome Env sequencing from plasma virus or proviral DNA from the cellular reservoir with a potentially new computational model that also incorporates glycan occupancy information for the prediction of bNAb binding and viral neutralization sensitivity to rapidly and accurately diagnose preexisting resistance to anti-HIV-neutralizing antibodies, therefore facilitating the selection of effective bNAb regimens.

Computational modeling of antibody neutralization activity has been mainly used for vaccine design to develop immunogens for elicitation of bNAbs, and these models have primarily focused on Env protein genetic information. Bui et al. presented a trained neural network approach to directly predict  $\text{IC}_{50}$  values of given antibodies against HIV-1, based on the Env sequence information (31), and Hake et al. built binary classifiers with nonlinear support vector machines and string kernels to distinguish between HIV-1 resistance and susceptibility to a bNAb based on different Env amino acid (AA) sequences (20). More recently, Magaret et al. used 2 machine-learning approaches, based on a set of predefined AA sequence features to predict several TZM-bl neutralization assay outcomes for the CD4bs antibody VRC01, including virus's resistance versus sensitivity status,  $\log\text{IC}_{50}/80$  values, and estimated neutralization slope of the dose-response curves (22). Bricault et al. used random forest for  $\text{IC}_{50}$  regression predictions that included information on AA, PNGS, clade, and variable loop characteristics, demonstrating that the model accuracy using all the information was superior compared with model accuracy including single parameters, i.e., contact-region-only signatures (21). All these approaches primarily rely on sequence information, and although the latter model included PNGS as a variable, information on glycan occupancy at N-linked glycan sites that can positively or negatively influence antibody binding was not included. Env glycans, however, influence antibody access to the underlying protein (16, 32–34), and addition or removal of a particular glycan at one site may have a substantial effect on epitope availability, antibody binding, and consequently neutralization. Indeed, adding glycan information to our Bayesian MCMC approach by incorporating sequence-based glycan occupancy approximations substantially enhanced the prediction accuracy of the model ( $R^2 = 0.9$ ), allowing rigorous prediction of antibody binding and neutralization potency across multiple anti-

**Table 3. Measured VRC01 neutralization susceptibility of HIV-1 pseudoviruses and the range of predicted IC50 values**

Pseudovirus from rebound plasma viruses	VRC01 IC50 (μg/ml)	Range of predicted proviral IC50 (μg/ml)
N02RP - 1A5.8	1.89	0.7–2.7
N02RP - 2D4.2	1.575	
N02RP - 2D10.6	2.24	
N02RP - 2G1.18	2.68	
N03NP - 1A11.12	2.22	0.6–3.1
N03NP - 1F10.23	1.79	
N03NP - 1F12.6	2.58	
N03NP - 2D5.10	2.57	
N04 EB - 1D1.1	1.96	0.04–11
N04EB - 1F8.7	3.56	
N04EB - 1G4.5	4.03	
N04EB - 2A8.4	3.00	
N04EB - 2F8.19	10.5	0.06–2.5
N06MO - 1D3.18	25.7	
N06MO - 1F2.7	0.778	
N06 MO - 2D8.7	0.131	
N08DA - 1E4.1	0.192	0.006–0.24
N08DA -1H8.6	0.203	
N08DA - 2C7.1	0.239	
N08DA - 2D12.A3	0.154	
N08DA 2H3.A9	0.293	0.02–7.3
N09JF - 1A1.2	50.000	
N09JF - 1D4.2	50.000	
N09JF - 1E10.3	50.000	

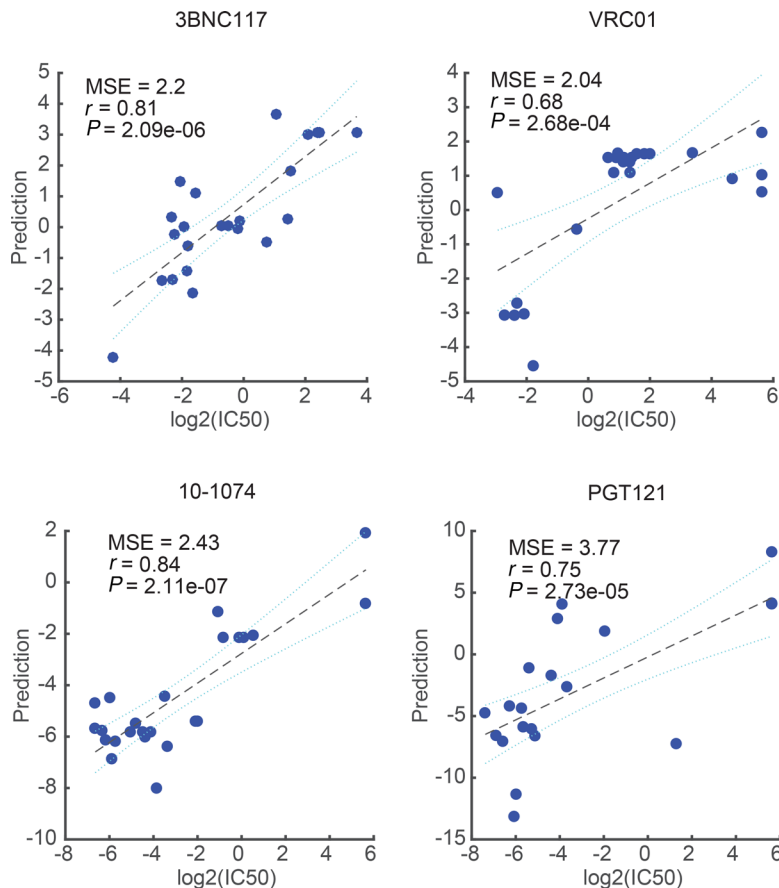
IC50 values for VRC01 neutralization of pseudoviruses based on HIV-1 envelope sequences from plasma rebound viruses during analytical antiretroviral treatment interruption from 6 participants in the NIH 15-I-0140 trial, as determined in a TZM-bl neutralization assay, and the range of predicted IC50 values based on proviral envelope sequences from CD4<sup>+</sup> T cells prior to antiretroviral treatment interruption.

body epitope specificities, demonstrating the robustness of the model. Moreover, using the predicted neutralization data, we were able to identify bNAb combinations that would theoretically cover the viral sequence diversity within an individual's reservoir using similar models, as previously reported (28, 29). Specifically, Wagh et al. had recently reported a mathematical modeling approach, the Bliss-Hill model, to accurately predict neutralization titers of bNAb combinations against HIV-1 clade C viruses (28). As, this model requires both IC50 and IC80 values, our bNAb combination analysis adapted the additive model as described by Kong et al., which had performed well in predicting potency of bNAb combinations using experimental data from single bNAbs (29). It should be emphasized that the optimal bNAb combinations predicted with our model are highly person specific, based purely on the available Env sequences obtained from the individual's reservoir, and that these combinations would not guarantee effectiveness against other HIV-1 isolates and clades, i.e., in the setting of a superinfection.

There are several potential limitations to this approach. Only a small fraction (~2%) of infected CD4<sup>+</sup> T cells carry replication-competent proviruses (35) that are inducible and able to produce plasma viremia and infect other target cells. Many Env sequences that we included in our analysis might therefore belong to defective proviruses, and their relevance for cure strategies is uncertain. Recent data, however, suggest that defective proviruses may produce HIV-1 antigens and perturb adaptive immunity in vivo, potentially contributing to the persistent immune activation observed in patients on

ART (36, 37). With the advances in full-length proviral sequencing technology (38, 39), we might be able to focus our analysis in the near future on proviruses that have been determined intact and therefore eliminate noise in the neutralization prediction by excluding irrelevant Env sequences that belong to defective proviruses. Furthermore, mapping the HIV reservoir at a single time point likely will not demonstrate the entire viral diversity that is present in an individual, and potential ongoing viral replication could lead to a change of the neutralization susceptibility pattern over time. It is, however, to some degree reassuring that longitudinal studies examining reactivatable latent viruses in patients under ART (27) have not revealed any evidence of viral evolution in the blood compartment, and a large proportion (>50%) of the inducible, replication-competent HIV reservoir derives from CD4<sup>+</sup> T cells carrying identical proviral sequences (40). Moreover, Lorenzi et al. has examined a large number of independent isolates of replication-competent virus from treated patients and found that >50% share sequence identity in the *env* gene with other isolates from the same patients (41). With adequate deep-sequencing approaches of the HIV reservoir producing significantly more single-genome sequencing than what we present here, minor viral variants could be identified and included in the neutralization sensitivity mapping. It needs to be determined if all, even very rare, sequences need to be considered, as they could theoretically generate systemic viral rebound following ART interruption, or if a cut-off frequency of viral variants might be feasible.

Although the model was developed using cross-clade Env sequences from the Los Alamos HIV database and from GenBank, paired with mAb neutralization sensitivity values from the CATNAP database, the available data in these databases are biased toward clade B. Training the model in the future with more non-clade B sequence/neutralization data could potentially strengthen its prediction accuracy against such clades. Nonetheless, the current model detected known differences in clade-specific bNAb susceptibilities, as seen by the relatively high predicted IC50 values for the V2 bNAbs (PGDM1400,



**Figure 7. Comparison between  $\text{IC}_{50}$  prediction and measured  $\text{IC}_{50}$  values using isolated plasma HIV-1 envelope sequences from 6 participants of the NIH 15-I-0140 trial following viral rebound during ATI.** Twenty-four pseudoviruses were generated based on plasma virus envelope (Env) sequences.  $\text{IC}_{50}$  values were predicted based on the Env sequences, and neutralization of the respective pseudoviruses against the broadly neutralizing antibodies (bNAb) 3BNC117, VRC01, 10-1074, and PGT121 was measured using a TZM/b1 neutralization assay. The regression lines with 95% confidence interval boundaries are indicated in black and blue dotted lines. The Spearman's rho and its 2-sided  $P$  value are labeled.

CAP256-VRC26.25, etc.) that are intrinsically less active against clade B (Figure 3). Finally, our approach requires advanced sequencing and data analysis capabilities that might not be readily available everywhere. Next-generation sequencing technology, however, is becoming more widely available, and the analytical algorithms can be formalized to facilitate access to a broader user base. Alternatively, a reference laboratory system could simplify access to reservoir bNAb susceptibility testing in the future.

In summary, sequence-based approaches to map the cellular reservoir for viral susceptibilities to antibody binding and neutralization are promising and might play a critical role in HIV prevention, but they are foremost promising for therapy strategies. Given the importance of glycans in the viral Env-antibody interaction, information on glycan occupancy should be a crucial component of prediction algorithms, and when included,

as we demonstrate here, prediction accuracy and reliability of such a model suggest its application for the selection of bNAb regimens, therefore supporting the advancement of bNAbs into the clinic.

## Methods

**Study participants.** Twenty-five HIV-1-infected adults  $\geq 18$  years old (23 on suppressive ART, with plasma HIV RNA  $< 50$  copies/mL, and 2 off ART, with plasma HIV RNA  $< 1000$  copies/mL) were included in this study. For the NIH study (NIH 15-I-0140), 9 HIV-1-infected individuals, ages  $\geq 18$  years old, with a  $\text{CD4}^+$  T cell count of  $\geq 450$  cells/mm<sup>3</sup> and an undetectable plasma viral load ( $< 50$  copies/mL) for at least or 3 years on continuous ART had been previously enrolled (11).

**Data set collection and preprocessing.** To develop a neutralization prediction model that identified the relationship between HIV Env sequence and mAb neutralization sensitivity, all available full-length Env protein sequences, spanning all different clades and neutralization tiers, were first collected from Los Alamos HIV database and GenBank. The neutralization sensitivity fingerprints of currently available anti-HIV monoclonal neutralizing antibodies were obtained from CATNAP (23), a web platform comprehensively collecting the published neutralization Env panel information. For those mAb neutralization values against the same Env panel repeatedly validated by multiple reports, the geometric median was taken among the values.

To accurately align the collected sequences, Clustal Omega (42), a multiple sequence alignment program that uses a hidden Markov model to help guide alignments, was implemented to align the sequences against the curated HIV Env alignments downloaded from Los Alamos HIV database (43) (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>). The positions of the protein residues for each aligned sequence were mapped back to the HXB2 sequence. The PNGSs were also identified in each sequence based on the consensus sequon sequence (N-X-S/T, X $\neq$ P). Additionally, to reduce the burden of computing and to relieve inconsistent alignments in highly variable V1 and V2 regions of Env affecting model prediction, the aligned residue positions in which more than 10% total collected sequences assigned as gap were removed for the later analysis.

*Glycan occupancy prediction for N-linked glycosylation sites.* Our previous report has shown that the presence of a PNGS is not sufficient to link the presence of the associated glycan (19) and the sequence usage signature around the PNGS was associated to different degrees of the site-specific glycan occupancy (19). Therefore, a predictive model modified from Li et al. (44) was developed to predict glycan occupancy based on the neighboring protein sequence of the PNGS. To build the model, the site-specific glycan occupancy profiles defined by mass spectrometry from a set of 149 selected Env proteins, including monomer gp120s and trimers, were collected from our previous study. Only the sites where the total spectral counts were larger than 300 were used in the model training to avoid instrumental noise due to low spectral signals, and therefore, a total of 766 PNGSs were collected, where the glycan occupancy ranged between 0 and 1.

Initially the neighboring protein sequences were extracted by a local sequence window between position  $-N$  and  $N$ , where the Asn of PNGS was at position 0 and  $N = 30$ . The extracted protein residues were encoded as a 1,220-dimensional vector (61 positions  $\times$  20 AA types), where every single residue was converted from a categorized AA into a binary vector. In addition to protein residue information, hydrophobicity, flexibility, polarity, and  $\beta$ -turn values of individual residues at every position were extracted from AAindex database (45). Elastic Net regularization was used for feature reduction, and the  $\alpha$  in Elastic Net optimization, controlling the weight of L1 norm and L2 norm, was determined by 10-fold cross validation. The selected features used in machine-learning algorithm can be accessed in Supplemental Data Set 1. To benchmark the prediction performance, several supervised machine-learning algorithms were tested, including SVR, k-nearest neighbors, random forest, and neural network. SVR showed the most robust prediction, outperforming other algorithms. The optimal hyperparameters for SVR were determined via grid search in the parameter space based on 10-fold cross validation, where kernel = radial basis function, gamma =  $2.28 \times 10^{-6}$ , Cost =  $1 \times 10^5$ , epsilon\_SVR = 9, epsilon = 0.001. The model prediction performance in each step was evaluated by  $R^2$  and mean squared error from repeated 10-fold CV framework.

The final SVR predictive model with the optimal parameters was trained by the complete glycan occupancy data set (766 PNGSs) with the selected features. The model was then used to predict the glycan occupancy of individual PNGSs from the collected HIV Env sequences based on the corresponding sequence features. The predicted occupancy served as part of input variables feeding into the mAb neutralization predictive models later.

*mAb neutralization sensitivity prediction.* The Bayesian MCMC support vector machine regression model developed in the previous study (19) was used for feature reduction and mAb neutralization sensitivity prediction. Bayesian MCMC-SVR integrates 2 computational methods. First, a Bayesian MCMC with the biased random walk Metropolis-Hastings algorithm (46, 47) is used to approximate the high-dimensional posterior distribution within the data to identify an optimal feature set that best fits the model predictions, and second, a SVR algorithm that implements support vector machines (SVMs) (48) is employed as a supervised learning model to predict neutralization sensitivity based on the selected feature profiles.

The feature matrix for model learning contained 2 parts. (a) The sequence features included every position of the aligned Env protein sequence, which was converted from a categorized AA residue into a binary vector. (b) The glycan features included the PNGSs from the sequence, which were given the predicted glycan occupancy values from 0 (not occupied) to 1 (fully occupied) estimated by the glycan predictive model in the previous section. The mAb neutralization values (IC<sub>50</sub>) for model regression were log<sub>2</sub> transformed to improve the normality of the data distribution. To determine the optimal feature set of the key protein residues and PNGSs, the statistical significance was calculated for individual features based on the permutation test that estimated null distribution. The detail notion of model construction, statistical calculation, and model benchmark can be found in the Materials and methods section in Yu et al. (19). The predictive model was built for each individual mAb neutralization profile available in CATNAP, with more than 70 Env panels. A total of 126 mAb models were constructed in this report (Supplemental Table 1), and the optimal hyperparameters for every mAb model was determined via grid search in the parameter space based on 10-fold cross validation. The optimal feature sets and hyperparameters of individual models can be accessed in Supplemental Data Set 2.

The customized Python scripts were coded to process Env sequence and neutralization data set. The codes for Bayesian MCMC-SVR can be accessed in Yu et al. (19), and they require Matlab LIBSVM package. The algorithm was run on Linux high-performance computing clusters.

*Virus culture.* Autologous virus was retrieved from HIV-1-infected individuals as previously described (49). Briefly, CD4<sup>+</sup> T cells were purified from freshly isolated PBMCs by negative selection using the

Rosette Sep CD4<sup>+</sup> T cell enrichment cocktail (Stemcell Technologies), depleting CD8<sup>+</sup> T cells, NK cells, B cells, macrophages, monocytes, and dendritic cells. CD4<sup>+</sup> T cells were then stimulated in IL-2 (50 units/ml) containing T cell medium in the presence of a bispecific anti-CD3/anti-CD8 mAb, which selectively activates CD4<sup>+</sup> T-lymphocytes, while simultaneously depleting all remaining CD8<sup>+</sup> T cells (50). CD4<sup>+</sup> T cell blasts, generated from HIV negative donors, were added every 7 days to maintain the cultures and provide additional targets for viral outgrowth. Every 2–3 days, p24 was measured in culture supernatants using a p24 based ELISA (PerkinElmer Life Sciences Inc., following the manufacturer's protocol).

*mAbs.* VRC01 was provided by the Vaccine Research Center (VRC) at NIH. 3BNC117, and 10-1074 was acquired from the NIH AIDS Reagent Program. PGT121 and PGDM1400 were provided by the Virology and Vaccine Research Center at Beth Israel Deaconess Medical Center. The same set of mAbs was used in Figure 2 and Figure 7.

*Neutralization assays.* Neutralization of autologous, PBMC-derived infectious isolates by VRC01, 3BNC117, PGDM1400, and PGT121 was assessed as previously described (51). TCID<sub>50</sub>s were determined for all HIV-1-containing supernatant by serial dilutions and infection of TZM-bl cells. For neutralization assays, mAb samples were tested using a primary concentration of 25 µg/ml with 5-fold dilution series. The viral isolates and mAbs were incubated in a 96-well plate at 37°C for 1 hour before addition of TZM-bl cells. After 2 days of incubation, cells were lysed and the viral infectivity was quantified by measuring luciferase activity using a Victor3 luminometer (Perkin-Elmer). Neutralization curves were fit by non-linear regression using a 5-parameter hill slope equation. The IC<sub>50</sub> values were reported as the antibody concentrations required to suppress infection by 50%.

*Single-genome amplification and analysis.* The Env gene was sequenced using single-genome amplification as described previously (52). Briefly, viral RNA was isolated from blood plasma using the QIAamp Viral RNA Mini kit (Qiagen) and subjected to cDNA synthesis using the primer Env3outR1 (TTGCTACTGTGATTGCTCCATGT). Newly synthesized cDNA was diluted to single-copy concentration and amplified by nested PCR using the primer sets Env3outR1 and Env5outF1 (TAGAGCCCTGGAAGCATCCAGGAAG) and Env3inR2 (GTCTCGAGATACTGCTCCACCC) and Env5inF2 (TTAGGCATCTCCTATGGCAGGAAGAAG). Resultant amplicons were sequenced using Sanger-based sequencing. In total, 727 sequences were obtained from the original 25 patients, and an additional 199 Env sequences were obtained from VRC study participants. The mean pairwise diversity measurements were obtained using DIVEIN (53). Phylogenetic trees were generated using the Tamura-Nei neighbor joining method. All 926 sequences were deposited in GenBank.

*Statistics.* Spearman's rank analysis (2 tailed) was used to test for correlations between variables. A 2-sided Wilcoxon rank-sum test was used for pairwise analyses comparing antibody neutralization responses across the groups. Comparisons between more than 2 groups were performed with 1-way ANOVA. A *P* value of less than 0.05 was considered significant. The Benjamini and Hochberg method was used to calculate FDR-adjusted *P* values for multiple-comparison post hoc correction.

*Study approval.* All study participants provided written informed consent prior to study participation in accordance with the Declaration of Helsinki, and studies were reviewed by the institutional review boards of each participating institution (Partners Human Research Committee, Somerville Massachusetts, USA, and National Institute of Allergy and Infectious Diseases Institutional Review Board).

## Author contributions

CMF and BFK performed single-genome amplification and HIV-1 sequencing. DS and JT performed viral outgrowth assays. MSS performed viral neutralization assays. RL, AS, and NDR generated pseudoviruses and neutralization data. TWC contributed PBMC samples. WHY, GA, and DAL developed the models. WHY, GA, DAL, and BJ conceived and designed the study and analyzed and interpreted the data. DAL and BJ supervised the study.

## Acknowledgments

BJ is supported by NIH grants AI106408 and AI138790, and by the Ragon Institute of MGH, MIT, and Harvard. DAL is supported by NIH grants AI142790, CA210180, CA215798, CA217377, and DK108056. RL is supported by NIH grant AI140788. GA is supported by NIH grants AI060354, AI080289, AI128751, AI142790, and AI129797. BJ and GA are associate members of the I4C-HIV Martin Delaney Collaboratory (UM1AI126603). The authors gratefully acknowledge John Mascola from the NIH Vaccine Research

Center for providing mAbs. This project has been funded in part with federal funds from the National Cancer Institute, NIH, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

Address correspondence to: Boris Julg, Ragon Institute of MGH, MIT and Harvard, 400 Technology Square, Cambridge, Massachusetts 02139, USA. Phone: 857.268.7088; Email: [bjulg@mgh.harvard.edu](mailto:bjulg@mgh.harvard.edu).

1. Corbeau P, Reynes J. Immune reconstitution under antiretroviral therapy: the new challenge in HIV-1 infection. *Blood*. 2011;117(21):5582–5590.
2. Hunt PW. HIV and inflammation: mechanisms and consequences. *Curr HIV/AIDS Rep*. 2012;9(2):139–147.
3. Lynch RM, et al. Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Sci Transl Med*. 2015;7(319):319ra206.
4. Barouch DH, et al. Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature*. 2013;503(7475):224–228.
5. Caskey M, et al. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*. 2015;522(7557):487–491.
6. Julg B, et al. Protection against a mixed SHIV challenge by a broadly neutralizing antibody cocktail. *Sci Transl Med*. 2017;9(408):eaao4235.
7. Julg B, et al. Virological Control by the CD4-Binding Site Antibody N6 in Simian-Human Immunodeficiency Virus-Infected Rhesus Monkeys. *J Virol*. 2017;91(16):e00498-17.
8. Julg B, et al. Protective Efficacy of Broadly Neutralizing Antibodies with Incomplete Neutralization Activity against Simian-Human Immunodeficiency Virus in Rhesus Monkeys. *J Virol*. 2017;91(20):e01187-17.
9. Julg B, et al. Broadly neutralizing antibodies targeting the HIV-1 envelope V2 apex confer protection against a clade C SHIV challenge. *Sci Transl Med*. 2017;9(406):eaal1321.
10. Caskey M, et al. Antibody 10-1074 suppresses viremia in HIV-1-infected individuals. *Nat Med*. 2017;23(2):185–191.
11. Bar KJ, et al. Effect of HIV Antibody VRC01 on Viral Rebound after Treatment Interruption. *N Engl J Med*. 2016;375(21):2037–2050.
12. Scheid JF, et al. HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. *Nature*. 2016;535(7613):556–560.
13. Mendoza P, et al. Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature*. 2018;561(7724):479–484.
14. Ho YC, et al. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell*. 2013;155(3):540–551.
15. Wang Z, Simonetti FR, Siliciano RF, Laird GM. Measuring replication competent HIV-1: advances and challenges in defining the latent reservoir. *Retrovirology*. 2018;15(1):21.
16. Stewart-Jones GB, et al. Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell*. 2016;165(4):813–826.
17. Lee JH, Ozorowski G, Ward AB. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science*. 2016;351(6277):1043–1048.
18. Panico M, et al. Mapping the complete glycoproteome of virion-derived HIV-1 gp120 provides insights into broadly neutralizing antibody binding. *Sci Rep*. 2016;6:32956.
19. Yu WH, et al. Exploiting glycan topography for computational design of Env glycoprotein antigenicity. *PLoS Comput Biol*. 2018;14(4):e1006093.
20. Hake A, Pfeifer N. Prediction of HIV-1 sensitivity to broadly neutralizing antibodies shows a trend towards resistance over time. *PLoS Comput Biol*. 2017;13(10):e1005789.
21. Bricault CA, et al. HIV-1 Neutralizing Antibody Signatures and Application to Epitope-Targeted Vaccine Design. *Cell Host Microbe*. 2019;25(1):59–72.e8.
22. Magaret CA, et al. Prediction of VRC01 neutralization sensitivity by HIV-1 gp160 sequence features. *PLoS Comput Biol*. 2019;15(4):e1006952.
23. Yoon H, et al. CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res*. 2015;43(W1):W213–W219.
24. Rothenberger MK, et al. Large number of rebounding/founder HIV variants emerge from multifocal infection in lymphatic tissues after treatment interruption. *Proc Natl Acad Sci USA*. 2015;112(10):E1126–E1134.
25. Huang J, et al. Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan Neutralization Breadth. *Immunity*. 2016;45(5):1108–1121.
26. Cheng HD, et al. Fine epitope signature of antibody neutralization breadth at the HIV-1 envelope CD4-binding site. *JCI Insight*. 2018;3(5):97018.
27. Brodin J, et al. Establishment and stability of the latent HIV-1 DNA reservoir. *Elife*. 2016;5:e18889.
28. Wagh K, et al. Optimal Combinations of Broadly Neutralizing Antibodies for Prevention and Treatment of HIV-1 Clade C Infection. *PLoS Pathog*. 2016;12(3):e1005520.
29. Kong R, et al. Improving neutralization potency and breadth by combining broadly reactive HIV-1 antibodies targeting major neutralization epitopes. *J Virol*. 2015;89(5):2659–2671.
30. Cohen YZ, et al. Relationship between latent and rebound viruses in a clinical trial of anti-HIV-1 antibody 3BNC117. *J Exp Med*. 2018;215(9):2311–2324.
31. Buiu C, Putz MV, Avram S. Learning the Relationship between the Primary Structure of HIV Envelope Glycoproteins and Neutralization Activity of Particular Antibodies by Using Artificial Neural Networks. *Int J Mol Sci*. 2016;17(10):E1710.
32. Sok D, et al. Promiscuous glycan site recognition by antibodies to the high-mannose patch of gp120 broadens neutralization of

- HIV. *Sci Transl Med*. 2014;6(236):236ra63.
33. Pritchard LK, et al. Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat Commun*. 2015;6:7479.
  34. Behrens AJ, et al. Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Rep*. 2016;14(11):2695–2706.
  35. Bruner KM, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med*. 2016;22(9):1043–1049.
  36. Pollack RA, et al. Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. *Cell Host Microbe*. 2017;21(4):494–506.e4.
  37. Laird GM, et al. Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog*. 2013;9(5):e1003398.
  38. Hiener B, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4<sup>+</sup> T Cells from Effectively Treated Participants. *Cell Rep*. 2017;21(3):813–822.
  39. Lee GQ, et al. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4<sup>+</sup> T cells. *J Clin Invest*. 2017;127(7):2689–2696.
  40. Bui JK, et al. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. *PLoS Pathog*. 2017;13(3):e1006283.
  41. Lorenzi JC, et al. Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA. *Proc Natl Acad Sci USA*. 2016;113(49):E7908–E7916.
  42. Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
  43. Gaschen B, Kuiken C, Korber B, Foley B. Retrieval and on-the-fly alignment of sequence fragments from the HIV database. *Bioinformatics*. 2001;17(5):415–418.
  44. Li F, et al. GlycoMine<sup>struct</sup>: a new bioinformatics tool for highly accurate mapping of the human N-linked and O-linked glycoproteomes by incorporating structural features. *Sci Rep*. 2016;6:34595.
  45. Kawashima S, Pokarowski P, Pokarowska M, Kolinski A, Katayama T, Kanehisa M. AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res*. 2008;36(Database issue):D202–D205.
  46. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970;57(1):97–109.
  47. Andrieu C, De Freitas N, Doucet A, Jordan MI. An introduction to MCMC for machine learning. *Machine learning*. 2003;50(1–2):5–43.
  48. Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995;20(3):273–297.
  49. Julg B, et al. Infrequent recovery of HIV from but robust exogenous infection of activated CD4(+) T cells in HIV elite controllers. *Clin Infect Dis*. 2010;51(2):233–238.
  50. Wilson CC, et al. Ex vivo expansion of CD4 lymphocytes from human immunodeficiency virus type 1-infected persons in the presence of combination antiretroviral agents. *J Infect Dis*. 1995;172(1):88–96.
  51. Sarzotti-Kelsoe M, et al. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J Immunol Methods*. 2014;409:131–146.
  52. Keele BF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA*. 2008;105(21):7552–7557.
  53. Deng W, et al. DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *BioTechniques*. 2010;48(5):405–408.