



Prevention efficacy of the broadly neutralizing antibody VRC01 depends on HIV-1 envelope sequence features

Michal Juraska^{a,1}, Hongjun Bai^{b,c}, Allan C. deCamp^a, Craig A. Magaret^a, Li Li^a, Kevin Gillespie^a, Lindsay N. Carpp^a, Elena E. Giorgi^a, James Ludwig^a, Cindy Molitor^a, Aaron Hudson^a, Brian D. Williamson^{a,d}, Nicole Espy^e, Brian Simpkins^f, Erika Rudnicki^a, Danica Shao^a, Raabya Rossenkhan^a, Paul T. Edlefsen^a, Dylan H. Westfall^g, Wenjie Deng^g, Lennie Chen^g, Hong Zhao^g, Tanmoy Bhattacharya^h, Alec Pankow^{i,2}, Ben Murrellⁱ, Anna Ysselⁱ, David Matter^j, Talita York^k, Nicolas Beaume^l, Asanda Gwashu-Nyangiwe^l, Nonkululeko Ndabambi^l, Ruwayhida Thebus^j, Shelly T. Karuna^a, Lynn Morris^{k,l,m}, David C. Montefioriⁿ, John A. Huralⁿ, Myron S. Cohen^o, Lawrence Corey^{p,q,r}, Morgane Rolland^{b,c}, Peter B. Gilbert^{a,s,t,3}, Carolyn Williamson^{h,3}, and James I. Mullins^{g,t,u,3}

Edited by Marina Caskey, The Rockefeller University, New York, NY; received June 9, 2023; accepted November 13, 2023 by Editorial Board Member Stephen P. Goff

In the Antibody Mediated Prevention (AMP) trials (HVTN 704/HPTN 085 and HVTN 703/HPTN 081), prevention efficacy (PE) of the monoclonal broadly neutralizing antibody (bnAb) VRC01 (vs. placebo) against HIV-1 acquisition diagnosis varied according to the HIV-1 Envelope (Env) neutralization sensitivity to VRC01, as measured by 80% inhibitory concentration (IC80). Here, we performed a genotypic sieve analysis, a complementary approach to gaining insight into correlates of protection that assesses how PE varies with HIV-1 sequence features. We analyzed HIV-1 Env amino acid (AA) sequences from the earliest available HIV-1 RNA-positive plasma samples from AMP participants diagnosed with HIV-1 and identified Env sequence features that associated with PE. The strongest Env AA sequence correlate in both trials was VRC01 epitope distance that quantifies the divergence of the VRC01 epitope in an acquired HIV-1 isolate from the VRC01 epitope of reference HIV-1 strains that were most sensitive to VRC01-mediated neutralization. In HVTN 704/HPTN 085, the Env sequence-based predicted probability that VRC01 IC80 against the acquired isolate exceeded 1 $\mu\text{g}/\text{mL}$ also significantly associated with PE. In HVTN 703/HPTN 081, a physicochemical-weighted Hamming distance across 50 VRC01 binding-associated Env AA positions of the acquired isolate from the most VRC01-sensitive HIV-1 strain significantly associated with PE. These results suggest that incorporating mutation scoring by BLOSUM62 and weighting by the strength of interactions at AA positions in the epitope:VRC01 interface can optimize performance of an Env sequence-based biomarker of VRC01 prevention efficacy. Future work could determine whether these results extend to other bnAbs and bnAb combinations.

sieve analysis | epitope | HIV diversity | PAR score | Hamming distance

HIV continues to pose a significant global public health challenge. In 2021, 38.4 million people were living with HIV, 1.5 million people newly acquired HIV, and 650,000 people died from AIDS-related illnesses (1). Considerable effort has focused on monoclonal broadly neutralizing antibodies (bnAbs) for HIV-1 prevention (2–5), and a proof of concept that a bnAb can prevent HIV-1 acquisition was established by the Antibody Mediated Prevention (AMP) trials [HVTN 704/HPTN 085 (NCT02716675) and HVTN 703/HPTN 081 (NCT02568215)] (6).

In the AMP trials, participants were randomized 1:1:1 to receive 10 intravenous infusions, administered at 8-wk intervals, of 10 mg/kg VRC01 [a bnAb that targets the CD4 binding site of the HIV-1 envelope glycoprotein (7)], 30 mg/kg VRC01, or placebo (saline). HVTN 704/HPTN 085 (hereafter referred to as “the Americas trial”) was conducted in Brazil, Peru, Switzerland, and the United States and enrolled 2699 HIV-uninfected men and transgender persons who have sex with men. HVTN 703/HPTN 081 (hereafter referred to as “the Africa trial”) was conducted in Botswana, Kenya, Malawi, Mozambique, South Africa, Tanzania, and Zimbabwe and enrolled 1924 HIV-uninfected sexually active women. There was no statistically significant overall prevention efficacy (PE) of VRC01 vs. placebo against documented HIV-1 acquisition (defined by an HIV-1 RNA PCR-positive test) by the week 80 visit in either trial. In the Americas trial, estimated PE was 26.6% (95% CI, –11.7 to 51.8; $P = 0.15$) for both VRC01 dose groups pooled, 22.4% (95% CI, –25.5 to 52.0) for the 10 mg/kg group, and 30.9% (95% CI, –13.9 to 58.0) for the 30 mg/kg group. In the Africa trial, estimated PE was 8.8% (95% CI, –45.1 to 42.6; $P = 0.70$) for both VRC01 dose groups pooled, –9.3% (95% CI, –85.3 to 35.5)

Significance

While there was no statistically significant overall prevention efficacy against HIV-1 diagnosis of the monoclonal broadly neutralizing antibody (bnAb) VRC01 vs. placebo in the Antibody Mediated Prevention trials, VRC01 prevented detection of HIV-1 sequences from viruses that were sensitive to VRC01-mediated neutralization. We found characteristics of HIV-1 Env AA sequences, obtained from AMP trial participants who were diagnosed with HIV-1, that associated with VRC01 prevention efficacy. One application of the Env sequence correlates is to improve ranking and selection of bnAb regimens by their predicted magnitude and breadth of prevention efficacy against a population of circulating HIV-1 strains in a geographic region where a future efficacy trial may be conducted.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.C. is a guest editor invited by the Editorial Board.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: mjuraska@fredhutch.org.

²Present address: Department of Microbiology, The Icahn School of Medicine at Mount Sinai, New York, NY 10029.

³P.B.G., C.W., and J.I.M. contributed equally to this work.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2308942121/-DCSupplemental>.

Published January 19, 2024.

Table 1. HIV-1 infection diagnosis primary end points and available HIV-1 Env sequence data by trial and treatment assignment

	HVTN 704/HPTN 085 (Americas Trial)				HVTN 703/HPTN 081 (Africa Trial)			
	VRC01 10 mg/kg	VRC01 30 mg/kg	Placebo	Total	VRC01 10 mg/kg	VRC01 30 mg/kg	Placebo	Total
No. of primary end points*	32	28	38	98	28	19	29	76
No. of primary end points with available first RNA+ sample Env sequence data ^{†,‡}	32	28	38	98	28	17	29	74
No. of primary end points with available first RNA+ sample Env sequence and IC80 data ^{†,‡}	28	27	35	90	26	17	29	72
No. of primary end points as single-lineage infections	23	21	25	69	21	10	18	49
No. of Env sequences per primary end point: mean (range) [†]	179 (16, 657)	147 (20, 496)	215 (25, 988)	184 (16, 988)	140 (5, 672)	172 (10, 591)	168 (9, 464)	158 (5, 672)

*HIV-1 infection diagnosis per the trial's testing algorithm by 595 d (85 wk) since enrollment.

[†]If the first positive sample contributed <20 viral RNA templates, the second positive sample was sequenced, and both sets of sequences were combined for sieve analysis. This occurred for five primary end points in the Americas trial and none in the Africa trial.

[‡]Sixteen percent of primary end points had sequences sampled from preseroconversion samples.

for the 10 mg/kg group, and 27.0% (95% CI, -30.7 to 59.3) for the 30 mg/kg group.

However, PE against new HIV-1 diagnosis significantly decreased with HIV-1 isolate resistance level to neutralization by the VRC01 clinical lot as measured by 80% inhibitory concentration (IC80) with the TZM-bl target cell assay (8, 9). First, prespecified analyses showed that PE against RNA-detectable HIV-1 isolates that were sensitive to VRC01-mediated neutralization [defined as IC80 ≤ 1 µg/mL in the TZM-bl target cell assay] was 75.4% (95% CI, 45.5 to 88.9) for both VRC01 dose groups pooled vs. placebo combining across both trials (6). This result also held in each individual trial: 73.0% (95% CI, 27.6 to 89.9) in the Americas trial and 78.6% (95% CI, 17.3 to 94.4) in the Africa trial. In contrast, PE against RNA-detectable virus with IC80 > 1 µg/mL was close to zero in each trial. Second, PE in each trial decreased steadily over the range of acquired isolate IC80 values 0.1 µg/mL to >10 µg/mL, as assessed by multiple statistical methods (6). These “neutralization sieve analyses” required measurements of in vitro neutralizing antibody potency of the clinical product against Env-pseudotyped viruses, which are relatively resource-intensive to obtain.

Genotypic sieve analysis is a complementary approach for obtaining additional mechanistic insights into correlates of protection and uses pathogen sequence data to assess how PE varies with sequence features of the pathogen (10, 11). Pertinent to the AMP trials, these sequence and neutralization data are related in that Bricault et al. (12) and our group (13) previously identified models that predicted the resistance level of a given HIV-1 Env-pseudotyped virus to VRC01-mediated neutralization based on HIV-1 Env amino acid (AA) sequence data. We also identified Env AA sequence features that ranked highly for predicting neutralization resistance of a given Env-pseudotyped virus to VRC01 (13). In a treatment interruption study with VRC01 infusions, Cale and colleagues showed that participants with short VRC01 epitope distances to most-VRC01-sensitive sequences rebounded later than participants with larger VRC01 epitope distances (14). Here, using Env sequences from the earliest available HIV-1 RNA-positive plasma samples from participants who were diagnosed with HIV-1 in the AMP trials, we assessed whether and how VRC01 PE varied with HIV-1 Env AA characteristics.

Results

HIV-1 Diagnosis Primary End Points and Available Sequence Data. The AMP trials' primary end point was HIV-1 infection diagnosis (per the trials' testing algorithm) by the week 80 visit (6). There were 98 and 74 HIV-1 primary end points with available HIV-1 sequence data from the earliest RNA-positive samples in the Americas and Africa trials, respectively (Table 1). Table 1 provides descriptive totals by trial and study group including numbers of HIV-1 primary end points with single (vs. multiple) HIV-1 lineages detected in the first virus-positive plasma sample.

Greater Subtype Diversity in the Americas than the Africa Trial.

Phylogenetic analysis of *env* nucleotide sequences showed that subtype B was dominant in the Americas trial, accounting for 77% of isolated viruses (*SI Appendix, Fig. S1A*). In Peru, where most of the HIV-1 diagnoses occurred, 26% of isolated viruses were non-B (mainly F or BF recombinant), compared to the United States where only 5% were non-B subtypes (*SI Appendix, Figs. S1A and S2A*). The isolated viruses of Peruvian participants tended to be more diverse than those from the United States., Brazil, and Switzerland, resulting in longer branch lengths in the midpoint-rooted phylogenetic tree (*SI Appendix, Fig. S2A*). In the Africa trial, 97% of the isolated viruses were subtype C, with one subtype G and one A1C recombinant identified (*SI Appendix, Fig. S1B*).

Prevention Efficacy Decreases with Increasing Sequence-Predicted Resistance to VRC01 Neutralization.

Using Los Alamos' National Laboratory's (LANL) CATNAP database (15), a SuperLearner modeling tool (13, 16) was used to predict in vitro VRC01 neutralization resistance of isolated viruses based on their Env AA sequences. Three different Env sequence-derived predictive measures of neutralization resistance, termed proteomic antibody resistance (PAR) scores (13), were calculated. The following analysis employed a PAR score defined as the logit (log odds) of the predicted probability that an isolated virus with the given Env sequence is resistant to neutralization by VRC01, with resistance defined as IC80 > 1 µg/mL. The PAR score was only weakly correlated with the experimentally measured in vitro IC80 (treatment group-adjusted Spearman rank correlation coefficient of 0.20 in each AMP trial) (Fig. 1 and *SI Appendix, Fig. S16*).

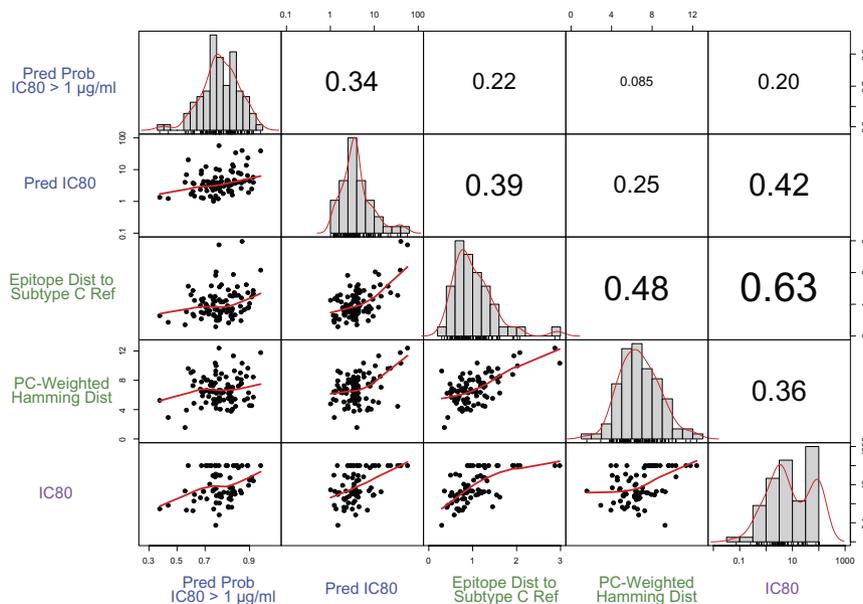


Fig. 1. Features discriminating prevention efficacy in HVTN 703/HPTN 081 (Africa trial). Treatment-adjusted Spearman rank correlations between pairs of the following five viral features of the predicted most resistant lineages isolated from the primary end point cases: predicted probability of IC80 > 1 µg/mL, predicted quantitative IC80 (µg/mL), VRC01 epitope distance from the subtype C VRC01-sensitive reference sequence, physicochemical-weighted Hamming distance from the VRC01-sensitive reference sequence, and IC80 measured by the TZM-bl target cell assay. The font sizes of the correlations are proportional to their magnitude. The scatter plots are superimposed with LOWESS curves and the histograms with probability density estimates.

Analyses of the two other PAR scores, predicted quantitative IC80 and predicted categorized IC80, are presented in supplement with consistent conclusions.

Whether and how VRC01 vs. placebo PE against HIV-1 diagnosis varied with the level of each PAR score was assessed. In the Americas trial, PE of both VRC01 dose groups pooled significantly declined with the PAR score, from 78% (95% CI, 20 to 94) against isolated viruses with predicted 33% probability of resistance to 0% (95% CI, -71 to 41) against isolated viruses with predicted 73% probability of resistance ($P = 0.021$) (Fig. 2A). In the Africa trial, the declining trend in PE of both VRC01 dose groups pooled with an increasing predicted probability of resistance was not significant ($P = 0.25$) (Fig. 2B). The trial difference in the strength of sieve effect evidence might be partially explained by more resistant isolated viruses (i.e., higher IC80s) and fewer primary end points observed among placebo recipients in the Africa than the Americas trial (Fig. 2). Most isolated viruses in the Africa trial were subtype C, which tend to be more resistant to VRC01 neutralization than other clades (12).

Fig. 2 C–F shows separate PE of the 30 vs. 10 mg/kg VRC01 dose regimens in each trial, with the steepest decline in PE with increasing predicted resistance found in the Americas trial's 30 mg/kg dose regimen (Fig. 2C), followed by similar intermediate sieve effects in the Americas trial's 10 mg/kg and Africa trial's 30 mg/kg dose regimens (Fig. 2 E and D). There was no variation in PE observed for the 10 mg/kg regimen in the Africa trial (Fig. 2F). These results suggest that a higher bnAb dose was required in the Africa trial to offset the greater resistance of subtype C viruses to yield similarly protective neutralization titers achieved by the lower dose in the Americas trial with predominantly subtype B viruses. *SI Appendix, Figs. S3 and S4* show that the results are similar for the two other PAR scores.

IC80 measurements were available from 94% of the primary end points with sequence data (92% in the Americas and 97% in the Africa trial) (Table 1). Hence, we assessed variation in PE with each quantitative PAR score for each of two VRC01 neutralization phenotypes defined by in vitro IC80 ≤ 1 or > 1 µg/mL to

investigate the relative contribution of the PAR score vs. IC80 in discriminating PE. In both trials, PE was estimated to be uniformly higher across all PAR score levels against diagnosis of IC80 ≤ 1 µg/mL viruses than against diagnosis of IC80 > 1 µg/mL viruses, supporting that the IC80 biomarker contains information discriminating PE beyond that captured by the PAR score (Fig. 3). Conversely, a consistent decline in the estimated PE with increasing predicted resistance within each of the in vitro dichotomized IC80 phenotypes was observed in the Americas trial only, which suggests that the PAR score carries PE-discriminating information about subtype B viruses not captured by the dichotomized IC80 biomarker. *SI Appendix, Fig. S5* shows that the results were similar for the predicted quantitative IC80 PAR score.

Lack of Significant PE Variation with Env Sequence Features Predictive of VRC01 Neutralization Resistance. We preidentified 24 Env sequence features with the highest variable importance measure for predicting VRC01 neutralization resistance [12 alignable AA positions, 5 potential N-linked glycosylation site (PNGS) motifs, and 7 viral geometry features] (Table 2, *SI Appendix, Figs. S6 and S7*). For each eligible feature, we evaluated variation in PE across distinct levels/genotypes of the feature, where eligibility required ≥ 6 primary end points representing each level of a binary feature. None of the 18 and 19 eligible features passed the multiple hypothesis testing adjusted significance bar for sieve effect evidence in the Americas and Africa trial, respectively, with all Q-values > 0.38 and > 0.58 (*SI Appendix, Figs. S8–S11*). Six and five binary features in the Americas and Africa trial, respectively, had low variability, limiting statistical power to detect sieve effects for those features. However, individual neutralization-associated sequence features are expected to have a weaker ability to discriminate PE than PAR scores, which aggregate over many features predictive of neutralization resistance.

Signal of VRC01 Protection against Diagnosis of Viruses with a PNGS Motif at Env Positions 230 to 232. Despite not meeting the higher significance bar required due to testing multiplicity,

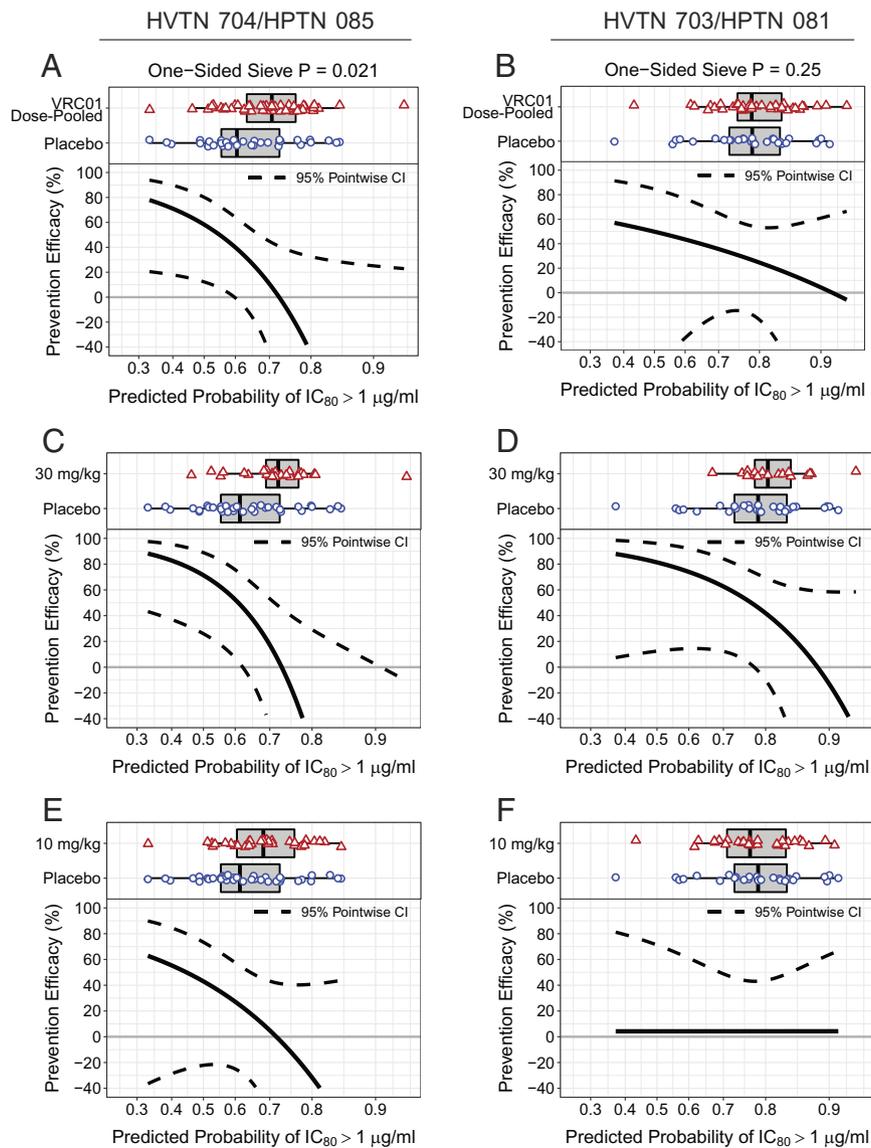


Fig. 2. Prevention efficacy (PE) of VRC01 treatment in the AMP trials by PAR score defined as the logit of the predicted probability that an isolated virus is resistant to neutralization by VRC01. The solid black curves show PE of both VRC01 dose groups pooled (A and B), the 30 mg/kg VRC01 group (C and D), and the 10 mg/kg VRC01 group (E and F) as a function of the predicted probability that the predicted most resistant lineage of the primary end point case in HVTN 704/HPTN 085 (Americas trial) (A, C, and E) or HVTN 703/HPTN 081 (Africa trial) (B, D, and F) has VRC01 80% inhibitory concentration (IC₈₀) > 1 µg/mL. The dashed black curves represent 95% pointwise CIs. Horizontal box plots at the top of each panel show the distributions of the predicted probability of IC₈₀ > 1 µg/mL PAR score for the predicted most resistant lineage, for primary end points in the designated VRC01 treatment group (open red triangles) or in the placebo group (open blue circles). “One-sided Sieve P” is a P-value from testing whether PE decreases with the predicted probability of IC₈₀ > 1 µg/mL.

we estimated in the Americas trial 64% PE (95% CI, 20 to 84) against viruses with a PNGS at Env positions 230 to 232 vs. -5% PE (95% CI, -70 to 36) against viruses with no PNGS at 230 to 232 (unadjusted $P = 0.027$ for differential PE) for both VRC01 dose groups pooled (Fig. 4A). Furthermore, VRC01 prevented acquisition of detectable viremia with both in vitro IC₈₀ ≤ 1 µg/mL and a PNGS at 230 to 232 [PE = 100% (95% CI, 36 to 100) using a binomial score method]: all six viruses with IC₈₀ ≤ 1 µg/mL isolated from VRC01 recipients lacked a PNGS at 230 to 232 compared to 75% of 12 viruses with IC₈₀ ≤ 1 µg/mL isolated from placebo recipients that lacked a PNGS at 230 to 232 in the Americas trial (Fig. 4B), associated primarily with a N to D amino acid substitution (Fig. 4C). We also observed a potential dose-response association between the VRC01 dose and PE against viruses exhibiting a PNGS at 230 to 232, with 50% PE (95% CI, -24 to 80) of the 10 mg/kg regimen and 79% PE (95% CI, 25 to 94) of the 30 mg/kg regimen (Fig. 4A). *SI Appendix, Fig. S12*

shows PE estimates restricted to subtype B viruses in the Americas trial. This result generates the hypothesis that, in viral subtypes observed in the Americas trial, a PNGS at 230 to 232 facilitates VRC01 binding. However, this sieve effect was not replicated in the Africa trial [PE 27% (95% CI, -27 to 58) vs. 13% (95% CI, -107 to 64) for isolated viruses with vs. without a PNGS at 230 to 232, respectively, for both VRC01 dose groups pooled] (Fig. 4A) despite a twofold higher probability of a PNGS at 230 to 232 among HIV-1 infected placebo recipients [70% (95% CI, 53 to 86) in the Africa vs. 35% (95% CI, 16 to 53) in the Americas trial] (*SI Appendix, Table S1*).

No evidence of Differential PE by Residue Presence vs. Absence in Comprehensive Site-Scanning Analysis. For each trial, we screened all AA positions located in the union of the VRC01 binding footprint (7) and the CD4 binding site (Table 2, *SI Appendix, Figs. S6 and S7*) with sufficient variability defined as ≥6 primary

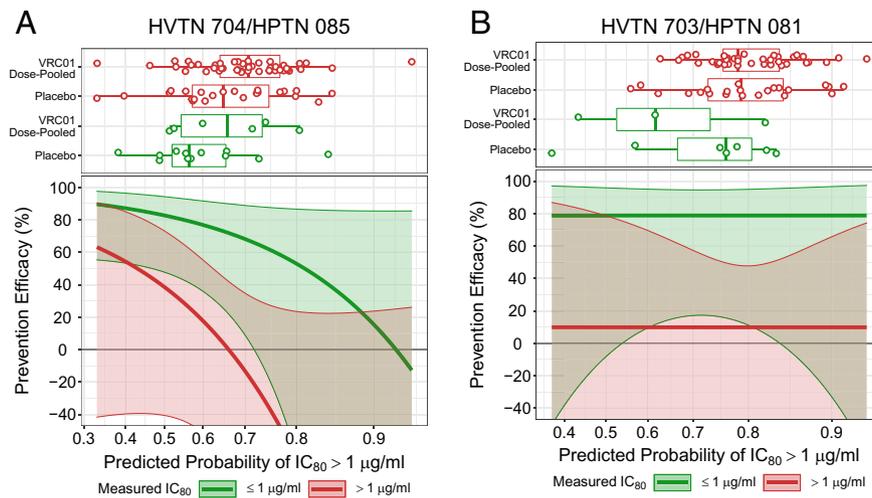


Fig. 3. Prevention efficacy (PE) of both VRC01 dose groups pooled vs. placebo by simultaneously i) the predicted probability that the predicted most resistant lineage of the primary end point case in HVTN 704/HPTN 085 (Americas trial) (A) or HVTN 703/HPTN 081 (Africa trial) (B) has VRC01 80% inhibitory concentration (IC₈₀) > 1 µg/mL and ii) the indicator of measured IC₈₀ > 1 µg/mL. The two curves show variation in PE with the predicted probability of IC₈₀ > 1 µg/mL for isolated viruses with measured IC₈₀ ≤ 1 µg/mL (green) or > 1 µg/mL (red). The green and red shaded areas represent 95% pointwise CIs. The horizontal box plots at the top of each panel show the distributions of the predicted probability of IC₈₀ > 1 µg/mL for the predicted most resistant lineage in both VRC01 dose groups pooled and the placebo group, stratified by IC₈₀ ≤ 1 µg/mL (green) vs. > 1 µg/mL (red).

end points having a nonmajority residue at the position (yielding 21 and 20 positions in the Americas and Africa trial, respectively). At each eligible position, we assessed whether PE differed by presence vs. absence of each residue including a gap observed in ≥ 6 primary end points (yielding 60 and 59 binary features in the Americas and Africa trial, respectively). The purpose of the analysis was to generate hypotheses about AA positions in the VRC01/CD4 binding region (Env binding to the Fab portion of VRC01) that may impact PE. We found no instance of the presence or absence of a specific residue that passed the significance bar for differential PE after multiple testing adjustment, with all Q-values >0.45 in the Americas trial and >0.64 in the Africa trial (*SI Appendix, Fig. S13*). Moreover, a phylogenetically corrected signature analysis using the LANL tool GenSig (12, 17, 18) (*SI Appendix, Supporting Information Text*) also did not find any significant associations of AA positions with randomized treatment assignment.

PE Decline in the Africa Trial with Increasing Weighted Hamming Distance from a VRC01-Sensitive Reference Sequence. Hamming distances, weighted by physicochemical (PC) properties of amino acids, are aggregate measures of deviation from VRC01-susceptible sequences and may exhibit greater power for detecting sieve effects than individual position-specific mutations. Hence, we calculated PC-weighted Hamming distances from the sequence identified in the CATNAP database against which the VRC01 IC₈₀ was lowest (serving as a “VRC01-sensitive reference sequence”), matching each isolated sequence by subtype and trial region (described in *Methods*), in two sets of AA positions: i) the 12 alignable positions and 5 PNGS position triplets predictive of neutralization resistance, and ii) the 50 positions constituting the union of the VRC01 binding footprint and the CD4 binding site (Table 2 and *SI Appendix, Figs. S6 and S7*). For the neutralization-associated distance in set (i), we found a point estimate trend of decreasing PE from 53 to -8% with the number of residue mismatches increasing from two to seven in the Africa trial (one-sided unadjusted $P = 0.13$, FWER $P = 0.13$, $Q = 0.13$); no decline in PE with an increasing distance was observed in the Americas trial (Fig. 5 A and B, respectively). For the VRC01/CD4 binding-associated distance in set (ii), we found significantly decreasing PE from

76% (95% CI, 21 to 93) to -9% (95% CI, -103 to 41) with the number of residue mismatches increasing from two to eight in the Africa trial (one-sided unadjusted $P = 0.02$, FWER $P = 0.033$, $Q = 0.027$); the declining point estimate trend in the Americas trial was not significant (Fig. 5 C and D, respectively). Hence, as measured by the number of accumulated VRC01/CD4 binding set residue mismatches from the VRC01-sensitive reference sequence corresponding to a point estimate of zero PE, eight residue mismatches abrogated PE in the Africa trial and 10 residue mismatches abrogated PE in the Americas trial.

Significant PE Decline with an Increasing VRC01 Epitope Distance from a VRC01-Sensitive Reference Sequence. As PC-weighted Hamming distances do not incorporate HIV-1 Env:VRC01 interface structural knowledge, we applied identical sieve analysis methods to weighted epitope distances that capture structural deviation of the VRC01 epitope when compared to the most VRC01-sensitive sequences using an approach adapted from ref. 14 (Table 2 and *SI Appendix, Figs. S6 and S7*). Specifically, the epitope distance was calculated as the minimum of distances from each of 10 sequences in the CATNAP database with the 10 lowest IC₅₀s of VRC01, with the set of reference sequences selected separately for subtype B (IC₅₀ 0.004 to 0.039 µg/mL), subtype C (IC₅₀ 0.01 to 0.05 µg/mL), and irrespective of subtype (IC₅₀ 0.004 to 0.02 µg/mL). Each epitope position was weighted by the strength of the interaction in the HIV-1 Env:VRC01 complex, i.e., structurally important sites carried extra weight. We previously showed that the strength of the Ab:epitope interaction was a feature that distinguished the broadest bnAbs (19). While bnAbs did not specifically target more conserved regions of HIV-1 Env, neutralization breadth depended on the conservation of key sites in the Ab:Env complex structure, and these key Env sites were defined by their stronger interaction with the bnAb. Here, we used a similar assessment of the structural VRC01:Env complex to assign weights to each epitope site [see *SI Appendix, Table S16* in the Statistical Analysis Plan (SAP), included in *SI Appendix, Supplementary Appendix*, for the set of position weights], and mutations were scored according to the BLOSUM62 matrix, with glycans given the highest score. For each trial and reference sequence, we found a substantial and highly significant decline in PE of both VRC01 dose groups pooled with an increasing epitope distance. For instance, in the

Table 2. Sets of AA positions and sequence features in HIV-1 Env used in the site-scanning and distance-based sieve analyses

Tier 1 Env features with highest variable importance for predicting TZM-bl VRC01 neutralization resistance

Alignable AA positions	60, 170, 230, 279, 280, 317, 365, 429, 456, 458, 459, and 471
PNGS motif AA position triplets	156 to 158, 229 to 231, 234 to 236, 616 to 618, and 824 to 826
Viral geometry features	Length of gp120, length of V1V2, length of V5, number of PNGS motifs in gp120, number of PNGS motifs in V1V2, number of PNGS motifs in V5, and number of cysteines in gp120
Union of VRC01 binding footprint (7) and CD4 binding site AA positions	97, 123 to 127, 196, 198, 276, 278 to 283, 365 to 371, 374, 425 to 432, 455 to 461, 463, 465 to 467, 469, and 471 to 477
VRC01 footprint AA positions* (14)	195 to 200, 274 to 285, 363 to 373, 425 to 433, and 453 to 476

*Used for computing epitope distances.
All AA positions are numbered using HXB2 coordinates.

Americas trial, PE dropped from 75% (95% CI, 39 to 90) against viruses with epitope distance 0.27 to -279% (95% CI, -1208 to -10) against viruses with epitope distance 1.7 from the subtype B reference sequence (unadjusted $P = 0.0026$, FWER $P = 0.0090$,

$Q = 0.010$) (Fig. 6A). Similarly, in the Africa trial, PE declined from 75% (95% CI, 38 to 90) against viruses with epitope distance 0.29 to -306% (95% CI, -1472 to -5) against viruses with epitope distance 2.0 from the subtype C reference sequence (unadjusted $P = 0.0030$,

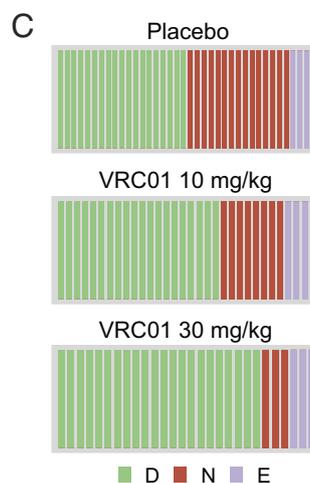
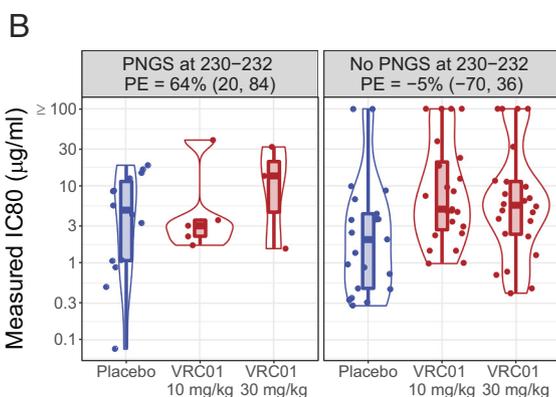
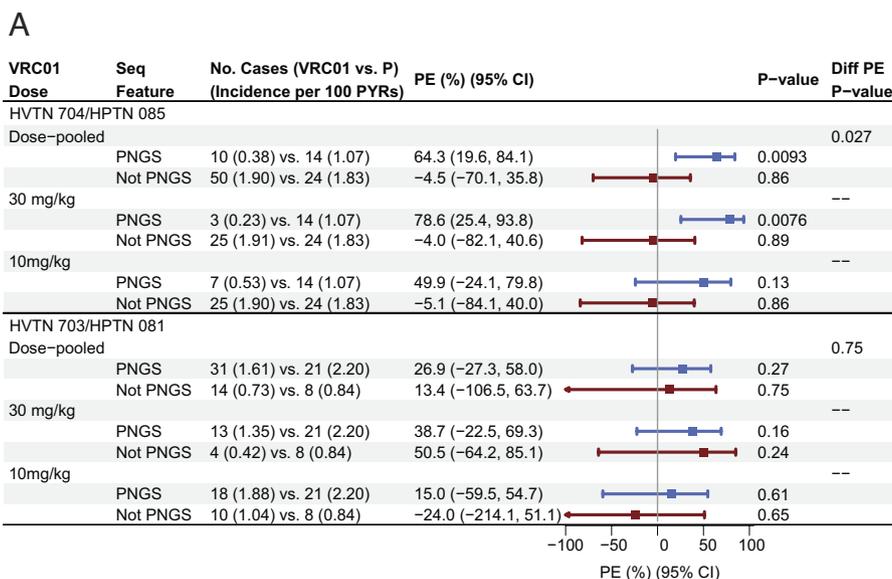


Fig. 4. VRC01 prevention efficacy associated with a N-linked glycosylation site motif. (A) Prevention efficacy by PNGS status at Env positions 230 to 232 of the predicted most resistant lineage in HVTN 704/HPTN 085 (Americas trial) (Top) or HVTN 703/HPTN 081 (Africa trial) (Lower). (B) Distribution of IC80 values of the most resistant synthesized variant among primary end points in the Americas trial, separated by PNGS status at positions 230 to 232 and treatment assignment. (C) Amino acid residue distributions among predicted most resistant lineages at amino acid position 230. Each vertical bar represents an analyzed primary end point.

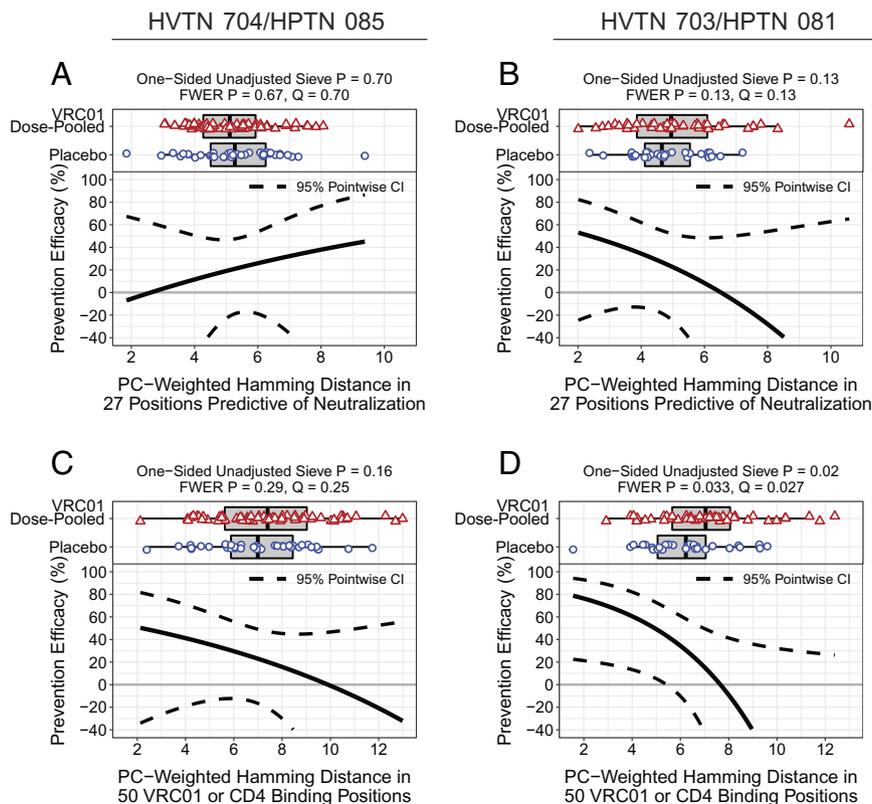


Fig. 5. Prevention efficacy (PE) by the physicochemical (PC)-weighted Hamming distance from the VRC01-sensitive reference sequence. The predicted most resistant lineage of observed sequences isolated from a primary end point case was matched to the reference sequence by subtype and trial region (see the text for details). Distance was calculated based on the 12 alignable positions and five PNGS position triplets predictive of neutralization resistance (A and B) or based on the 50 positions constituting the union of the VRC01 binding footprint and the CD4 binding site (C and D). The solid black curves show PE of both VRC01 dose groups pooled in HVTN 704/HPTN 085 (Americas trial) in panels (A) and (C) and in HVTN 703/HPTN 081 (Africa trial) in panels (B) and (D). The dashed black curves represent the 95% pointwise CIs. The horizontal box plots at the top of each panel show the distributions of the PC-weighted Hamming distance in the corresponding amino acid position set, for primary end points in both VRC01 dose groups pooled (open red triangles) or in the placebo group (open blue circles). “One-Sided Unadjusted Sieve P ” is a P -value from testing whether PE decreases with the distance on the X axis.

FWER $P = 0.0030$, $Q = 0.0030$) (Fig. 6B). Fig. 6C–F indicates that the epitope distance might also effectively discriminate PE in comparisons of individual 30 mg/kg and 10 mg/kg VRC01 dose regimens vs. placebo. *SI Appendix*, Fig. S14 shows Env positions 460 and 461 in both trials, as well as 459 in the Americas trial, as being most influential on values of the epitope distance across the HIV-1 diagnosis primary end points.

No Evidence of Differential PE against Single vs. Multiple Lineage HIV-1 Diagnosis Primary End Points. We found no evidence in either trial for a difference in PE against HIV-1 diagnosis with a single vs. multiple early virus lineages (*SI Appendix*, Fig. S15).

Greater Mean Epitope Distances in Participants Diagnosed with HIV-1 for the VRC01 vs. Placebo Group. While sieve analysis assesses how prospectively interpretable prevention efficacy in the entire study cohort depends on viral sequence features, it is also informative to restrict the analysis to individuals who were diagnosed with HIV-1 during the trial and compare Env sequence features between the VRC01 and placebo case groups. To this end, for each prespecified sequence feature, we estimated the mean feature value and tested for a mean difference (pooled VRC01 minus placebo) using a covariate-adjusted doubly robust method (20) (*SI Appendix*, Tables S2 and S3). For each trial, the VRC01 epitope distance exhibited the greatest and most statistically significant contrast between HIV-1 diagnosis primary end point cases in both VRC01 dose groups pooled vs. the placebo group. In the Americas trial, the mean epitope distance was 1.0 (95% CI, 0.91 to 1.09) among VRC01 cases and 0.76 (95% CI, 0.65 to 0.87) among placebo cases ($P = 0.0012$ for a difference).

In the Africa trial, the mean distance was 1.14 (95% CI, 0.93 to 1.35) among VRC01 cases and 0.84 (95% CI, 0.73 to 0.95) among placebo cases ($P = 0.012$ for a difference). The epitope distance was also the only studied feature for which there was a significant mean difference in the 10 mg/kg VRC01 dose vs. placebo comparison (*SI Appendix*, Table S4).

Epitope Distance Superior in Discriminating Treatment Assignment to VRC01 vs. Placebo. Given the low correlations among the amino acid sequence features that discriminated prevention efficacy as well as between each of these features and IC80 (Fig. 1 and *SI Appendix*, Fig. S16), we used three variable importance analysis approaches, restricting to HIV-1 diagnosis primary end points, to examine the features’ ability to discriminate the treatment assignment to VRC01 vs. placebo. First, using logistic models adjusted for pairs of features and geographic region in the Americas trial, both the epitope distance and predicted resistance probability carried treatment-predictive information beyond that captured by any single other sequence feature or IC80, whereas the PC-weighted Hamming distance carried no additional predictive information (*SI Appendix*, Fig. S17A). In contrast, in the Africa trial, the epitope distance and, to a lesser degree, the PC-weighted Hamming distance, predicted the treatment assignment while adjusting for any single other sequence feature or IC80, whereas the predicted resistance probability carried no additional predictive information (*SI Appendix*, Fig. S17B). In the model adjusted for the full set of sequence features and IC80, the epitope distance outperformed all other features in

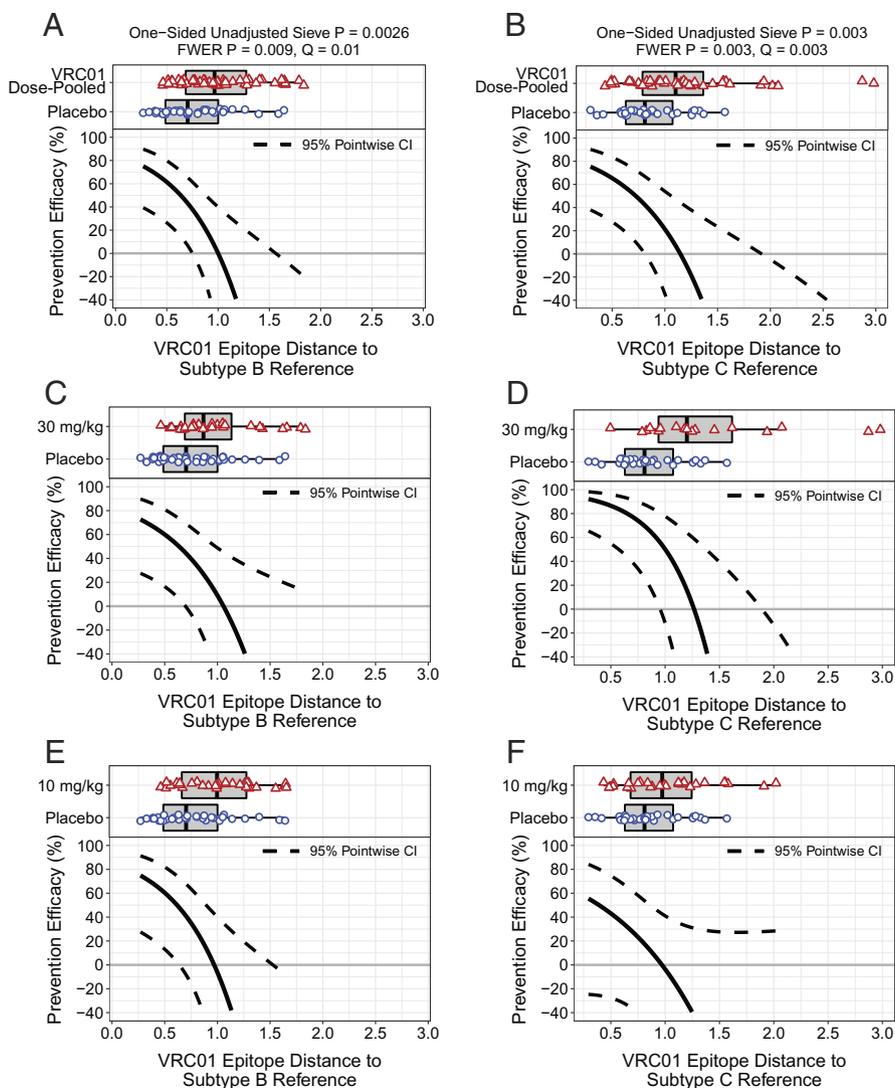


Fig. 6. Prevention efficacy (PE) by VRC01 epitope distance from the VRC01-sensitive subtype-specific reference sequence. The solid black curves show PE of both VRC01 dose groups pooled (A and B), the 30 mg/kg VRC01 group (C and D), and the 10 mg/kg VRC01 group in (E and F). The dashed black curves represent 95% pointwise CIs. Data from HVTN 704/HPTN 085 (Americas trial) are shown in panels (A), (C), and (E) and from HVTN 703/HPTN 081 (Africa trial) in panels (B), (D), and (F). The horizontal box plots at the top of each panel show the distributions of the epitope distance from the designated reference sequence, for primary end points in the designated VRC01 treatment group (open red triangles) or in the placebo group (open blue circles). “One-Sided Unadjusted Sieve P” is a *P*-value from testing whether PE decreases with the VRC01 epitope distance.

discriminating treatment assignment to VRC01 vs. placebo in both trials.

Given the performance of the epitope distance, we constructed additional distance measures by isolating individual components of the PC-weighted and epitope distances and fit logistic models adjusted for pairs of the distances as described above to assess the relative importance of the distinct components making up the epitope distance in predicting the treatment assignment. We found that the use of BLOSUM62 markedly outperformed physicochemical weighting in the Americas trial and the inclusion of VRC01:Env interaction weights with BLOSUM62 improved the prediction of treatment assignment in each trial (*SI Appendix, Fig. S18*). This supports the epitope distance as a superior marker overall in the AMP trials compared to other analyzed distance measures. An adjustment for another distance measure besides the epitope distance did not improve prediction of treatment assignment.

Next, using a more flexible penalized spline regression with cross-validation, we assessed in each trial the ability of any single Env

feature to predict treatment assignment to VRC01 while adjusting also for geographic region. The epitope distance and predicted IC80 were found to be predictive of treatment assignment in the Americas trial (*SI Appendix, Table S5*). In the Africa trial, no feature was found to be predictive of treatment assignment. There appeared to be no notable improvement in prediction accuracy comparing the model using all features with each of the models using a single feature (*SI Appendix, Table S6*). Additionally, in each trial, we assessed the strength of conditional associations between treatment assignment and any single feature given remaining features and geographic region. The epitope distance exhibited the strongest and statistically significant conditional association with treatment assignment (*SI Appendix, Table S7*).

Last, a more exhaustive multivariable machine-learning variable importance analysis (21) was conducted to assess how well different sets (*SI Appendix, Table S8*) of viral phenotypic and AA sequence features in HIV-1 diagnosis primary end points predicted treatment assignment to VRC01 vs. placebo beyond that

provided by baseline risk factors. We found no evidence of differential prediction performance based on cross-validated AUC after multiple testing adjustment, with all *Q*-values >0.3 in each trial (*SI Appendix, Figs. S19 and S20*).

Discussion

VRC01 prevention efficacy (PE) against HIV-1 diagnosis through 80 wk in the AMP trials varied with HIV-1 Env amino acid sequence features. The VRC01 epitope distance exhibited the greatest ability to discriminate PE in each trial, followed by the sequence-predicted probability of IC80 > 1 µg/mL in the Americas trial and the PC-weighted Hamming distance in the 50 binding-associated positions in the Africa trial. Furthermore, VRC01 epitope distance showed the greatest contrast between primary end point cases in both VRC01 dose groups pooled vs. the placebo group and the strongest ability to predict treatment assignment to VRC01 vs. placebo. Proteomic antibody resistance (PAR) scores of isolated viruses among placebo group cases were higher in the Africa trial, reflecting greater VRC01 neutralization resistance among circulating subtype C viruses. However, VRC01 epitope distance had a similar distribution among placebo group cases in the two trials, suggesting that it may be less subtype-dependent (*SI Appendix, Fig. S21 and Tables S2 and S3*).

The analysis generated a hypothesis that VRC01 confers greater prevention efficacy against diagnosis with subtype B viruses with a PNGS at Env positions 230 to 232. Among VRC01 recipients, no isolated viruses with IC80 ≤ 1 µg/mL and a PNGS at 230 to 232 broke through compared to three such viruses among placebo recipients in the Americas trial. The presence of a PNGS at 230 to 232 was not associated with a sieve effect in the Africa trial. Bricault et al. (12) found that a PNGS at 230 to 232 was associated with changes in VRC01 neutralization across multiple subtypes but not in a clade C only virus panel.

A limitation of the study was small treatment-pooled HIV-1 diagnosis end point counts for one of two levels of many binary sequence features preidentified as predictive of VRC01 neutralization resistance, which limited precision in estimation of feature-specific PE and implied low power to detect differential PE across the two feature levels. An additional limitation was that most statistical approaches evaluated a single Env sequence per HIV-1 diagnosis primary end point selected from the set of multiple measured sequences. There was also imperfect concordance between the sequence synthesized for measuring IC80 in the TZM-bl neutralization assay and the selected sequence used for predicting IC80 via SuperLearner modeling (*SI Appendix, Fig. S22*) due to differences in sequence availability at the time of conducting each analysis, which may partly contribute to the low correlations between IC80 and PAR scores.

A challenge posed to sieve analysis is the formidable difficulties in discriminating between acquisition (sterilizing immunity) and postacquisition VRC01 sieve effects, given that VRC01 could impact viral evolution between the time of initial viral acquisition and detection, which is unobservable. This problem was less severe in AMP with its monthly HIV testing schedule compared to previous vaccine efficacy trials that used 3- or 6-monthly HIV testing schedules. However, the fact that VRC01-sensitive isolated viruses (IC80 ≤ 1 µg/mL) had lower viral load in the VRC01 vs. placebo arms at the first RNA-positive time point (6) suggests that VRC01 could have converted some RNA-positive acquisition events to subviremic infections. If this occurred, such infections could conceivably be kept subviremic for a prolonged period of time by the repeated VRC01 infusions during which the virus could evolve. In this regard,

epitope distances in VRC01 cases that were greater than the maximal epitope distance in placebo cases (Fig. 6) may reflect VRC01-driven postacquisition evolution (whether the infection remained subclinical or not). An alternative explanation for the absence of placebo cases with epitope distances as large as observed in VRC01 cases is that if viruses with smaller epitope distances were more likely to establish an infection in placebo cases due to a fitness advantage, then a blockage of acquisition of such viruses by VRC01 may have created an opportunity for viruses with larger epitope distances and reduced fitness to establish breakthrough infections. Irrespective of whether the VRC01 effect can be explained as a result of blockage of acquisition of viruses with smaller epitope distances and/or postacquisition selection of viruses with larger epitope distances, after discarding such outlying VRC01 cases, the unadjusted one-sided *P*-value for a declining PE changed from 0.0026 to 0.016 in the Americas trial and from 0.003 to 0.10 in the Africa trial indicating a contribution of this effect to the former sieve effect observation (*SI Appendix, Fig. S23*).

Anticipating the challenge of disentangling sterilizing immunity vs. postacquisition effects of VRC01, the AMP studies were designed to follow participants 32 wk after the last infusion (week 104), a time point selected such that VRC01 serum concentration by that time would be zero or near zero for all participants. This design element implies that RNA-detectable HIV acquisition events converted to subviremic HIV acquisition events by VRC01 would become detectable by an HIV test at week 104. If VRC01 provided no sterilizing immunity whatsoever (i.e., no blockage of initial infections and hence no acquisition sieve effects), then overall cumulative prevention efficacy against RNA-detectable HIV-1 (including all strains) to week 104 [PE(104)] would be 0%. Moreover, any sieve effect for a given viral feature would “teeter-totter around PE = 0%,” that is, PE(104) against RNA-detectable HIV-1s with small PAR scores/distances would be positive and PE(104) against large PAR scores/distances would be negative, where such an effect would presumably be caused by postacquisition viral escape during subviremic infection to an increased level of VRC01 neutralization resistance. On the other hand, if VRC01 provided some sterilizing immunity, then PE(104) would exceed 0%, albeit it would be expected to be closer to zero than PE(80) due to waning prevention efficacy and to heterogeneity in exposure and risk (22); moreover, PE(104) against viruses with small PAR scores/distances would exceed 0% more than expected under a “teeter-totter around PE = 0%” result. Therefore, estimation of PE(104) overall and by PAR score/distance provides information to help discriminate the likelihood of acquisition vs. postacquisition sieve effects. Overall PE(104) was near zero (6% [95% CI, -35 to 35] in the Americas trial and 4% [95% CI, -46 to 38] in the Africa trial) (*SI Appendix, Fig. S24*), supporting postacquisition sieve effects. On the other hand, PE(104) against RNA-detectable IC80 < 1 µg/mL viruses through to week 104 for the trials pooled (*SI Appendix, Fig. S25*) was 66% (95% CI, 32 to 83), suggesting that some sieve effects could be sterilizing immunity effects. Sampling variability [e.g., considering 95% CIs around PE(104)] limits the ability to draw definitive conclusions about the relative correctness of different interpretations of results as VRC01 acquisition vs. postacquisition sieve effects. Future studies that compare viral evolution at longitudinal time points postacquisition between the VRC01 and placebo arms would shed further light on this question given that these analyses can address postacquisition sieve effects in a more isolated fashion.

The potent sieve effects captured by the VRC01 epitope distance were equally strong as the sieve effects captured by the neutralization resistance biomarker IC80. The low correlation of

epitope distance with IC80 and logistic models in *SI Appendix, Fig. S17* support that the epitope distance biomarker carries additional information about prevention efficacy beyond the TZM-bl assay IC80 biomarker. Consequently, while IC80 as a direct measure of neutralization constitutes a primary criterion for comparing the potential prevention efficacy of different bnAbs and bnAb combinations, epitope distance as well as PAR scores may provide useful supplemental criteria. For example, when planning a potential efficacy trial, bnAbs with smaller epitope distances of circulating strains in the future trial region from strains that are highly sensitive to the bnAb may be ranked higher. The epitope distance and PAR score biomarkers may be especially useful when the available database of HIV-1 Env sequences is much larger than the available database of bnAb IC80s against circulating strains, a typical situation given the greater resources that are required to generate IC80 data.

Methods

Ethics Statement. All participants in the AMP trials provided written informed consent. A sample informed consent form is provided in the Appendix of ref. 6. New consent was obtained for each version of the protocol. For the AMP trials, central and site-specific institutional review boards and ethics committees reviewed and approved the initial protocol and each subsequent version. In addition, the HIV-1 sequencing work was approved by the University of Cape Town Human Research Ethics Committee (University of Cape Town) through HREF ref. no. 176/2017. The TZM-bl target cell neutralization assay work was approved by the Duke University Health System Institutional Review Board (Duke University) through protocol no. Pro00093087 and by the University of the Witwatersrand Human Research Ethics Committee (National Institute for Communicable Diseases) through protocol no. M201105.

Analysis Cohort and HIV-1 End point. We conducted the sieve analysis in each AMP trial separately and combined, in the modified intention-to-treat (MITT) cohort who were HIV-negative at enrollment and received the first infusion. Given the marked differences in results pertaining to the Americas vs. Africa trial, the trial-pooled results did not yield any additional insights beyond those from the trial-specific results, and therefore only the results of the trial-specific analyses are presented.

For primary end points (i.e., HIV-1 infection diagnoses per the trials' testing algorithm by the week 80 visit), the event of interest was the collection of the first HIV-1 RNA-positive sample by either a standard HIV-1 RNA PCR assay (Abbott m2000 RealTime in the Americas trial and Roche COBAS TaqMan in the Africa trial) or a low-copy assay [iSCA v2.0 (23) in both trials]. Given the trials' dynamic visit windows determined according to the prior infusion visit date, the time from enrollment to the event of interest was right-censored at either the time to the last HIV-negative sample collection or 595 d (85 wk) since enrollment, whichever occurred earlier.

Participants who met the HIV-1 primary end point but had missing sequence data were excluded (none in the Americas trial and 2 in the 30 mg/kg dose group in the Africa trial). In total, the analysis cohort comprised 2,687 and 1,922 MITT participants in the Americas and Africa trial, respectively.

HIV-1 Sequencing and Lineage Identification. In each trial, nucleotide sequences of the *rev-env-nef* and *gag-pol* genomic regions were measured in the first HIV-1 RNA-positive plasma sample from all HIV-1 diagnosis primary end point cases using the unique molecular identifier-tagged PacBio sequencing methodology (24). If the first positive sample contributed <20 viral RNA templates, the second positive sample was included for sequencing, and the sequences obtained from these two samples were combined for use in the sieve analysis. This occurred for five primary end points in the Americas trial but did not occur in any in the Africa trial. The raw sequences were processed using a standardized error filtration pipeline (24). Subsequently, nucleotide and amino acid (AA) alignments were approved by three independent reviews, where sequences with evidence of postinfection recombination or hypermutation were excluded from the analysis. The approved alignments were next used by an expert panel for determining the number of virus lineages in each

sample and assigning each measured sequence to an identified lineage. A viral lineage refers to a phylogenetic cluster of closely related sequences that emerged over the early sampling period.

In this analysis, we used aligned Env AA sequences only and observed an average of 196 and 155 Env sequences per primary end point in the Americas and Africa trial, respectively (Table 1). Hypervariable AA positions in the V1, V2, V4, and V5 regions of Env were deemed unalignable and excluded from the analysis.

Representative Sequence Selection for Analysis. Sieve analysis methods presented herein analyzed a single AA sequence selected for analysis from each primary end point case. Given the potential multiplicity of acquired virus lineages and a lack of knowledge guiding the relative importance of different lineages, we adopted a sensitivity analysis strategy where the same suite of analyses was conducted for three distinct definitions of a representative isolated sequence from very early in infection. To that end, we first identified a "mindist" sequence for each lineage defined as an individual's closest observed sequence to their lineage-specific consensus sequence. Second, among the possibly multiple lineage-representing mindist sequences in a sample, we selected three types of representative sequences: a) the mindist sequence corresponding to the lineage with the largest number of sequences (termed the most frequent lineage), b) the mindist sequence with the minimal predicted IC80 (termed the predicted most sensitive lineage), and c) the mindist sequence with the maximal predicted IC80 (termed the predicted most resistant lineage), with predicted IC80 values being PAR scores from the SuperLearner model for a quantitative outcome. For consistency with Corey et al. (6) and because blocking HIV-1 diagnosis requires blocking the most resistant exposing strain, all presented analyses used lineage type (c). Given the very high correlations among the three lineage types for each analyzed feature, results for lineage types (a) and (b) are similar and not presented but available upon request.

HIV-1 Env AA Sequence Features. All Env sequence features were prespecified before treatment unblinding and divided into two analysis tiers. Preidentified features that had previously been shown to predict *in vitro* VRC01 neutralization resistance (based on the TZM-bl target cell assay) were analyzed in the hypothesis-driven tier 1. A broader feature set, agnostic to neutralization data, was prespecified for a comprehensive exploratory tier 2 analysis, allowing hypothesis generation about signatures attributable to VRC01-mediated effector functions other than neutralization as captured by the TZM-bl assay. Tier 1 comprised three PAR scores and a set of 24 Env features with the highest variable importance in predicting VRC01 neutralization resistance (see *SI Appendix, section 4.6* of the SAP for details). Two of the PAR scores corresponded to SuperLearner prediction models for a binary (≤ 1 vs. > 1 $\mu\text{g}/\text{mL}$) and quantitative IC80 outcome. The third PAR score was obtained by categorizing predicted quantitative IC80s as ≤ 1 vs. 1 to ≤ 3 vs. > 3 $\mu\text{g}/\text{mL}$. The main manuscript reports results for the PAR score generated by the binary-outcome model, on the logit of the predicted probability scale, with results for the predicted quantitative IC80 and categorical IC80 PAR scores presented in supplement. Tier 2 comprised a comprehensive set of binary residue presence vs. absence features for all alignable Env positions with sufficient residue variability to potentially infer differential PE, two measures of virus lineage multiplicity, PC-weighted Hamming distances, and VRC01 epitope BLOSUM62 distances with positions weighted by the strength of the Env:VRC01 interaction. PC-weighted Hamming distances were calculated to the most VRC01 neutralization-sensitive sequences available in the CATNAP database shortly before the analysis was carried out (July 2022). Epitope distances were defined as the minimum of distances calculated to 10 CATNAP sequences with the lowest VRC01 IC50s, with the same three sets of 10 reference sequences selected for all primary end point cases: a set of subtype B sequences with the lowest IC50s, a set of subtype C sequences with the lowest IC50s, and a set of sequences of any subtype with the lowest IC50s. Detailed feature descriptions are provided in the SAP.

PE by an HIV-1 Env AA Sequence Feature. We defined "feature-specific PE" as 1 minus the hazard ratio (dose-pooled or dose-specific VRC01/placebo) of the primary end point exhibiting a particular sequence feature by the selected lineage type. A sieve effect refers to statistically significant evidence for variation in feature-specific

PE across different levels of the analyzed feature. For binary features such as the presence/absence of a residue at a given AA position, we estimated feature-specific PE by a competing risks Cox model, with differential PE assessed using a Wald test (25). For quantitative features such as the predicted IC80 PAR score, length of gp120, or a PC-weighted Hamming distance, feature-specific PE was estimated using a continuous mark-specific hazard-ratio model (26), with a Wald test employed for assessing variation in PE. Bivariate features with a quantitative and binary component were analyzed using the same methods as univariate quantitative features, assuming a constant difference in the log quantitative feature-specific hazard ratio for the two levels of the binary component. The competing risks Cox model, the duplication method Cox model underlying (25), and the Cox model portion of the product estimator of ref. 26 were stratified by the same four-level geographic region covariate as described above. One-sided testing for varying PE was performed for quantitative PAR scores, and the Hamming and epitope distances, under the rationale that only one direction of the sieve effect is scientifically relevant. Two-sided testing for feature-varying PE was performed for binary features and all other quantitative features such as loop lengths or numbers of PNGSs because both directions of the sieve effect are scientifically relevant.

Infection-Conditional Analysis Methods. Besides analyses of feature-specific PE in the MITT cohort, we conducted analyses of sequence features restricting to primary end point cases. To estimate and compare marginal mean values of sequence features between VRC01 and placebo group end points, we used the doubly robust targeted minimum loss-based (DRTMLE) estimation (20), with 95% Wald CIs and a two-sided Wald test of zero mean difference. The DRTMLE method used adjustment for the geographic region covariate described above since region was judged to possibly be associated with both HIV-1 diagnosis and the sequence features under analysis. This covariate adjustment was needed in order that the estimated contrast in means across treatment groups assesses a causal effect of VRC01 on the analyzed feature in the subpopulation of participants who would acquire HIV-1 under either treatment assignment (VRC01 or placebo) (27). More aggressive covariate adjustment was not done because adjusting for covariates predictive of HIV-1 diagnosis but not the feature value would reduce precision.

Variable importance analyses were used to assess the ability of sequence features and IC80 to predict treatment assignment to VRC01 vs. placebo using three statistical approaches. First, we used logistic regression models adjusted for pairs of features and the four-level geographic region covariate as described above, with 95% Wald CIs for the odds ratio of assignment to VRC01. Next, we assessed treatment prediction accuracy of individual features by modeling the probability of VRC01 assignment using penalized spline regression, with prediction accuracy measured using cross-validated area under the receiver operating curve (AUC). We then fit a logistic generalized additive model adjusted for all features including geographic region and measured prediction accuracy using cross-validated AUC.

We additionally estimated the strength of conditional association between VRC01 assignment and any single feature given remaining features and geographic region. We measured the strength of association using the scaled expected conditional covariance (28), which can be expressed as the Pearson correlation between the following two quantities: i) the difference between observed VRC01 assignment and the conditional probability of VRC01 assignment given all features except for the feature of interest and ii) the difference between the feature of interest and its conditional mean given all remaining features. We estimated each of the conditional means using generalized additive models.

Finally, a machine-learning analysis was conducted using the SuperLearner ensemble method for predicting the treatment assignment to VRC01 vs. placebo based on predefined sets of viral features (*SI Appendix, Table S8*); this analysis was then repeated for predicting assignment to 30 mg/kg VRC01 vs. placebo and again for predicting assignment to 10 mg/kg VRC01 vs. placebo. As part of this analysis, differential prediction of treatment assignment based on viral features vs. baseline risk factors alone was assessed (21).

Multiplicity Adjustment. Adjustments for multiplicity of hypothesis tests for differential PE (sieve effect tests) were applied separately to each individual or combined trials and the three selected lineage types within each trial due to

their high correlations (*SI Appendix, Figs. S26 and S27*). No multiple testing correction was applied to PAR scores given their Tier 1 status as primary features of interest. The following four separate multiple comparison procedures were specified: i) the 18 and 19 eligible Tier 1 features in the Americas and Africa trial, respectively, with the highest variable importance for predicting VRC01 neutralization resistance, ii) the 60 and 59 Tier 2 Env-scanning eligible residue match/mismatch features in the Americas and Africa trial, respectively, iii) the five Tier 2 PC-weighted Hamming distances and VRC01 epitope distances, and iv) the two different measures of single vs. multiple lineage infections. For each multiplicity set, a permutation-based method (29, 30), implemented in ref. 31, was used to control the family-wise error rate (FWER) and the false discovery rate (FDR) (32) (Q-values). FWER statistical significance was defined as an FWER-adjusted P -value ≤ 0.05 , and FDR statistical significance as a Q-value ≤ 0.2 and an unadjusted P -value ≤ 0.05 .

Data, Materials, and Software Availability. The analyzed Env AA mindist sequence alignment with the clinical and sequence feature data is available at <https://atlas.scharp.org/cpas/project/HVTN%20Public%20Data/HVTN%20704%20HPTN%20085%20and%20HVTN%20703%20HPTN%20081%20AMP/begin.view> (33). All computer code used for the statistical analysis, including code used for deriving sequence features from the AA mindist sequence alignment, is available at <https://github.com/mjuraska/AMPSieve-public> (34).

ACKNOWLEDGMENTS. We thank all study participants and clinical staff involved in the trials. This work was supported by the National Institute of Allergy and Infectious Diseases of the NIH, under awards R37AI054165 (to P.B.G.) and UM1AI068635 (to P.B.G.) and by the Office of Research Infrastructure Programs, NIH (award S10OD028685 to Fred Hutchinson Cancer Center). H.B. and M.R. are supported by a cooperative agreement between The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the US Department of the Army [W81XWH-18-2-0040]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, the US Army, the Department of Defense, or the Department of Health and Human Services.

Author affiliations: ^aVaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA 98109; ^bU.S. Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, MD 20910; ^cHenry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD 20817; ^dBiostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101; ^eScience and Technology Policy Fellowships, American Association for the Advancement of Science, Washington, DC 20005; ^fDepartment of Computer Science, Pitzer College, Claremont, CA 91711; ^gDepartment of Microbiology, University of Washington School of Medicine, Seattle, WA 98195; ^hTheoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87544; ⁱDepartment of Microbiology, Tumor, and Cell Biology, Karolinska Institutet, Solna 171 77, Sweden; ^jInstitute of Infectious Disease and Molecular Medicine, and Wellcome Centre for Infectious Diseases Research in Africa, Department of Pathology, Faculty of Health Sciences, University of Cape Town and National Health Laboratory Service, Cape Town 7701, South Africa; ^kHIV Virology Section, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg 2192, South Africa; ^lAntibody Immunity Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2000, South Africa; ^mCentre for the AIDS Programme of Research in South Africa, University of KwaZulu-Natal, Durban 4041, South Africa; ⁿDepartment of Surgery, Duke University Medical Center, Durham, NC 27710; ^oInstitute of Global Health and Infectious Diseases, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; ^pDepartment of Medicine, University of Washington, Seattle, WA 98195; ^qDepartment of Laboratory Medicine, University of Washington, Seattle, WA 98195; ^rPublic Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; ^sDepartment of Biostatistics, University of Washington, Seattle, WA 98195; ^tDepartment of Global Health, University of Washington, Seattle, WA 98195; and ^uDepartment of Microbiology, University of Washington, Seattle, WA 98109

Author contributions: M.J., H.B., C.A.M., C.M., S.T.K., L.M., D.C.M., J.A.H., M.S.C., L. Corey, M.R., P.B.G., C.W., and J.I.M. designed research; M.J., H.B., D.H.W., W.D., L. Chen, H.Z., T.B., A.P., B.M., A.Y., D.M., T.Y., N.B., A.G.-N., N.N., and R.T. performed research; M.J., H.B., A.C.d., C.A.M., L.L., K.G., E.E.G., J.L., C.M., A.H., B.D.W., N.E., B.S., E.R., D.S., R.R., P.T.E., M.R., and P.B.G. analyzed data; M.J., H.B., A.C.d., C.A.M., L.L., J.L., N.E., R.R., and P.T.E. wrote the statistical analysis plan, curated data, and reviewed the manuscript and approved the final version for submission; K.G., E.E.G., C.M., A.H., B.D.W., B.S., E.R., and D.S. curated data and reviewed the manuscript and approved the final version for submission; L.N.C., D.H.W., W.D., L. Chen, H.Z., T.B., A.P., B.M., A.Y., D.M., T.Y., N.B., A.G.-N., N.N., and R.T. reviewed the manuscript and approved the final version for submission; S.T.K., L.M., D.C.M., J.A.H., M.S.C., L. Corey, M.R., C.W., and J.I.M. performed supervision and reviewed the manuscript and approved the final version for submission; P.B.G. wrote the statistical analysis plan, curated data, performed supervision, provided oversight of the statistical analyses, and reviewed the manuscript and approved the final version for submission; and M.J., L.N.C., and P.B.G. wrote the paper.

1. UNAIDS, Global HIV & AIDS statistics – Fact Sheet 2022. <https://www.unaids.org/en/resources/fact-sheet>. Accessed 15 December 2022.
2. S. Mahomed, N. Garrett, C. Baxter, O. A. Karim, S. S. A. Karim, Clinical trials of broadly neutralizing monoclonal antibodies for human immunodeficiency virus prevention. *J. Infect Dis.* **223**, 370–380 (2021).
3. B. Julg, D. Barouch, Broadly neutralizing antibodies for HIV-1 prevention and therapy. *Semin. Immunol.* **51**, 101475 (2021). [10.1016/j.smim.2021.101475](https://doi.org/10.1016/j.smim.2021.101475).
4. S. T. Karuna, L. Corey, Broadly neutralizing antibodies for HIV prevention. *Annu. Rev. Med.* **71**, 329–346 (2020).
5. S. R. Walsh, M. S. Seaman, Broadly neutralizing antibodies for HIV-1 prevention. *Front. Immunol.* **12**, 712122 (2021).
6. L. Corey *et al.*, Two randomized trials of neutralizing antibodies to prevent HIV-1 acquisition. *N. Engl. J. Med.* **384**, 1003–1014 (2021).
7. T. Zhou *et al.*, Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811–817 (2010).
8. M. Sarzotti-Kelsoe *et al.*, Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J. Immunol. Methods* **409**, 131–146 (2014).
9. D. C. Montefiori, Measuring HIV neutralization in a luciferase reporter gene assay. *Methods Mol. Biol.* **485**, 395–405 (2009).
10. P. Gilbert, S. Self, M. Rao, A. Naficy, J. Clemens, Sieve analysis: Methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. *J. Clin. Epidemiol.* **54**, 68–85 (2001).
11. M. Juraska *et al.*, Viral genetic diversity and protective efficacy of a tetravalent dengue vaccine in two phase 3 trials. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8378–E8387 (2018).
12. C. A. Bricault *et al.*, HIV-1 neutralizing antibody signatures and application to epitope-targeted vaccine design. *Cell Host Microbe* **25**, 59–72.e58 (2019).
13. C. A. Magaret *et al.*, Prediction of VRC01 neutralization sensitivity by HIV-1 gp160 sequence features. *PLoS Comput. Biol.* **15**, e1006952 (2019).
14. E. M. Cale *et al.*, Neutralizing antibody VRC01 failed to select for HIV-1 mutations upon viral rebound. *J. Clin. Invest.* **130**, 3299–3304 (2020).
15. H. Yoon *et al.*, CATNAP: A tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res* **43**, W213–W219 (2015).
16. B. D. Williamson *et al.*, Super LeArner Prediction of NAb Panels (SLAPNAP): A containerized tool for predicting combination monoclonal broadly neutralizing antibody sensitivity. *Bioinformatics* **37**, 4187–4192 (2021).
17. T. Bhattacharya *et al.*, Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* **315**, 1583–1586 (2007).
18. Los Alamos National Laboratory (LANL), GenSig explanation. <https://www.hiv.lanl.gov/content/sequence/GENETICSIGNATURES/help.html>. Accessed 16 January 2023.
19. H. Bai, Y. Li, N. L. Michael, M. L. Robb, M. Rolland, The breadth of HIV-1 neutralizing antibodies depends on the conservation of key sites in their epitopes. *PLoS Comput. Biol.* **15**, e1007056 (2019).
20. D. Benkeser, M. Carone, M. J. V. Laan, P. B. Gilbert, Doubly robust nonparametric inference on the average treatment effect. *Biometrika* **104**, 863–880 (2017).
21. B. Williamson, P. B. Gilbert, N. Simon, M. Carone, A general framework for inference on algorithm-agnostic variable importance. *J. Am. Stat. Assoc.* **118**, 1645–1658 (2022). [10.1080/01621459.2021.2003200](https://doi.org/10.1080/01621459.2021.2003200).
22. J. J. O'Hagan, M. A. Hernan, R. P. Walensky, M. Lipsitch, Apparent declining efficacy in randomized trials: Examples of the Thai RV144 HIV vaccine and South African CAPRISA 004 microbicide trials. *AIDS* **26**, 123–126 (2012).
23. M. A. Tosiano, J. L. Jacobs, K. A. Shutt, J. C. Cyktor, J. W. Mellors, A simpler and more sensitive single-copy HIV-1 RNA assay for quantification of persistent HIV-1 viremia in individuals on suppressive antiretroviral therapy. *J. Clin. Microbiol.* **57**, e01714-18 (2019).
24. D. H. Westfall *et al.*, Optimized SMRT-UMI protocol produces highly accurate sequence datasets from diverse populations – application to HIV-1 quasispecies. *bioRxiv* [Preprint] Posted 24 Feb, 2023 (2023). <https://doi.org/10.1101/2023.02.23.529831> (Accessed 5 March 2023).
25. M. Lunn, D. McNeil, Applying Cox regression to competing risks. *Biometrics* **51**, 524–532 (1995).
26. M. Juraska, P. B. Gilbert, Mark-specific hazard ratio model with multivariate continuous marks: An application to vaccine efficacy. *Biometrics* **69**, 328–337 (2013).
27. B. E. Shepherd, P. B. Gilbert, Y. Jemai, A. Rotnitzky, Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics* **62**, 332–342 (2006).
28. Y. Xiang, N. Simon, A flexible framework for nonparametric graphical modeling that accommodates machine learning. *Proc. Mach. Learn. Res.* **119**, 10442–10451 (2020).
29. P. H. Westfall, S. S. Young, p value adjustments for multiple tests in multivariate binomial models. *J. Am. Stat. Assoc.* **84**, 780–786 (1989).
30. P. H. Westfall, S. S. Young, *Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment* (John Wiley & Sons, 1993), **vol. 279**.
31. Y. Fong, K. Sebeten, *kyotil*: Utility functions for statistical analysis report generation and monte carlo studies (2023), <https://CRAN.R-project.org/package=kyotil>. Accessed 18 August 2022.
32. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B (Methodological)* **57**, 289–300 (1995).
33. C. A. Magaret, M. Juraska, J. Ludwig, C. Molitor, H. Bai, HVTN 704/HPTN 085 and HVTN 703/HPTN 081 (AMP): Juraska *et al.* PNAS 2024. Atlas. <https://atlas.scharp.org/cpas/project/HVTN%20Public%20Data/HVTN%20704%20HPTN%20085%20and%20HVTN%20703%20HPTN%20081%20AMP/begin.view>. Deposited 5 August 2022.
34. M. Juraska *et al.*, R code used for amino acid sequence sieve analysis of data from the Antibody Mediated Prevention (AMP) trials. GitHub. <https://github.com/mjuraska/AMPSieve-public>. Deposited 31 May 2023.