bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TITLE

Functional Improbable Antibody Mutations Critical for HIV Broadly Neutralizing Antibody Development

AUTHORS

Kevin Wiehe[1*], Todd Bradley[1], R. Ryan Meyerhoff[1], Connor Hart[1], Wilton Williams[1], David Easterhoff[1], William J. Faison[1], Thomas B. Kepler[2], Kevin Saunders[1], S. Munir Alam[1], Mattia Bonsignori[1], and Barton F. Haynes[1*]

AFFILIATIONS

[1]Duke Human Vaccine Institute, Duke University, Durham NC 27710

[2]Department of Microbiology and Immunology, Boston University, Boston, MA

*Corresponding Authors: Email: kevin.wiehe@duke.edu (K.W.)(Lead Contact); barton.haynes@duke.edu (B.F.H.)

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUMMARY

HIV-1 broadly neutralizing antibodies (bnAbs) require high levels of activation-induced cytidine deaminase (AID) catalyzed somatic mutations for optimal neutralization potency. Probable mutations occur at sites of frequent AID activity, while improbable mutations occur where AID activity is infrequent. One bottleneck for induction of bnAbs is the evolution of viral envelopes (Envs) that can select bnAb B cell receptors (BCR) with improbable mutations. Here we define the probability of bnAb mutations and demonstrate the functional significance of key improbable mutations in three bnAb B cell lineages. We show that bnAbs are enriched for improbable mutations, implying their elicitation will be critical for successful vaccine induction of potent bnAb B cell lineages. We outline a mutation-guided vaccine strategy for identification of Envs that can select B cells with BCRs with key improbable mutations required for bnAb development. Our analysis suggests that through generations of viral escape, Env trimers evolved to hide in low probability regions of antibody sequence space.

INTRODUCTION

The goal of HIV-1 vaccine development is the reproducible elicitation of potent, broadly neutralizing antibodies (bnAbs) that can protect against infection of transmitted/founder (TF) viruses (Haynes and Burton, 2017). While ~50% of HIV-infected individuals generate bnAbs (Hraber et al., 2014), bnAbs in this setting only arise after years of infection (Bonsignori et al., 2016; Doria-Rose et al., 2014; Liao et al., 2013). BnAbs isolated from infected individuals have one or more unusual traits, including long third complementarity determining regions (CDR3s)(Yu and Guan, 2014), autoreactivity (Kelsoe and Haynes, 2017), large insertions and

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

deletions (Kepler et al., 2014), and high frequencies of somatic mutations (Burton and Hangartner, 2016). Somatic hypermutation of the B cell receptor (BCR) heavy and light chain genes is the primary diversification method during antibody affinity maturation - the evolutionary process that drives antibody development after initial BCR rearrangement and leads to high affinity antigen recognition (Teng and Papavasiliou, 2007). Not all somatic mutations acquired during antibody maturation are necessary for bnAb development; rather high mutational levels may reflect the length of time required to elicit bnAbs (Georgiev et al., 2014; Jardine et al., 2016b). Therefore, shorter maturation pathways to neutralization breadth involving a critical subset of mutations is desirable because antibody mutation levels induced by vaccines seldom reach the mutation frequencies observed in bnAbs (Easterhoff et al., 2017; Moody et al., 2011). Importantly, within this subset of critical mutations some mutations may be probable and easy to elicit, while other mutations may be improbable and very challenging to induce due to biases in how mutations arise during affinity maturation.

Somatic hypermutation occurs prior to antigen affinity-based selection during affinity maturation (De Silva and Klein, 2015; Victora and Nussenzweig, 2012). The principal enzyme that mediates somatic hypermutation is activation-induced cytidine deaminase (AID) (Di Noia and Neuberger, 2007). AID preferentially targets specific nucleotide sequence motifs (referred to as "AID hot spots") while targeting of other nucleotide motifs is disfavored (referred to as "AID cold spots") (Betz et al., 1993; Pham et al., 2003; Yaari et al., 2013). AID initiates DNA lesions and their subsequent repair results in a bias for which bases are substituted at the targeted position (Cowell and Kepler, 2000). The consequence of this non-uniformly random mutation process is that specific amino acid substitutions occur with varying frequencies prior to antigenic selection. Mutations at AID hot spots can occur frequently in the absence of antigen selection

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

due to immune activation-associated AID activity (Bonsignori et al., 2016; Yeap et al., 2015). Amino acid substitutions that occur infrequently generally require strong antigenic selection to arise during maturation (Brown et al., 1992; Kocks and Rajewsky, 1988). Such rare amino acid substitutions are improbable prior to selection for two reasons: 1) base mutations must occur at AID cold spots and 2) due to codon mapping, multiple base substitutions must occur for a specific amino acid change to take place. Within the critical subset of mutations that grant broad neutralization capacity to a bnAb lineage, those key mutations that are also improbable prior to selection may represent important events in bnAb maturation and are thus compelling targets for selection in a vaccine setting. We recently described a rare mutation in an early ancestor of a V3-glycan bnAb that allowed broad neutralization of transmitted/founder viruses demonstrating in one bnAb lineage that functionally important, improbable mutations can be roadblocks in HIV-1 bnAb development (Bonsignori et al., 2017). However, what has remained unclear is whether and to what extent such roadblocks are a general problem for bnAb elicitation. Here we describe the identification of improbable mutations in three bnAb B cell lineages, determine the functional relevance of these mutations for development of bnAb neutralization potency, and outline a vaccine design strategy for choosing sequential Envs capable of selecting B cells with BCRs with functionally important improbable mutations.

RESULTS

Identification of Functional Improbable Antibody Mutations

To determine the role of rare mutational events in bnAb development, we developed a computational program to identify improbable antibody mutations. Our program, Antigen Receptor Mutation Analyzer for Detection of Low Likelihood Occurrences (ARMADiLLO) simulates the somatic hypermutation process using a model of AID targeting and base

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

substitution via DNA repair (Yaari et al., 2013) and estimates the probability of any amino acid substitution in an antibody based on the frequencies observed in the computational simulation (**Figure 1**).

First, we applied ARMADiLLO retrospectively to the analysis of a mutation in a bnAb lineage that occurred at an AID cold spot that we have previously shown was functionally important for neutralization (Bonsignori et al., 2017). The DH270 V3-glycan bnAb lineage developed a variable heavy chain (V_H) complementary determining region 2 (CDR H2) G57R mutation that when analyzed with the ARMADiLLO program was predicted to occur with <1% frequency prior to selection (**Figure S1**). This mutation was functionally critical because reversion back to G57 in the DH270 bnAb lineage resulted in total loss of neutralization potency and breadth (Bonsignori et al., 2017). Thus, the ARMADiLLO program can identify a known, key improbable mutation.

All BCR mutations arise during the stochastic process of somatic hypermutation prior to antigenic selection. In HIV-1 infection, antibody heterologous breadth is not directly selected for during bnAb development because BCRs only interact with autologous virus Envs. Since improbable bnAb mutations can confer heterologous breadth, they represent critical events in bnAb development, and make compelling targets for focusing selection with immunogens. To test this hypothesis, we analyzed three additional bnAb lineages with ARMADiLLO to identify improbable mutations (defined as <2% estimated probability of occurring prior to selection; see Materials and Methods) and then tested for the effect of these mutations on bnAb neutralization during bnAb B cell lineage development. We chose three lineages that allowed for study of different levels of maturation in bnAb development: CH235, mid-stage bnAb development

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Bonsignori et al., 2017); VRC01, late stage bnAb development (Wu et al., 2015); and BF520.1, early stage bnAb development in an infant (Simonich et al., 2016).

A Subset of Improbable Mutations Confer Heterologous Neutralization in CD4 Binding Site bnAb Lineages

CH235 is a CD4-binding site, CD4-mimicking (Gao et al., 2014) bnAb B cell lineage that evolved to 90% neutralization breadth and high potency over 5-6 years of infection and acquired 44 V_H amino acid mutations (Bonsignori et al., 2016). We identified improbable mutations in the heavy chain of an early intermediate member of the lineage (also termed CH235), reverted each to their respective germline-encoded amino acid, and then tested antibody mutants for neutralization against the heterologous, difficult-to-neutralize (tier 2) (Seaman et al., 2010) TRO.11 HIV-1 strain (**Figures 2A and S2A**). Single amino acid reversion mutations resulted in either a reduction or total loss of heterologous HIV-1 TRO.11 neutralization for each of three improbable mutations, K19T, W47L and G55W demonstrating that improbable mutations in the CH235 lineage were indeed critical and could confer heterologous neutralization.

Identification of the K19T mutation was of particular interest because the mutation was observed in all but one member of the CH235 bnAb lineage and was also present in two other CD4-binding site bnAbs (Scheid et al., 2011) from different individuals that shared the same VH gene segment (VH1-46) as CH235 (**Figure S3A**). We performed genomic sequencing of the individual from which the CH235 lineage was isolated and confirmed that K19T was a mutation and was not due to allelic variation in gene segment VH1-46 (**Figure S4**). Superposition of the CH235 complex into a fully-glycosylated trimer (Stewart-Jones et al., 2016) showed that the K19T mutation position was in close proximity to the N197 glycan site on the Env trimer

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(**Figures S3B and S3C**). The K19T mutation shortened the amino acid at this position thus allowing for larger glycan forms at the heterogeneously glycosylated N197 position (Behrens et al., 2016) providing a structural rationale for the effect of this mutation on heterologous breadth. Consistent with this hypothesis, CH235 neutralization of HIV-1 JR-FL, a tier 2 heterologous virus lacking the N197 glycan site, was unaffected by the T19K reversion mutation (**Table S1 and Figure S2A**). Moreover, we introduced the K19T mutation into the CH235 UCA and observed improved binding to an early autologous Env suggesting that the improbable K19T mutation may have been selected for by an early autologous virus variant (**Figure S3D**).

We next asked what role improbable mutations played in the maturation of a second broad and potent CD4 binding site-targeting bnAb lineage, termed VRC01, that acquired 43 V_H amino acid mutations (Zhou et al., 2010). We reverted improbable mutations in the VRC01 antibody and tested for their effects on neutralization of heterologous tier 2 HIV-1 JR-FL (**Figure 2B**). Reversion of improbable mutations reduced potency of heterologous neutralization of HIV-1 JRFL demonstrating that in the VRC01 CD4 binding site B cell lineage, single improbable amino acid substitutions can also have functional consequences for heterologous neutralization capacity. Improbable mutations identified by ARMADiLLO in the VRC01 light chain showed an even larger effect on reducing neutralization than heavy chain mutations (**Table S1 and Figure S2B**), further underscoring, along with an atypically short CDRL3 and a critical CDRL1 deletion (Zhou et al., 2013), the importance of key improbable events in the maturation of the VRC01 bnAb lineage.

An Improbable Mutation Associated with Accelerated BnAb Development

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Babies develop bnAbs earlier after HIV-1 infection than adults (Goo et al., 2014; Muenchhoff et al., 2016). We analyzed the glycan-V3 epitope targeting BF520.1 bnAb, isolated from an HIV-1 infected infant with many fewer mutations (12 V_H amino acid mutations) compared to VRC01 and CH235 bnAbs (Simonich et al., 2016). We identified an improbable mutation, N52A, located in the CDR H2 of BF520.1, reverted it to germline, and expressed the resultant antibody mutant (A52N). Heterologous neutralization of the A52N reversion mutant against tier 2 JR-FL virus was markedly reduced relative to wildtype BF520.1 (**Figure 2C**). The A52N reversion mutation antibody reduced neutralization potency for all tier 2 viruses that the BF520.1 bnAb could neutralize (**Table S1 and Figure S2C**) demonstrating that the N52A mutation was critical to the neutralization potency of BF520.1 and suggested the acquisition of this improbable mutation may have played a role in the relatively early elicitation (<15 months) of a bnAb with limited mutation frequency.

Improbable Antibody Mutations Are Enriched in BnAbs

While not all improbable mutations will be critical for bnAb development, a subset will be key, as demonstrated in the examples above. To provide a view of the scope of the problem for many bnAb B cell lineages, we estimated the number of improbable mutations for a representative set of known bnAb lineages spanning all known sites of vulnerability on the Env trimer (**Figures 3A and S5, Table S2**). Study of a representative sample of bnAb lineages is plausible because of commonalities of Env recognition by bnAb germline precursors (Andrabi et al., 2015; Bonsignori et al., 2011; Gorman et al., 2016; Zhou et al., 2013). Compared to Env-reactive antibodies induced by an HIV-1 vaccine candidate (the RV144 vaccine) (Rerks-Ngarm et al., 2009) or antibodies isolated from non-HIV-1 infected individuals (Williams et al., 2015),

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the broadest and most potent HIV-1 bnAbs had the highest numbers of improbable mutations (**Figure 3B**). This result may follow directly from the observations that bnAbs tend to be highly mutated (**Figure S6A**) (Burton and Hangartner, 2016), and the number of improbable mutations an antibody possesses is correlated with its mutation frequency (**Figure S6B**) (Sheng et al., 2017). However, it is not known why most bnAbs are highly mutated. Recent work has shown that not all mutations in bnAbs are essential for neutralization activity (Jardine et al., 2016b). One hypothesis is that high mutation frequency is due to the extended number of rounds of somatic hypermutation required for a lineage to acquire a specific subset of mutations (Klein et al., 2013). If some of those specific mutations are also improbable, it is very likely that more probable mutations would be acquired prior to attaining key improbable ones. We found that for many bnAbs the number of improbable mutations exceeded what would be expected by chance given their high mutation frequency (**Table S3**). This observation, along with our experimental observations demonstrating that a subset of improbable mutations are important for neutralization capacity, is consistent with the notion that improbable mutations may act as key bottlenecks in the development of bnAb neutralization breadth. Thus, during chronic HIV-1 infection with persistent high viral loads that are required for bnAbs with improbable mutations to develop (Gray et al., 2011), excess numbers of probable mutations also accumulate. Probable mutations arise easily from the intrinsic mutability of antibody genes, and unlike improbable mutations, may not require Env selection (Bonsignori et al., 2016; Hwang et al., 2017; Neuberger et al., 1998). Thus, if the selection of critical improbable mutations can be targeted with Env immunogens, it should be possible to accelerate bnAb maturation and result in the induction of bnAb lineages with fewer mutations than those that occur in the setting of chronic HIV-1 infection.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Implications for Vaccine Design

The ability to identify functional improbable bnAb mutations using the ARMADiLLO program and antibody functional studies informs a mutation-guided vaccine design and immunization strategy (**Figure 4**). The principal goal is to be able to choose the correct sequential Envs to precisely focus selection towards the most difficult-to-induce mutations, while allowing the key easier, more probable mutations to occur due to antibody intrinsic mutability from immune activation-associated AID activity. In this strategy, improbable mutations are identified computationally using the ARMADiLLO program. Next, improbable mutations identified are expressed as single amino acid substitution mutant antibodies and their functional importance validated by Env binding and neutralization assays. Mutations in both maturation and reversion directions (UCA antibody with the single mutation added and mature antibody with the mutation reverted to germline, respectively) may be necessary to define the functional role of each mutation. Env mutants that bind to the UCA mutant and subsequent mutated intermediates with high affinity are chosen as immunogens to select for these functional improbable mutations. Last, sequential immunization with the chosen immunogens are studied for optimization of regimens to select for B cells with BCRs with the required improbable mutations. The vaccine strategy outlined here requires an adequate priming immunogen capable of engaging bnAb precursors (Dosenovic et al., 2015; Escolano et al., 2016; Jardine et al., 2016a; Jardine et al., 2015; McGuire et al., 2016; Sok et al., 2016; Williams et al., 2017; Zhang et al., 2016) prior to boosting to select for improbable mutations.

DISCUSSION

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Here we report the identification of improbable mutations in three bnAb lineages spanning multiple epitope specificities and representing distinct stages of bnAb development. We showed through functional antibody analysis that certain improbable mutations in these lineages were critical for the development of neutralization breadth and potency. In addition, we showed that bnAbs from HIV infected individuals were enriched for the number of improbable mutations relative to antibodies isolated from uninfected or HIV-1 vaccinated subjects. We conclude that improbable mutations that are functionally important for neutralization capacity can act as barriers for bnAb development in many bnAb lineages. Thus, the elicitation of important improbable mutations should be a high priority in vaccination. To this end, we proposed a vaccine design strategy that aims to target selection of critical improbable mutations with the goal of circumventing these barriers and accelerating the maturation of a bnAb response.

Vaccine strategies aiming to select for specific maturation mutations in a clonal lineage first requires adequate priming immunogens capable of engaging bnAb precursors. Recent progress has been made in the design of immunogens that bind bnAb UCAs (Jardine et al., 2016a; McGuire et al., 2013; Steichen et al., 2016; Zhang et al., 2016). Once UCAs are engaged, a vaccine must elicit mutations that keep antibody maturation on a pathway towards heterologous breadth. Immunization studies with fully germline-reverted bnAb UCA knock-in mice have so far failed to demonstrate the acquisition of the specific mutations that lead to full bnAb neutralization capacity (Briney et al., 2016; Tian et al., 2016) suggesting that the pathways that lead to breadth are blocked and not circumvented with the current generation of immunogens. We propose here that one significant barrier could be the acquisition of important

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mutations that are highly improbable and thus infrequent prior to selection in the germinal center.

If a mutation occurs prior to selection with less than 2% probability, the cutoff we use here to define an improbable mutation, its expected frequency in a germinal center would be less than 1 in 50 clonally related B cells. Estimates of the number of clonally related B cells range from 10-100 B cells per germinal center (Jacob et al., 1991; Tas et al., 2016) implying that the frequency of a single improbable mutation in any given germinal center will be two or fewer, per round of mutation. Thus, in order for improbable mutations to arise and be available for antigenic selection, bnAb B cell clones must either fortuitously acquire an important improbable mutation in the first round of mutation or must survive long enough in the germinal center to undergo many rounds of mutations and selection to increase the opportunities to acquire that improbable mutation. Moreover, B cell survival is dictated not just by having higher affinity to antigen than other clonally related members, but also by having higher affinity than other clones in the germinal center. In addition, heterologous neutralization breadth is never selected for directly and thus the competition is confounded by selecting for autologous antigen affinity. If an improbable mutation important for heterologous neutralization does not simultaneously increase affinity for the antigen present in the germinal center, it will not be selected. Thus, a vaccine design strategy must select for mutations that increase antibody affinity, however the sequence of acquisition of mutations is important and will require experimental determination. An additional caveat is that improbable mutations leading to different but chemically similar amino acids may be functionally equivalent.

Because improbable mutations arise as either rare intrinsic mutations or by selection by antigens derived from autologous virus, not all improbable mutations are required for mediation

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of heterologous neutralization (**Table S1**). Thus, the probability of a mutation is not predictive of its essential nature for neutralization breadth. While the recurrence of an improbable mutation within many clonal members can indicate positive selection, the selection criteria in B cell maturation is not heterologous neutralization capacity but rather is autologous antigen affinity. Consequently, improbable mutations must be evaluated for their functional relevance, and only the subset of improbable mutations that are also important for acquisition of neutralization breadth are considered potential bottlenecks for bnAb development. Thus, while we showed that bnAbs are enriched for the number of improbable mutations relative to antibodies from uninfected or vaccinated individuals, work remains to functionally assess all of the improbable mutations to define those required for bnAb affinity maturation. It is important to note as well that intrinsically mutable positions (Neuberger et al., 1998) can also be capable of conferring heterologous breadth. In this regard we identified one such functionally important probable intrinsic mutation in the CH235 lineage, S57R (**Table S1**). However, such highly probable mutations, by definition, are easily inducible and are not likely to represent barriers in bnAb development.

Interestingly, bnAbs that demonstrate relatively low numbers of improbable single somatic mutations (**Figure 3A**) possessed other unusual antibody characteristics that were due to additional improbable events such as insertion/deletions (indels) or extraordinary CDR H3 lengths. For example, the bnAbs with the two lowest number of improbable mutations were PGT128 and CAP256-VRC26.25. These bnAbs are notable for having the largest indels (PGT128; 11 aas) or the longest CDR H3 (CAP256-VRC26.25; 38 aas) of the bnAbs we analyzed. In summary, our data presented here suggest Env trimers evolved to evade neutralizing B-cell responses by hiding within low probability regions of antibody sequence

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

space. The ARMADiLLO program and mutation-guided vaccine design strategy presented here should be broadly applicable for vaccine design for other mutating pathogens.

EXPERIMENTAL PROCEDURES

Study Design

Malawian individual CH505, from whom the CH235 lineage was isolated and VH1-46 gene segment alleles determined by genomic sequencing, was enrolled in the Center for HIV/AIDS Vaccine Immunology (CHAVI) 001, nonblinded, nonrandomized, observational study protocol at a CHAVI clinical site in Malawi after informed consent was obtained under protocols approved by the Institutional Review Board of the Duke University Health System, the National Institutes of Health, and a clinical site review board in Malawi (Tomaras et al., 2011).

Analysis of the Probability of Antibody Mutations

The probability of an amino acid substitution at any given position in the antibody sequence of an antibody of interest was estimated using the ARMADiLLO program. The algorithm and the analysis performed using ARMADiLLO are described in Supplemental Experimental Procedures.

Antibody Site-directed Mutagenesis

BF520.1 mutant antibody genes were synthesized by Genscript and recombinantly produced. Mutations into antibody genes for CH235 and VRC01 mutants were introduced using the QuikChange II Lightning site-directed mutagenesis kit (Agilent Technologies) following the manufacturer's protocol. Single-colony sequencing was used to confirm the sequences of the

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mutant plasmid products. Primers used for introducing mutations are listed in the Supplemental Experimental Procedures.

Recombinant Antibody Production

Antibodies were recombinantly produced as previously described (Saunders et al., 2017).

HIV-1 Neutralization

Antibody neutralization was measured in TZM-bl cell-based neutralization assays as previously described (Li et al., 2005; Sarzotti-Kelsoe et al., 2014). CH235 and BF520.1 and selected mutants were assayed for neutralization using a global panel of 12 HIV-1 Env reference strains (deCamp et al., 2014). Neutralization values are reported as inhibitory concentrations of antibody in which 50% of virus was neutralized (IC_{50}) with units in μg/ml.

Antibody Binding Measurements

Binding of CH235.UCA and mutants to the monomeric CH505 transmitted/founder (T/F) delta7 gp120 and monomeric CH505 M5 (early autologous virus variant) delta8 gp120 (Bonsignori et al., 2016; Gao et al., 2014) was measured by surface plasmon resonance assays (SPR) on a Biacore S200 instrument and data analysis was performed with the S200 BIAevaluation software (Biacore/GE Healthcare) as previously described (Alam et al., 2013; Dennison et al., 2011).

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

Alam, S.M., Liao, H.X., Tomaras, G.D., Bonsignori, M., Tsao, C.Y., Hwang, K.K., Chen, H., Lloyd, K.E., Bowman, C., Sutherland, L., *et al.* (2013). Antigenicity and immunogenicity of RV144 vaccine AIDSVAX clade E envelope immunogen is enhanced by a gp120 N-terminal deletion. J Virol *87*, 1554-1568.

Andrabi, R., Voss, J.E., Liang, C.H., Briney, B., McCoy, L.E., Wu, C.Y., Wong, C.H., Poignard, P., and Burton, D.R. (2015). Identification of Common Features in Prototype Broadly Neutralizing Antibodies to HIV Envelope V2 Apex to Facilitate Vaccine Design. Immunity *43*, 959-973.

Behrens, A.J., Vasiljevic, S., Pritchard, L.K., Harvey, D.J., Andev, R.S., Krumm, S.A., Struwe, W.B., Cupo, A., Kumar, A., Zitzmann, N., *et al.* (2016). Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. Cell Rep *14*, 2695-2706.

Betz, A.G., Rada, C., Pannell, R., Milstein, C., and Neuberger, M.S. (1993). Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. Proc Natl Acad Sci U S A *90*, 2385-2388.

Bonsignori, M., Hwang, K.K., Chen, X., Tsao, C.Y., Morris, L., Gray, E., Marshall, D.J., Crump, J.A., Kapiga, S.H., Sam, N.E., *et al.* (2011). Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. J Virol *85*, 9998-10009.

Bonsignori, M., Kreider, E.F., Fera, D., Meyerhoff, R.R., Bradley, T., Wiehe, K., Alam, S.M., Aussedat, B., Walkowicz, W.E., Hwang, K.K., *et al.* (2017). Staged induction of HIV-1 glycan-dependent broadly neutralizing antibodies. Sci Transl Med *9*.

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

Briney, B., Sok, D., Jardine, J.G., Kulp, D.W., Skog, P., Menis, S., Jacak, R., Kalyuzhniy, O., de Val, N., Sesterhenn, F., *et al.* (2016). Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies. Cell *166*, 1459-1470 e1411.

Brown, M., Stenzel-Poore, M., Stevens, S., Kondoleon, S.K., Ng, J., Bachinger, H.P., and Rittenberg, M.B. (1992). Immunologic memory to phosphocholine keyhole limpet hemocyanin. Recurrent mutations in the lambda 1 light chain increase affinity for antigen. J Immunol *148*, 339-346.

Burton, D.R., and Hangartner, L. (2016). Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. Annu Rev Immunol *34*, 635-659.

Cowell, L.G., and Kepler, T.B. (2000). The nucleotide-replacement spectrum under somatic hypermutation exhibits microsequence dependence that is strand-symmetric and distinct from that under germline mutation. J Immunol *164*, 1971-1976.

De Silva, N.S., and Klein, U. (2015). Dynamics of B cells in germinal centres. Nat Rev Immunol *15*, 137-148.

deCamp, A., Hraber, P., Bailer, R.T., Seaman, M.S., Ochsenbauer, C., Kappes, J., Gottardo, R., Edlefsen, P., Self, S., Tang, H., *et al.* (2014). Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. J Virol *88*, 2489-2507.

Dennison, S.M., Anasti, K., Scearce, R.M., Sutherland, L., Parks, R., Xia, S.M., Liao, H.X., Gorny, M.K., Zolla-Pazner, S., Haynes, B.F., and Alam, S.M. (2011). Nonneutralizing HIV-1 gp41 envelope cluster II human monoclonal antibodies show polyreactivity for binding to phospholipids and protein autoantigens. J Virol *85*, 1340-1347.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem *76*, 1-22.

Doria-Rose, N.A., Schramm, C.A., Gorman, J., Moore, P.L., Bhiman, J.N., DeKosky, B.J., Ernandes, M.J., Georgiev, I.S., Kim, H.J., Pancera, M., *et al.* (2014). Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature *509*, 55-62.

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

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Easterhoff, D., Moody, M.A., Fera, D., Cheng, H., Ackerman, M., Wiehe, K., Saunders, K.O., Pollara, J., Vandergrift, N., Parks, R., *et al.* (2017). Boosting of HIV envelope CD4 binding site antibodies with long variable heavy third complementarity determining region in the randomized double blind RV305 HIV-1 vaccine trial. PLoS Pathog *13*, e1006182.

Escolano, A., Steichen, J.M., Dosenovic, P., Kulp, D.W., Golijanin, J., Sok, D., Freund, N.T., Gitlin, A.D., Oliveira, T., Araki, T., *et al.* (2016). Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice. Cell *166*, 1445-1458 e1412.

Gao, F., Bonsignori, M., Liao, H.X., Kumar, A., Xia, S.M., Lu, X., Cai, F., Hwang, K.K., Song, H., Zhou, T., *et al.* (2014). Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. Cell *158*, 481-491.

Georgiev, I.S., Rudicell, R.S., Saunders, K.O., Shi, W., Kirys, T., McKee, K., O'Dell, S., Chuang, G.Y., Yang, Z.Y., Ofek, G., *et al.* (2014). Antibodies VRC01 and 10E8 neutralize HIV-1 with high breadth and potency even with Ig-framework regions substantially reverted to germline. J Immunol *192*, 1100-1106.

Goo, L., Chohan, V., Nduati, R., and Overbaugh, J. (2014). Early development of broadly neutralizing antibodies in HIV-1-infected infants. Nat Med *20*, 655-658.

Gorman, J., Soto, C., Yang, M.M., Davenport, T.M., Guttman, M., Bailer, R.T., Chambers, M., Chuang, G.Y., DeKosky, B.J., Doria-Rose, N.A., *et al.* (2016). Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. Nat Struct Mol Biol *23*, 81-90.

Gray, E.S., Madiga, M.C., Hermanus, T., Moore, P.L., Wibmer, C.K., Tumba, N.L., Werner, L., Mlisana, K., Sibeko, S., Williamson, C., *et al.* (2011). The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. J Virol *85*, 4828-4840.

Haynes, B.F., and Burton, D.R. (2017). Developing an HIV vaccine. Science *355*, 1129-1130.

Hraber, P., Seaman, M.S., Bailer, R.T., Mascola, J.R., Montefiori, D.C., and Korber, B.T. (2014). Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. AIDS *28*, 163-169.

Hwang, J.K., Wang, C., Du, Z., Meyers, R.M., Kepler, T.B., Neuberg, D., Kwong, P.D., Mascola, J.R., Joyce, M.G., Bonsignori, M., *et al.* (2017). Sequence Intrinsic Somatic Mutation Mechanisms Contribute to Affinity Maturation of VRC01-class HIV-1 Broadly Neutralizing Antibodies. Proc Natl Acad Sci U S A *In Press*.

Jacob, J., Kelsoe, G., Rajewsky, K., and Weiss, U. (1991). Intraclonal generation of antibody mutants in germinal centres. Nature *354*, 389-392.

Jardine, J.G., Kulp, D.W., Havenar-Daughton, C., Sarkar, A., Briney, B., Sok, D., Sesterhenn, F., Ereno-Orbea, J., Kalyuzhniy, O., Deresa, I., *et al.* (2016a). HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. Science *351*, 1458-1463.

Jardine, J.G., Ota, T., Sok, D., Pauthner, M., Kulp, D.W., Kalyuzhniy, O., Skog, P.D., Thinnes, T.C., Bhullar, D., Briney, B., *et al.* (2015). HIV-1 VACCINES. Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. Science *349*, 156-161.

Jardine, J.G., Sok, D., Julien, J.P., Briney, B., Sarkar, A., Liang, C.H., Scherer, E.A., Henry Dunand, C.J., Adachi, Y., Diwanji, D., *et al.* (2016b). Minimally Mutated HIV-1 Broadly Neutralizing Antibodies to Guide Reductionist Vaccine Design. PLoS Pathog *12*, e1005815.

Kelsoe, G., and Haynes, B.F. (2017). Host controls of HIV broadly neutralizing antibody development. Immunol Rev *275*, 79-88.

Kepler, T.B., Liao, H.X., Alam, S.M., Bhaskarabhatla, R., Zhang, R., Yandava, C., Stewart, S., Anasti, K., Kelsoe, G., Parks, R., *et al.* (2014). Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. Cell Host Microbe *16*, 304-313.

Klein, F., Diskin, R., Scheid, J.F., Gaebler, C., Mouquet, H., Georgiev, I.S., Pancera, M., Zhou, T., Incesu, R.B., Fu, B.Z., *et al.* (2013). Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. Cell *153*, 126-138.

Kocks, C., and Rajewsky, K. (1988). Stepwise intraclonal maturation of antibody affinity through somatic hypermutation. Proc Natl Acad Sci U S A *85*, 8206-8210.

Li, M., Gao, F., Mascola, J.R., Stamatatos, L., Polonis, V.R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K.M., *et al.* (2005). Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. J Virol *79*, 10108-10125.

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

McGuire, A.T., Gray, M.D., Dosenovic, P., Gitlin, A.D., Freund, N.T., Petersen, J., Correnti, C., Johnsen, W., Kegel, R., Stuart, A.B., *et al.* (2016). Specifically modified Env immunogens activate B-cell precursors of broadly neutralizing HIV-1 antibodies in transgenic mice. Nat Commun *7*, 10618.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

McGuire, A.T., Hoot, S., Dreyer, A.M., Lippy, A., Stuart, A., Cohen, K.W., Jardine, J., Menis, S., Scheid, J.F., West, A.P., *et al.* (2013). Engineering HIV envelope protein to activate germline B cell receptors of broadly neutralizing anti-CD4 binding site antibodies. J Exp Med *210*, 655-663.

Moody, M.A., Zhang, R., Walter, E.B., Woods, C.W., Ginsburg, G.S., McClain, M.T., Denny, T.N., Chen, X., Munshaw, S., Marshall, D.J., *et al.* (2011). H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin antibodies than influenza vaccination. PLoS One *6*, e25797.

Muenchhoff, M., Adland, E., Karimanzira, O., Crowther, C., Pace, M., Csala, A., Leitman, E., Moonsamy, A., McGregor, C., Hurst, J., *et al.* (2016). Nonprogressing HIV-infected children share fundamental immunological features of nonpathogenic SIV infection. Sci Transl Med *8*, 358ra125.

Neuberger, M.S., Ehrenstein, M.R., Klix, N., Jolly, C.J., Yelamos, J., Rada, C., and Milstein, C. (1998). Monitoring and interpreting the intrinsic features of somatic hypermutation. Immunol Rev *162*, 107-116.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103-107.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., *et al.* (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med *361*, 2209-2220.

Sarzotti-Kelsoe, M., Bailer, R.T., Turk, E., Lin, C.L., Bilska, M., Greene, K.M., Gao, H., Todd, C.A., Ozaki, D.A., Seaman, M.S., *et al.* (2014). Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. J Immunol Methods *409*, 131-146.

Saunders, K.O., Nicely, N.I., Wiehe, K., Bonsignori, M., Meyerhoff, R.R., Parks, R., Walkowicz, W.E., Aussedat, B., Wu, N.R., Cai, F., *et al.* (2017). Vaccine Elicitation of High Mannose-Dependent Neutralizing Antibodies against the V3-Glycan Broadly Neutralizing Epitope in Nonhuman Primates. Cell Rep *18*, 2175-2188.

Scheid, J.F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T.Y., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., *et al.* (2011). Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. Science *333*, 1633-1637.

Seaman, M.S., Janes, H., Hawkins, N., Grandpre, L.E., Devoy, C., Giri, A., Coffey, R.T., Harris, L., Wood, B., Daniels, M.G., *et al.* (2010). Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. J Virol *84*, 1439-1452.

Sheng, Z., Schramm, C.A., Kong, R., Program, N.C.S., Mullikin, J.C., Mascola, J.R., Kwong, P.D., and Shapiro, L. (2017). Gene-Specific Substitution Profiles Describe the Types and Frequencies of Amino Acid Changes during Antibody Somatic Hypermutation. Front Immunol *8*, 537.

Simonich, C.A., Williams, K.L., Verkerke, H.P., Williams, J.A., Nduati, R., Lee, K.K., and Overbaugh, J. (2016). HIV-1 Neutralizing Antibodies with Limited Hypermutation from an Infant. Cell *166*, 77-87.

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

Steichen, J.M., Kulp, D.W., Tokatlian, T., Escolano, A., Dosenovic, P., Stanfield, R.L., McCoy, L.E., Ozorowski, G., Hu, X., Kalyuzhniy, O., *et al.* (2016). HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. Immunity *45*, 483-496.

Stewart-Jones, G.B., Soto, C., Lemmin, T., Chuang, G.Y., Druz, A., Kong, R., Thomas, P.V., Wagh, K., Zhou, T., Behrens, A.J., *et al.* (2016). Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. Cell *165*, 813-826.

Tas, J.M., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J.T., Mano, Y.M., Chen, C.S., Weill, J.C., Reynaud, C.A., Browne, E.P., *et al.* (2016). Visualizing antibody affinity maturation in germinal centers. Science *351*, 1048-1054.

Teng, G., and Papavasiliou, F.N. (2007). Immunoglobulin somatic hypermutation. Annu Rev Genet *41*, 107-120.

Tian, M., Cheng, C., Chen, X., Duan, H., Cheng, H.L., Dao, M., Sheng, Z., Kimble, M., Wang, L., Lin, S., *et al.* (2016). Induction of HIV Neutralizing Antibody Lineages in Mice with Diverse Precursor Repertoires. Cell *166*, 1471-1484 e1418.

Tomaras, G.D., Binley, J.M., Gray, E.S., Crooks, E.T., Osawa, K., Moore, P.L., Tumba, N., Tong, T., Shen, X., Yates, N.L., *et al.* (2011). Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals. J Virol *85*, 11502-11519.

Victora, G.D., and Nussenzweig, M.C. (2012). Germinal centers. Annu Rev Immunol *30*, 429-457.

Williams, W.B., Liao, H.X., Moody, M.A., Kepler, T.B., Alam, S.M., Gao, F., Wiehe, K., Trama, A.M., Jones, K., Zhang, R., *et al.* (2015). HIV-1 VACCINES. Diversion of HIV-1 vaccine-induced immunity by gp41-microbiota cross-reactive antibodies. Science *349*, aab1253.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Williams, W.B., Zhang, J., Jiang, C., Nicely, N.I., Fera, D., Luo, K., Moody, M.A., Liao, H.X., Alam, S.M., Kepler, T.B., *et al.* (2017). Initiation of HIV neutralizing B cell lineages with sequential envelope immunizations. Nat Commun *8*, 1732.

Wu, X., Zhang, Z., Schramm, C.A., Joyce, M.G., Kwon, Y.D., Zhou, T., Sheng, Z., Zhang, B., O'Dell, S., McKee, K., *et al.* (2015). Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-1 Infection. Cell *161*, 470-485.

Yaari, G., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Gupta, N., Stern, J.N., O'Connor, K.C., Hafler, D.A., Laserson, U., Vigneault, F., and Kleinstein, S.H. (2013). Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. Front Immunol *4*, 358.

Yeap, L.S., Hwang, J.K., Du, Z., Meyers, R.M., Meng, F.L., Jakubauskaite, A., Liu, M., Mani, V., Neuberg, D., Kepler, T.B., *et al.* (2015). Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes. Cell *163*, 1124-1137.

Yu, L., and Guan, Y. (2014). Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. Front Immunol *5*, 250.

Zhang, R., Verkoczy, L., Wiehe, K., Munir Alam, S., Nicely, N.I., Santra, S., Bradley, T., Pemble, C.W.t., Zhang, J., Gao, F., *et al.* (2016). Initiation of immune tolerance-controlled HIV gp41 neutralizing B cell lineages. Sci Transl Med *8*, 336ra362.

Zhou, T., Georgiev, I., Wu, X., Yang, Z.Y., Dai, K., Finzi, A., Kwon, Y.D., Scheid, J.F., Shi, W., Xu, L., *et al.* (2010). Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science *329*, 811-817.

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

ACKNOWLEDGEMENTS:

We would like to thank Bette Korber for many helpful discussions. We thank Shia-Mao Xia, Melissa Cooper, Kara Anasti, and Amanda Eaton for technical assistance. This work was funded by UM1 AI100645 from the Duke CHAVI–Immunogen Discovery, Division of AIDS, NIAID, NIH (to B.F.H.). T.B. and K.W. were supported by the Duke University Center for AIDS Research (CFAR), an NIH funded program (5P30 AI064518). R.R.M is supported by a Medical Scientist Training Program (MSTP) training grant (T32GM007171) and the Ruth L. Kirschstein National Research Service Award F30-AI122982-0, NIAID.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIGURES

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

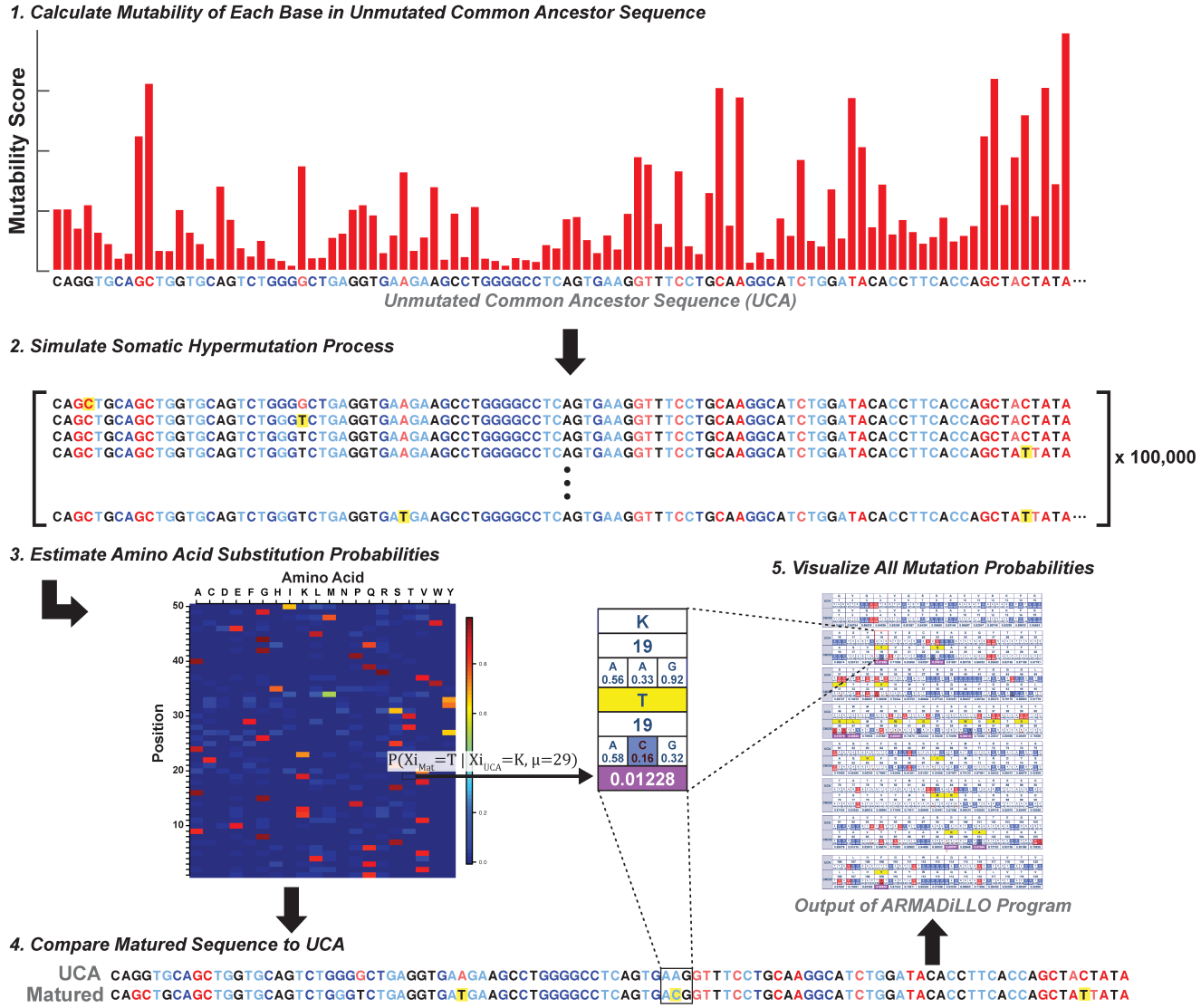


Figure 1. Computational method for estimating the probability of antibody mutations

The probability of an amino acid substitution during B cell maturation in the absence of selection is estimated by simulating the somatic hypermutation process. 1) The inferred unmutated common ancestor sequence (UCA) of the antibody of interest is assigned mutability scores according to a statistical model of AID targeting. 2) Bases in the sequence are then drawn randomly according to these scores and mutated according to a base substitution model (see methods). Rounds of single base mutation continue for the number of mutations observed in the antibody of interest with mutability scores updated as the simulation proceeds. The simulation is then repeated 100,000 times to generate a set of synthetic matured sequences. 3) An amino acid positional frequency matrix is constructed from the simulated sequences and utilized to estimate the probability of amino acid substitutions. 4) The UCA and matured sequence are aligned and 5) the estimated probability of amino acid substitutions identified in the matured sequence are outputted.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

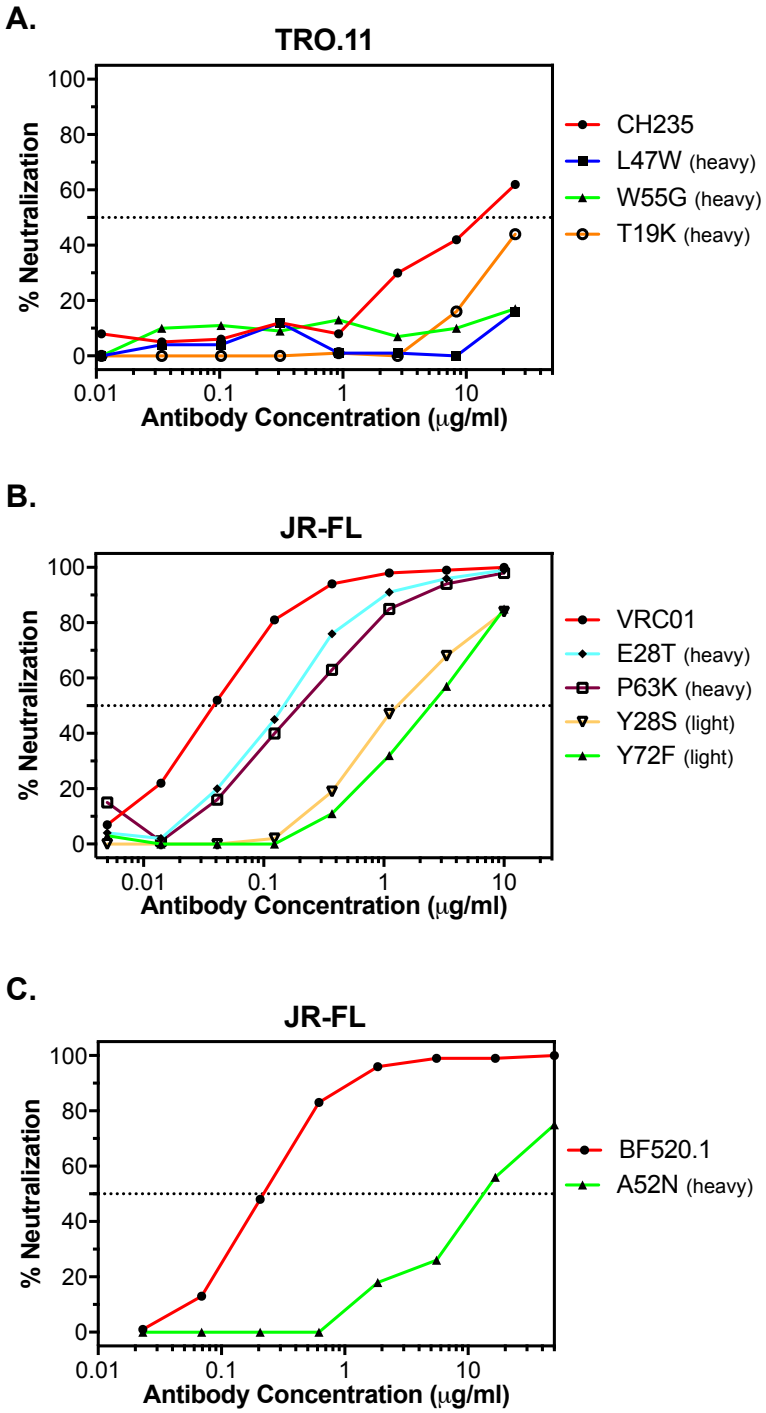


Figure 2. Improbable mutations confer heterologous neutralization in bnAb development

BnAbs A) CH235, B) VRC01 and C) BF520.1 and their corresponding mutants with reverted improbable mutations were tested for neutralization against heterologous viruses. The reversion of improbable mutations in all three bnAbs diminished neutralization potency.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A.

			Antibody Characteristics				Neutralization	
Epitope	Type	HIV BNAb	# Improbable Mutations[a]	#AA Muts[b]	# Indels[c]	CDR3 Length[d]	Geo Mean Potency[e]	Breadth[f]
CD4	CD4 mimicking VH1-46 derived	CH235.12	22	67	0	21	0.69	100%
		8ANC131	17	69	3	25	10.23	100%*
		1B2530	13	60	0	27	2.74	100%*
	CD4 mimicking VH1-2 derived	VRC01	23	71	3	17	0.59	92%
		VRC-CH31	15	70	9	18	0.20	92%
		3BNC117	16	68	8	15	0.29	92%
		N6	19	75	0	18	0.08	100%*
	CDRH3 Binder	CH103	9	49	3	23	3.11	80%*
Glycan-V3		DH270.6	11	34	0	28	0.59	67%
		BF520.1	12	24	0	29	5.28	82%*
		PGT128	7	53	11	29	0.04	83%
		PGT121	15	55	3	36	0.30	64%
		PGT135	18	64	5	27	9.36	50%*
Glycan-V1V2		PG9	14	28	0	38	0.17	92%
		CH01	16	44	0	33	2.26	75%
		PGT145	12	52	0	40	0.05	86%*
		CAP256-VRC26.25	8	27	0	48	0.01	78%*
		PGDM1400	14	67	0	41	0.02	100%*
MPER		2F5	10	31	0	31	17.40	50%
		4E10	10	30	0	27	3.09	100%
		DH511.2	19	46	0	32	0.25	100%
		10E8	14	45	0	32	0.21	100%
gp120-gp41 Interface		PGT151	17	41	0	35	0.26	83%
		35022	23	78	8	24	2.50	58%
		N123-VRC34	10	37	0	22	7.49	42%
		8ANC195	20	77	8	29	1.60	83%

a Probability of mutations in the absence of selection estimated by the ARMADiLLO program. Due to inconsistent uncertainity levels in the inference of the CDRH3 region between BNAbs, only mutations outside the CDR3 were used.

b Number of total amino acids mutations

c Number of amino acids inserted or deleted relative to UCA

d Sum of heavy and light chain CDR3 length

e Geometric mean neutralization potency (μg/ml) of BNAb against a tier-2 12-virus global panel

f Neutralization breadth of BNAb against a standard tier-2 12-virus global panel

* The number of viruses run in panel was <12

B.

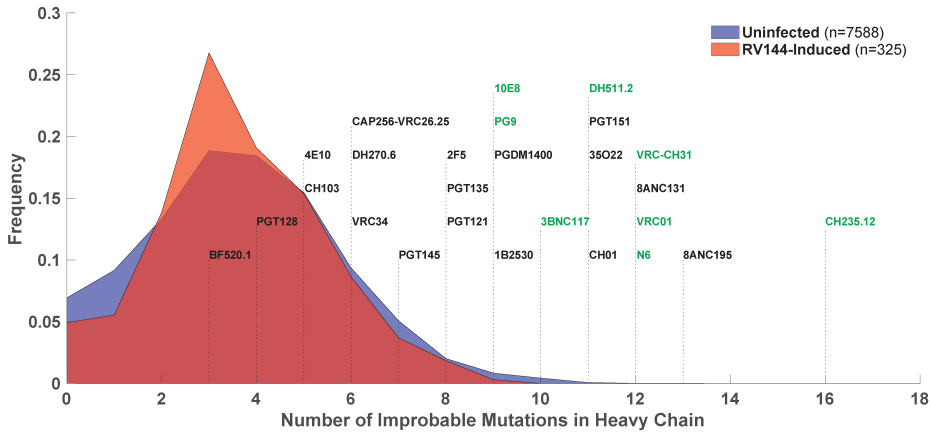


Figure 3. BnAbs are enriched for improbable antibody mutations

A) Table of improbable mutations for a representative set of bnAbs B) Histograms of the distributions of number of improbable mutations from antibody heavy chain sequences from three groups: "RV144-induced" antibodies were isolated from RV144 vaccinated subjects by antigenically sorting with RV144 immunogens (red shaded area); "Uninfected" antibodies correspond to duplicated NGS reads from IgG antibodies isolated from PBMC samples from 8 HIV-uninfected individuals (blue shaded area; see methods for details on sampling); a representative set of published bnAb antibody sequences are shown labeled above dotted lines that correspond to their number of improbable mutations (at the <2% level).

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

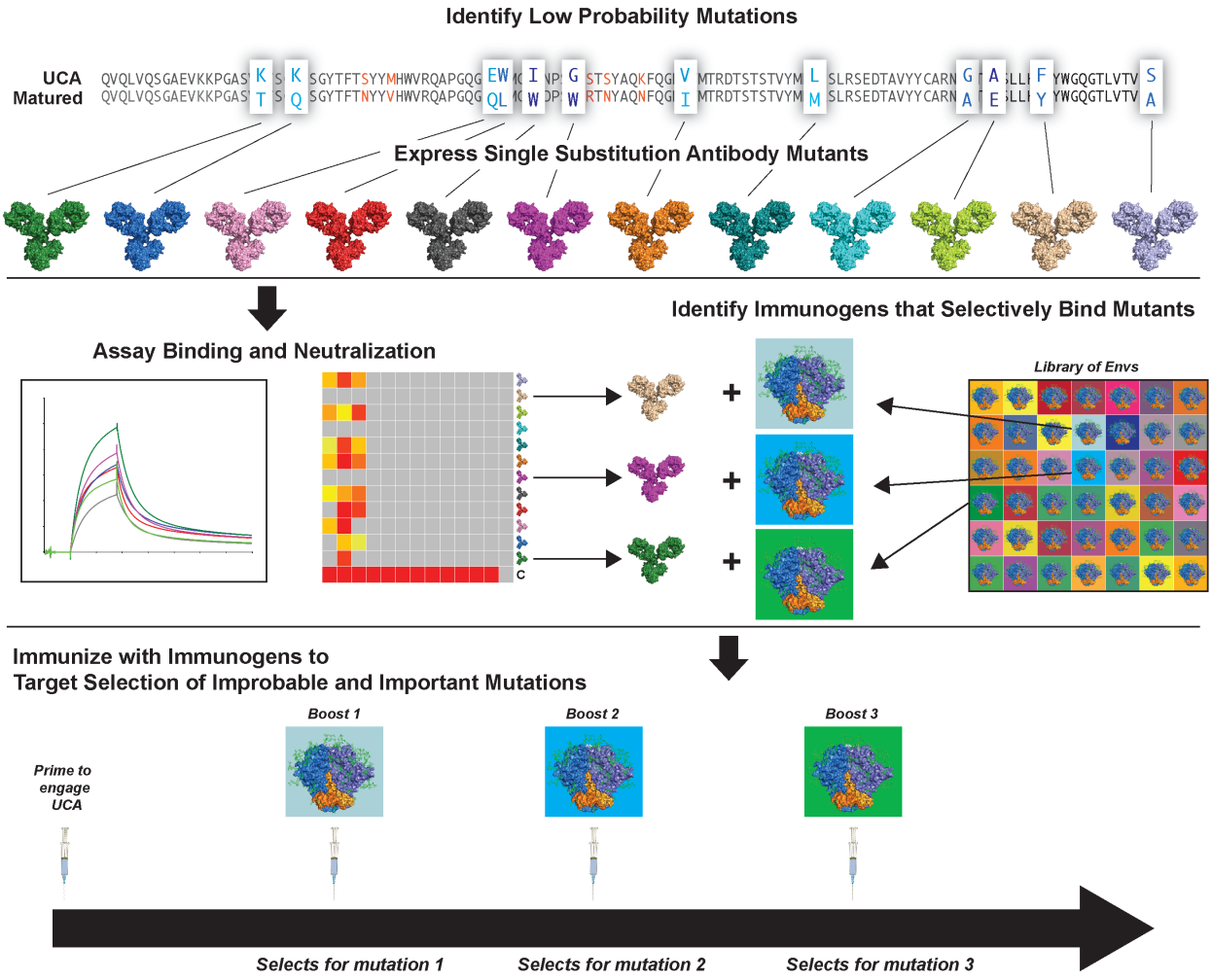


Figure 4. Mutation Guided Lineage Design Vaccine Strategy

Improbable mutations can act as important bottlenecks in the development of bnAbs and we propose here a strategy to specifically target those mutations for selection through vaccination. First, for a specific bnAb lineage, low probability mutations are identified computationally and recombinant antibody mutants corresponding to these mutations are produced (top panel). Binding and neutralization assays are performed to validate which of the improbable mutations are functionally important for lineage development (middle panel, left) and Envs are chosen that can specifically bind the corresponding antibody mutants (middle panel, right). These Envs are then used in a sequential immunization regimen to select the most difficult-to-induce, critical mutations thus potentially alleviating key bottlenecks in bnAb elicitation.

bioRxiv preprint doi: https://doi.org/10.1101/262592; this version posted February 9, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPLEMENTAL INFORMATION

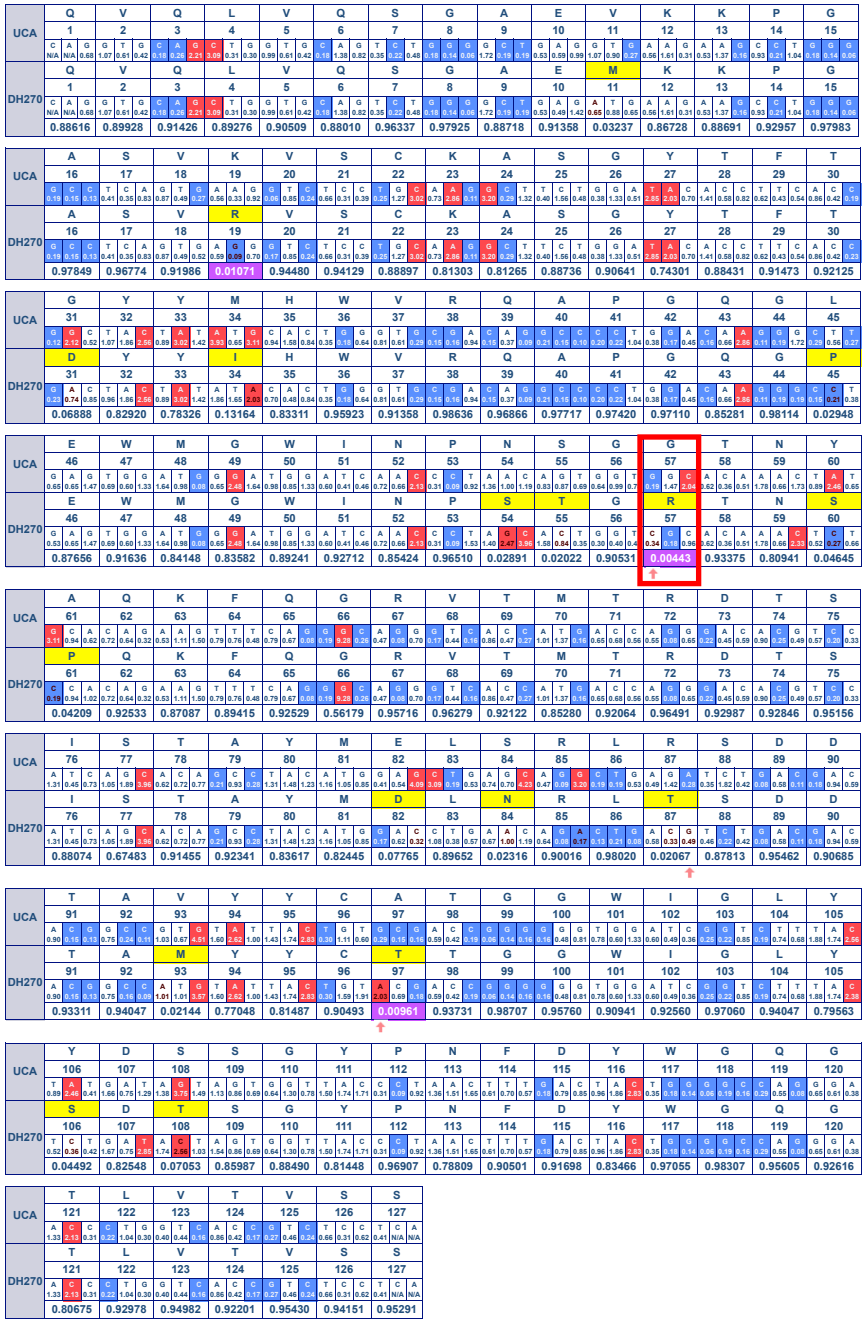


Figure S1. ARMADiLLO output for DH270 heavy chain shows G57R mutation is improbable

ARMADiLLO output for the DH270 heavy chain. The first three rows of each block corresponds to the DH270 UCA sequence and the following four rows correspond to the matured DH270 sequenced. The first row is the amino acid sequence for the DH270 UCA. The second row is the amino acid numbering (consecutively numbered starting at 1 for the first residue) for the DH270 UCA. The third row is the nucleotide sequence with each codon falling under the amino acid designated in row 1. The mutability score calculated with the S5F model is shown below the base in each box in this row. Each box is highlighted at AID hot spots (red; mutability score>2) and cold spots (blue; mutability score <0.3). Row 8 is the estimated probability of the amino acid observed in the matured sequence (see methods for how this is calculated). The formatting pattern of rows 1-3 is repeated for the matured DH270 in rows 4-7. Amino acid substitutions are highlighted in yellow in row 4. Nucleotide mutations are shown in dark red text. Nucleotide mutations that are the result of mutations at AID cold spots are shown with an arrow below.

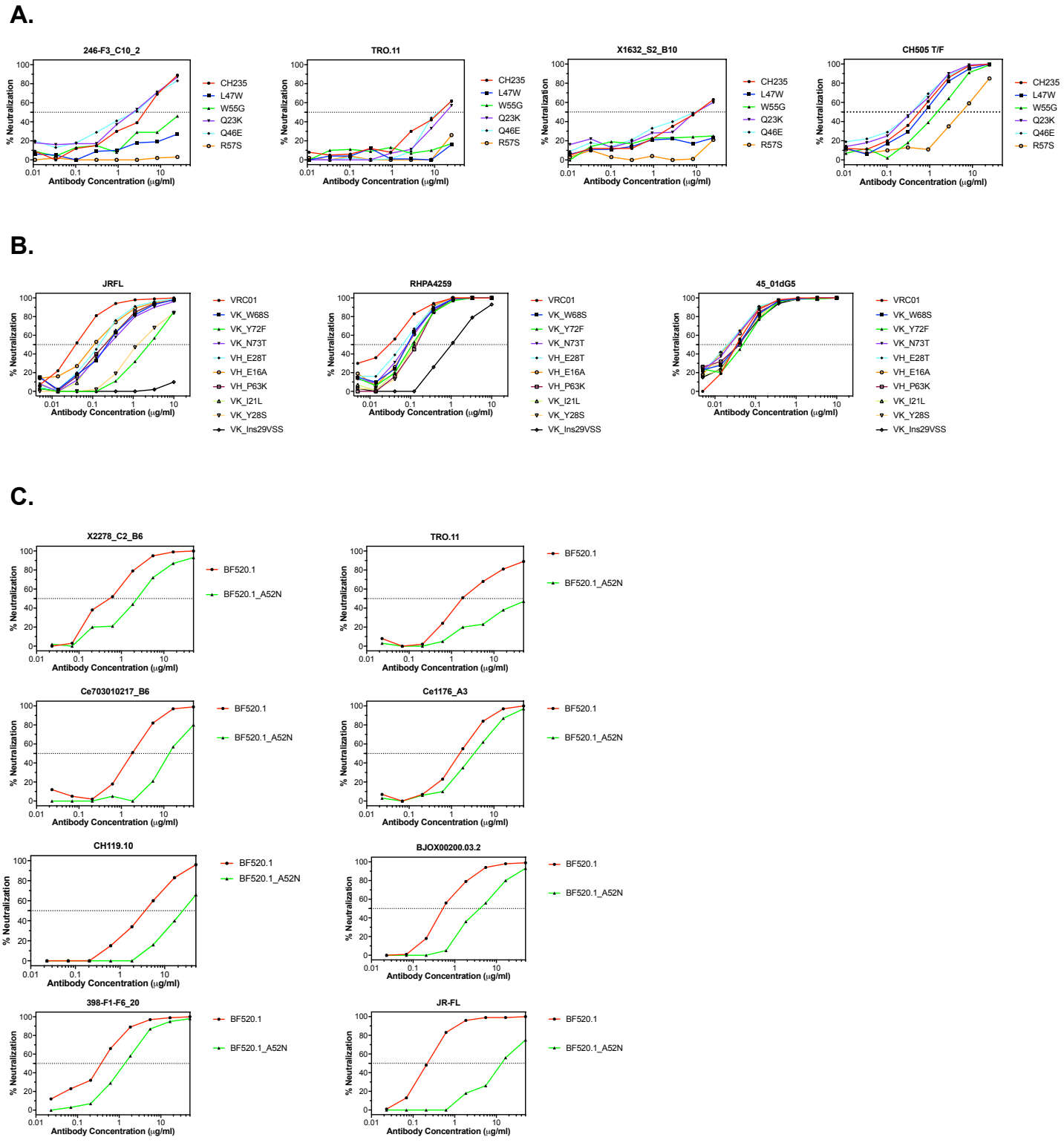


Figure S2. Neutralization of improbable mutation reversion mutants for CH235, VRC01, and BF520.1

Curves of the percent neutralization of WT (red line) **A)** CH235 **B)** VRC01 and **C)** BF520.1 and mutants containing reversions of identified improbable mutations against heterologous and autologous (CH505 T/F and 4501dG5 for CH235 and VRC01, respectively) viruses. 50% neutralization is denoted by a dotted line.

A.

```
CH235.UCA      QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYYMHWVRQAPGQGLEWMGIINPSGGSTSYAQKFQGRVTM
CH235          ................T...Q.......N..V...........QL..W.D..W.R.N...N....I..
CH236          .........A..R.....TI..R.......T..I.........R..L..M.D..R.R.D..........
CH239          ..........R......T..........NNF.V............C..W.D..V.RIN..........
CH240          ........T..R......TI.........NNF.V............C..W.D..V.RI..G.........
CH241          .........A........R............SHI..........P....M.D..V.RPTT.G........
CH235.13       ........GG..R..STTTI..V....S.ND..I............VL.F.D..N.R.N..GA.GD.FS.
CH235.12       ....A.Y.GG..RL..TMTL..V......ND..I..........F.LL.Y.D.AN.RPD..GALRE.LSF
CH235.11       .........GT..S..T..TL...T...N.ID..I....R....RP.L..Y.D..H.RPD.EG..RD.ISL
CH235.10       .........T....R...TL..RT...N.ID.FI....R....R..V..Y.D..R.RPD..PN.RD..SL
8ANC131        .G......GGL....T..TI..L..E...NEFVI..I.......PL.L.L.KR.-.RLMT.YN..D.LSL
1B2530         ....E...TA.R......TL..Q....N.VK.II.....K..L.F..V.M.D.YR.RPWS.H.....LSL

CH235.UCA      TRDTSTSTVYMELSSLRSEDTAVYYCARNVGTAGS---LLHFDYWGQGTLVTVSS---------
CH235          ............MR..............A.E........Y............A.........
CH236          S.......L....R...PD...L...V.....E........Y.....................
CH239          ...R.......G..........I...V.D...E..........H.......I..A.........
CH240          ...R.......G..........M...V.D...E..........H.......I..A.........
CH241          ...RY...A..D...............S.E.T.......Y.........I.............
CH235.13       Y..K.ME.L..D.RN....D...M...V.............Y.H..T.SKII............
CH235.12       Y..K.ME.L..D.R...YD...M...V.............Y.H..S.SP.I............
CH235.11       Y......V...DVRG..LD...L...V.GG.VEV....SN.Y.H..P..M.F..P.........
CH235.10       Y....M.I..LD.RD.TPD...I...V.SE..E.T...V..Y.H..P..R....P.........
8ANC131        R..R..G..F...RG..PD..........DGLGEVAPDYRYGI.V....ST.I.T.ASTKGPSVF
1B2530         S....MEIL..T.T..K.D...T.F...AEAASD.HSRPIM..H....S......ASTKGPSVF
```

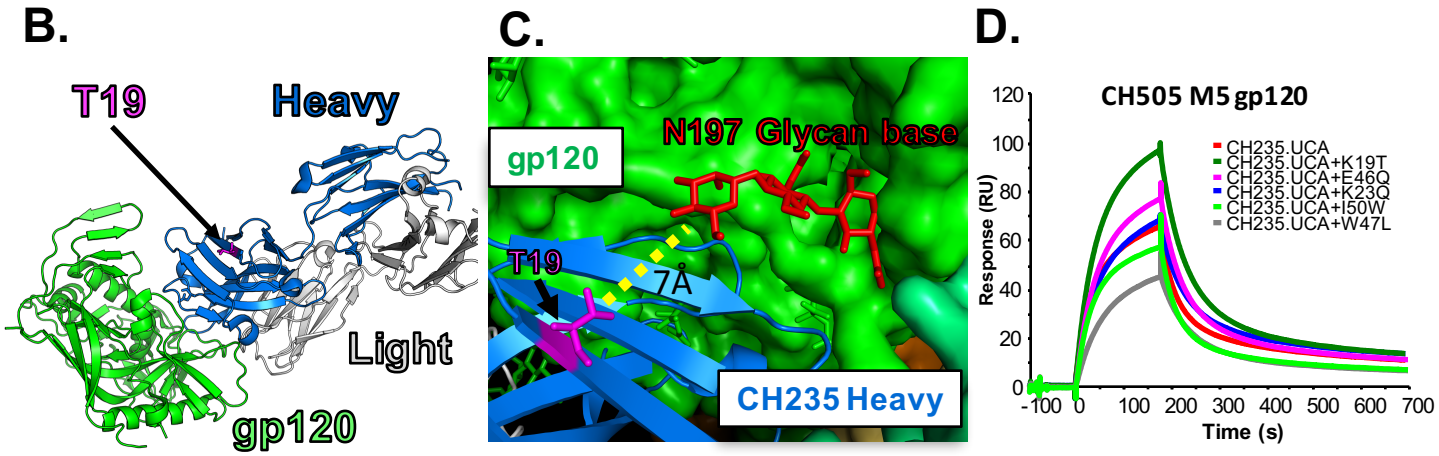


Figure S3. K19T mutation is conserved across all VH1-46 derived bnAb lineages and T19 position is proximal to N197 glycan site

A) Amino acid multiple sequence alignment of the heavy chains of the three known VH1-46 gene segment-derived CD4 binding site bnAbs: 8ANC131, 1B2530, and the multiple member CH235 lineage aligned to the CH235 UCA. The K19T mutation (red) is observed in all three lineages suggesting convergence of this mutation in three distinct individuals. Dots denote an amino acid match with the CH235 UCA in that position. **B)** The T19 position (magenta) in the CH235/gp120 complex structure (PDB: 5F9W) is outside of the CH235 (heavy chain, blue; light chain, gray) binding site. The complex structure was determined with monomeric gp120 (green) and only minimal glycosylation (not shown) was resolved. **C)** Superposition of the CH235 complex onto a fully glycosylated SOSIP trimer (5FYL) revealed that T19 (magenta) is in close proximity (7Å) to the N197 glycan base (red) resolved in the trimer structure (green). A longer Lys residue in the 19th position may sterically clash with longer glycans, providing a structural rationale for the conservation of the K19T mutation in VH1-46 derived CD4 binding site bnAbs. **D)** SPR sensorgrams for wildtype CH235 UCA and 5 UCA mutants containing improbable mutations show binding response to M5, a gp120 construct featuring a single amino acid mutation from the CH505 T/F that makes it more favorable for binding the CH235.UCA.

A

B

```
IGHV1-46*01     QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYYMHWVRQAPGQGLEWMGI
CH505.VH1-46    QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYYMHWVRQAPGQGLEWMGI

IGHV1-46*01     INPSGGSTSYAQKFQGRVTMTRDTSTSTVYMELSSLRSEDTAVYYCAR
CH505.VH1-46    INPSGGSTSYAQKFQGRVTMTRDTSTSTVYMELSSLRSEDTAVYYCAR
```

Figure S4. Genomic sequencing of the VH1-46 gene in the CH505 subject detected no allelic variation in the K19 codon confirming K19T was a mutation

A) Sequence logo plot of VH1-46 reads from genomic sequencing of the CH505 subject. No polymorphisms due to allelic variation were detected in the K19 codon (red box). Ambiguous bases in the primer region are denoted in gray. **B)** Consensus of translated VH1-46 reads from the CH505 subject showed no variation in amino acid sequence from the reference allele of IGHV1-46*01 confirming that K19T was a mutation and not due to an allelic variant.

Table S1. Neutralization of bnAbs and Mutants

		Probability‡	CH505 T/F†	DU156.12	JR-FL	Q23	TRO.11*	246-F3_C10 2*	X1632_ S2_B10*	398-F1-F6_20*	CNE8*	X2278_ C2_B6*	CNE55*	CH119.10*	BJOX002000.03.2*	25710-2.43*	Ce703010217*	Ce1176_ A3*	SVA
CH235	**CH235**	-	0.572	18.385	2.438	1.927	11.933	3.649	8.907	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	T19K	0.0116	0.888	>50	1.635	7.39	>50	6.148	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	L47W	0.00864	0.699	NT	NT	NT	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	W55G	0.00017	1.663	NT	NT	NT	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	Q23K	0.01612	0.398	NT	NT	NT	19.869	2.641	11.821	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	Q46E	0.01847	0.41	NT	NT	NT	12.142	2.178	7.125	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	R57S	0.22522	6.102	NT	NT	NT	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25	>25
	T19N	0.033	23.249	>50	20.289	>50	>50	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	>25
	T19R	0.021	0.587	35.681	2.175	3.715	40.392	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	NT	>25
	CH01+CH31	-	< 0.011	3.078	<0.023	<0.023	0.093	0.034	<0.011	0.039	0.059	0.025	0.035	0.468	0.977	0.16	< 0.011	0.018	>25

		Probability‡	JR-FL	TRO.11*	246-F3_C10 2*	X1632_S 2_B10*	398-F1-F6_20*	CNE8*	X2278_ C2_B6*	CNE55*	CH119.10*	BJOX002000.03.2*	25710-2.43*	Ce703010217*	Ce1176_ A3*	SVA
BF520.1	**BF520.1**	-	1.315	1.866	NT	>50	0.39	>50	0.467	>50	3.894	0.598	40.449	1.901	1.472	>50
	A52N	0.0005	12.81	>50	NT	>50	1.247	>50	2.238	>50	22.203	3.971	>50	13.288	3.634	>50
	CH01+CH31	-	0.031	0.38	NT	0.078	0.108	0.183	0.055	0.113	2.29	10.806	0.895	0.102	0.11	>50

		Probability‡	JR-FL	45_01dG5†	RHPA4259	SVA
VRC01	**VRC01**	-	0.04	0.036	0.033	>10
	VK_W68S	0.0007	0.215	0.038	0.089	>10
	VK_Y72F	0.009	2.095	0.053	0.109	>10
	VK_N73T	0.0085	0.261	0.025	0.088	>10
	VH_E28T	0.0009	0.141	0.019	0.069	>10
	VH_E16A	0.0012	0.12	0.022	0.092	>10
	VH_P63K	0.007	0.209	0.038	0.133	>10
	VK_I21L	0.0095	0.196	0.042	0.121	>10
	VK_Y28S	0.001	1.387	0.042	0.128	>10
	VK_Ins29VSS	ND	> 10	0.042	1.054	>10
	CH01+CH31	-	< 0.023	< 0.023	0.136	>10

Neutralization (IC50 µg/ml)		
50 to 1	1 to .1	<0.1

*From global panel of 12 Tier-2 reference strains
† Autologous virus
‡ Estimated probability of mutation prior to selection as calculated by the ARMADiLLO program
NT Not Tested

Fig. S5

A.

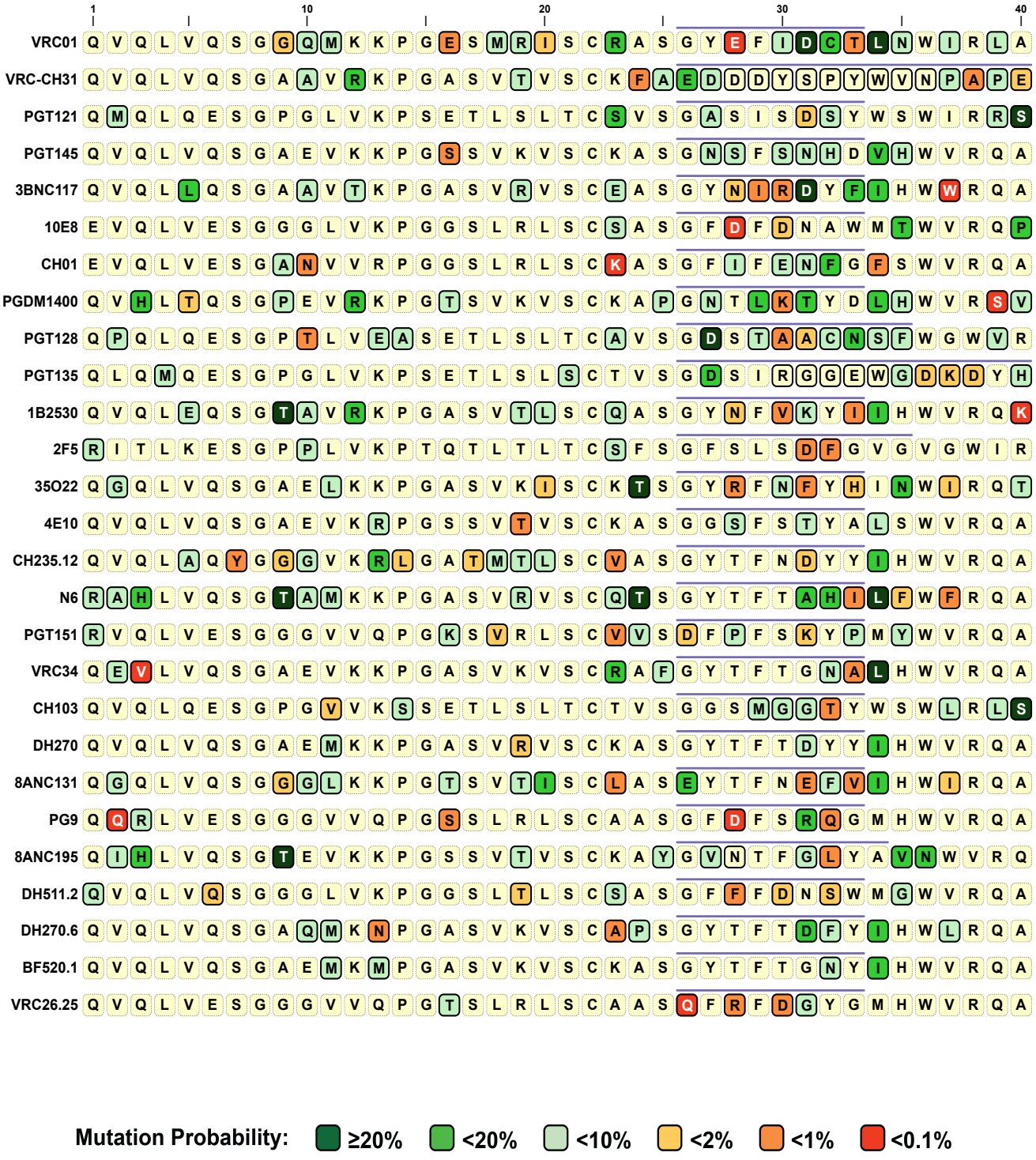

1
10
20
30
40
VRC01 Q V Q L V Q S G G Q M K K P G E S M R I S C R A S G Y E F I D C T L N W I R L A
VRC-CH31 Q V Q L V Q S G A A V R K P G A S V T V S C K F A E D D D Y S P Y W V N P A P E
PGT121 Q M Q L Q E S G P G L V K P S E T L S L T C S V S G A S I S D S Y W S W I R R S
PGT145 Q V Q L V Q S G A E V K K P G S S V K V S C K A S G N S F S N H D V H W V R Q A
3BNC117 Q V Q L L Q S G A A V T K P G A S V R V S C E A S G Y N I R D Y F I H W W R Q A
10E8 E V Q L V E S G G G L V K P G G S L R L S C S A S G F D F D N A W M T W V R Q P
CH01 E V Q L V E S G A N V V R P G G S L R L S C K A S G F I F E N F G F S W V R Q A
PGDM1400 Q V H L T Q S G P E V R K P G T S V K V S C K A P G N T L K T Y D L H W V R S V
PGT128 Q P Q L Q E S G P T L V E A S E T L S L T C A V S G D S T A A C N S F W G W V R
PGT135 Q L Q M Q E S G P G L V K P S E T L S L S C T V S G D S I R G G E W G D K D Y H
1B2530 Q V Q L E Q S G T A V R K P G A S V T L S C Q A S G Y N F V K Y I I H W V R Q K
2F5 R I T L K E S G P P L V K P T Q T L T L T C S F S G F S L S D F G V G V G W I R
35O22 Q G Q L V Q S G A E L K K P G A S V K I S C K T S G Y R F N F Y H I N W I R Q T
4E10 Q V Q L V Q S G A E V K R P G S S V T V S C K A S G G S F S T Y A L S W V R Q A
CH235.12 Q V Q L A Q Y G G G V K R L G A T M T L S C V A S G Y T F N D Y Y I H W V R Q A
N6 R A H L V Q S G T A M K K P G A S V R V S C Q T S G Y T F T A H I L F W F R Q A
PGT151 R V Q L V E S G G G V V Q P G K S V R L S C V V S D F P F S K Y P M Y W V R Q A
VRC34 Q E V L V Q S G A E V K K P G A S V K V S C R A F G Y T F T G N A L H W V R Q A
CH103 Q V Q L Q E S G P G V V K S S E T L S L T C T V S G G S M G G T Y W S W L R L S
DH270 Q V Q L V Q S G A E M K K P G A S V R V S C K A S G Y T F T D Y Y I H W V R Q A
8ANC131 Q G Q L V Q S G G G L K K P G T S V T I S C L A S E Y T F N E F V I H W I R Q A
PG9 Q Q R L V E S G G G V V Q P G S S L R L S C A A S G F D F S R Q G M H W V R Q A
8ANC195 Q I H L V Q S G T E V K K P G S S V T V S C K A Y G V N T F G L Y A V N W V R Q
DH511.2 Q V Q L V Q S G G G L V K P G G S L T L S C S A S G F F F D N S W M G W V R Q A
DH270.6 Q V Q L V Q S G A Q M K N P G A S V K V S C A P S G Y T F T D F Y I H W L R Q A
BF520.1 Q V Q L V Q S G A E M K M P G A S V K V S C K A S G Y T F T G N Y I H W V R Q A
VRC26.25 Q V Q L V E S G G G V V Q P G T S L R L S C A A S Q F R F D G Y G M H W V R Q A
Mutation Probability: ≥20% <20% <10% <2% <1% <0.1%

Fig. S5 (cont.)

A.

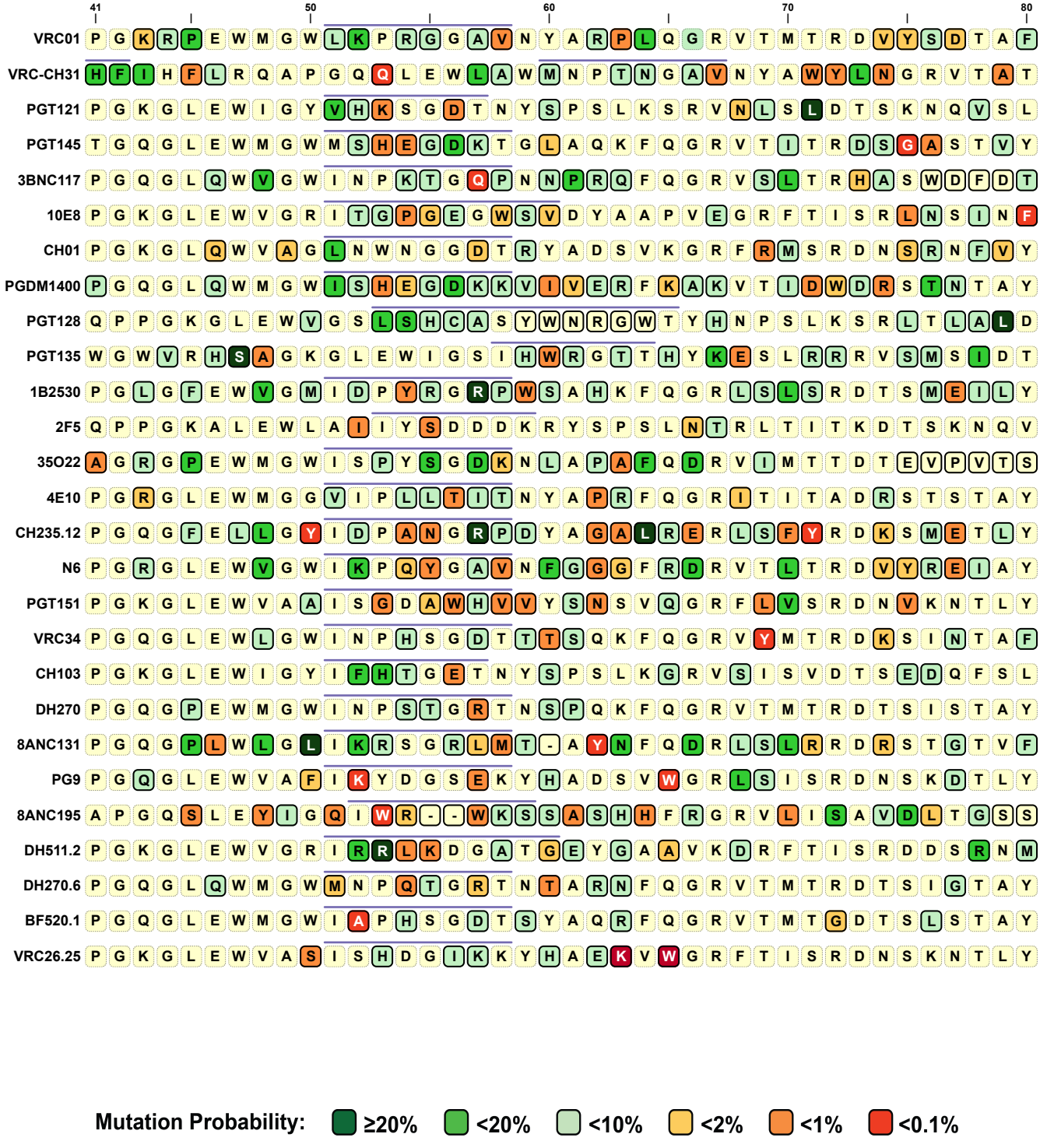

41
50
60
70
80
VRC01 P G K R P E W M G W L K P R G G A V N Y A R P L Q G R V T M T R D V Y S D T A F
VRC-CH31 H F I H F L R Q A P G Q Q L E W L A W M N P T N G A V N Y A W Y L N G R V T A T
PGT121 P G K G L E W I G Y V H K S G D T N Y S P S L K S R V N L S L D T S K N Q V S L
PGT145 T G Q G L E W M G W M S H E G D K T G L A Q K F Q G R V T I T R D S G A S T V Y
3BNC117 P G Q G L Q W V G W I N P K T G Q P N N P R Q F Q G R V S L T R H A S W D F D T
10E8 P G K G L E W V G R I T G P G E G W S V D Y A A P V E G R F T I S R L N S I N F
CH01 P G K G L Q W V A G L N W N G G D T R Y A D S V K G R F R M S R D N S R N F V Y
PGDM1400 P G Q G L Q W M G W I S H E G D K K V I V E R F K A K V T I D W D R S T N T A Y
PGT128 Q P P G K G L E W V G S L S H C A S Y W N R G W T Y H N P S L K S R L T L A L D
PGT135 W G W V R H S A G K G L E W I G S I H W R G T T H Y K E S L R R R V S M S I D T
1B2530 P G L G F E W V G M I D P Y R G R P W S A H K F Q G R L S L S R D T S M E I L Y
2F5 Q P P G K A L E W L A I I Y S D D D K R Y S P S L N T R L T I T K D T S K N Q V
35O22 A G R G P E W M G W I S P Y S G D K N L A P A F Q D R V I M T T D T E V P V T S
4E10 P G R G L E W M G G V I P L L T I T N Y A P R F Q G R I T I T A D R S T S T A Y
CH235.12 P G Q G F E L L G Y I D P A N G R P D Y A G A L R E R L S F Y R D K S M E T L Y
N6 P G R G L E W V G W I K P Q Y G A V N F G G G F R D R V T L T R D V Y R E I A Y
PGT151 P G K G L E W V A A I S G D A W H V V Y S N S V Q G R F L V S R D N V K N T L Y
VRC34 P G Q G L E W L G W I N P H S G D T T T S Q K F Q G R V Y M T R D K S I N T A F
CH103 P G K G L E W I G Y I F H T G E T N Y S P S L K G R V S I S V D T S E D Q F S L
DH270 P G Q G P E W M G W I N P S T G R T N S P Q K F Q G R V T M T R D T S I S T A Y
8ANC131 P G Q G P L W L G L I K R S G R L M T - A Y N F Q D R L S L R R D R S T G T V F
PG9 P G Q G L E W V A F I K Y D G S E K Y H A D S V W G R L S I S R D N S K D T L Y
8ANC195 A P G Q S L E Y I G Q I W R - - W K S S A S H H F R G R V L I S A V D L T G S S
DH511.2 P G K G L E W V G R I R R L K D G A T G E Y G A A V K D R F T I S R D D S R N M
DH270.6 P G Q G L Q W M G W M N P Q T G R T N T A R N F Q G R V T M T R D T S I G T A Y
BF520.1 P G Q G L E W M G W I A P H S G D T S Y A Q R F Q G R V T M T G D T S L S T A Y
VRC26.25 P G K G L E W V A S I S H D G I K K Y H A E K V W G R F T I S R D N S K N T L Y
Mutation Probability:
≥20%
<20%
<10%
<2%
<1%
<0.1%

Fig. S5 (cont.)

A.

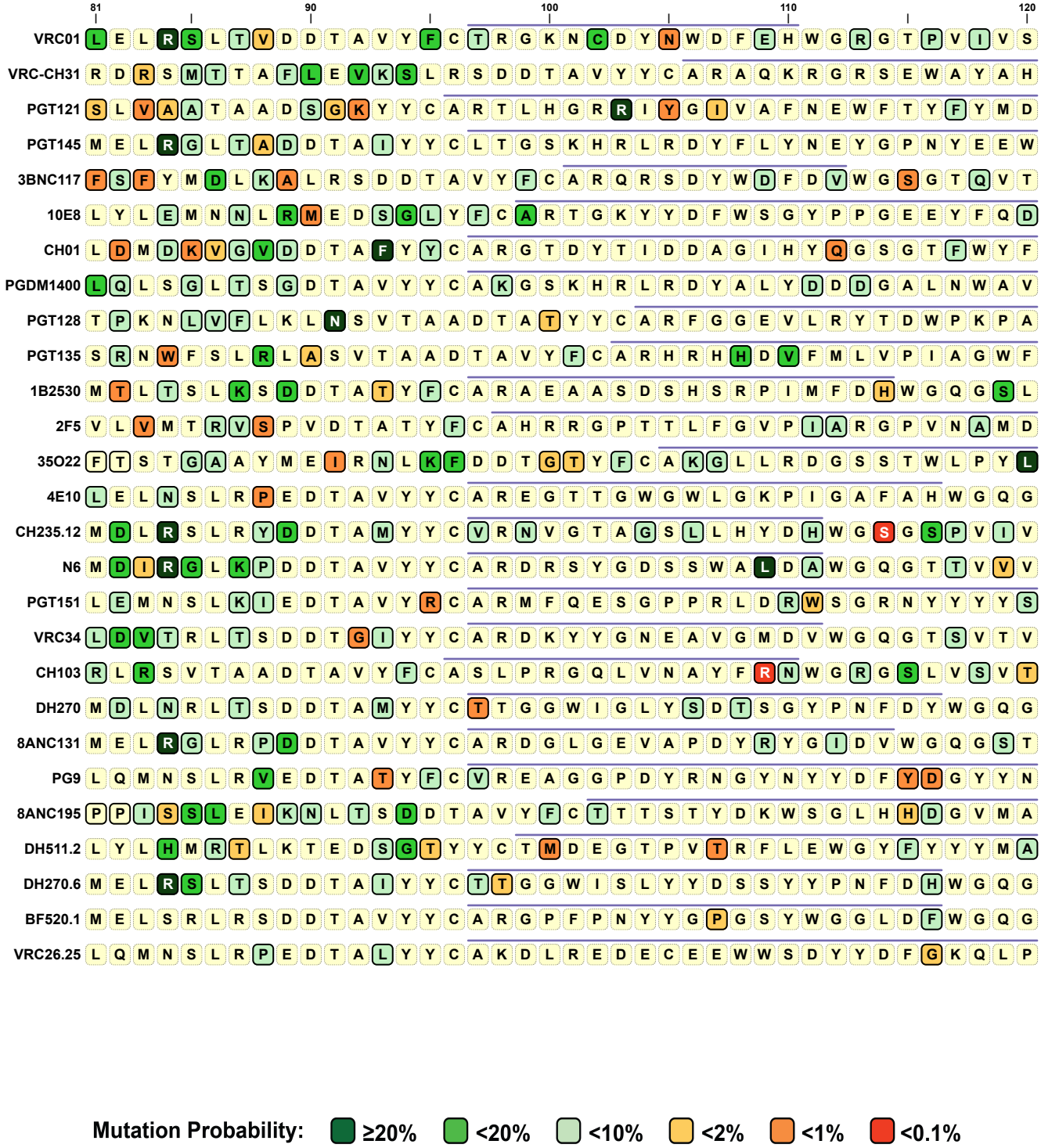

81
90
100
110
120
VRC01 L E L R S L T V D D T A V Y F C T R G K N C D Y N W D F E H W G R G T P V I V S
VRC-CH31 R D R S M T T A F L E V K S L R S D D T A V Y Y C A R A Q K R G R S E W A Y A H
PGT121 S L V A A T A A D S G K Y Y C A R T L H G R R I Y G I V A F N E W F T Y F Y M D
PGT145 M E L R G L T A D D T A I Y Y C L T G S K H R L R D Y F L Y N E Y G P N Y E E W
3BNC117 F S F Y M D L K A L R S D D T A V Y F C A R Q R S D Y W D F D V W G S G T Q V T
10E8 L Y L E M N N L R M E D S G L Y F C A R T G K Y Y D F W S G Y P P G E E Y F Q D
CH01 L D M D K V G V D D T A F Y Y C A R G T D Y T I D D A G I H Y Q G S G T F W Y F
PGDM1400 L Q L S G L T S G D T A V Y Y C A K G S K H R L R D Y A L Y D D D G A L N W A V
PGT128 T P K N L V F L K L N S V T A A D T A T Y Y C A R F G G E V L R Y T D W P K P A
PGT135 S R N W F S L R L A S V T A A D T A V Y F C A R H R H H D V F M L V P I A G W F
1B2530 M T L T S L K S D D T A T Y F C A R A E A A S D S H S R P I M F D H W G Q G S L
2F5 V L V M T R V S P V D T A T Y F C A H R R G P T T L F G V P I A R G P V N A M D
35O22 F T S T G A A Y M E I R N L K F D D T G T Y F C A K G L L R D G S S T W L P Y L
4E10 L E L N S L R P E D T A V Y Y C A R E G T T G W G W L G K P I G A F A H W G Q G
CH235.12 M D L R S L R Y D D T A M Y Y C V R N V G T A G S L L H Y D H W G S G S P V I V
N6 M D I R G L K P D D T A V Y Y C A R D R S Y G D S S W A L D A W G Q G T T V V V
PGT151 L E M N S L K I E D T A V Y R C A R M F Q E S G P P R L D R W S G R N Y Y Y Y S
VRC34 L D V T R L T S D D T G I Y Y C A R D K Y Y G N E A V G M D V W G Q G T S V T V
CH103 R L R S V T A A D T A V Y F C A S L P R G Q L V N A Y F R N W G R G S L V S V T
DH270 M D L N R L T S D D T A M Y Y C T T G G W I G L Y S D T S G Y P N F D Y W G Q G
8ANC131 M E L R G L R P D D T A V Y Y C A R D G L G E V A P D Y R Y G I D V W G Q G S T
PG9 L Q M N S L R V E D T A T Y F C V R E A G G P D Y R N G Y N Y Y D F Y D G Y Y N
8ANC195 P P I S S L E I K N L T S D D T A V Y F C T T T S T Y D K W S G L H H D G V M A
DH511.2 L Y L H M R T L K T E D S G T Y Y C T M D E G T P V T R F L E W G Y F Y Y Y M A
DH270.6 M E L R S L T S D D T A I Y Y C T T G G W I S L Y Y D S S Y Y P N F D H W G Q G
BF520.1 M E L S R L R S D D T A V Y Y C A R G P F P N Y Y G P G S Y W G G L D F W G Q G
VRC26.25 L Q M N S L R P E D T A L Y Y C A K D L R E D E C E E W W S D Y Y D F G K Q L P
Mutation Probability:
≥20%
<20%
<10%
<2%
<1%
<0.1%

Fig. S5 (cont.)

A.

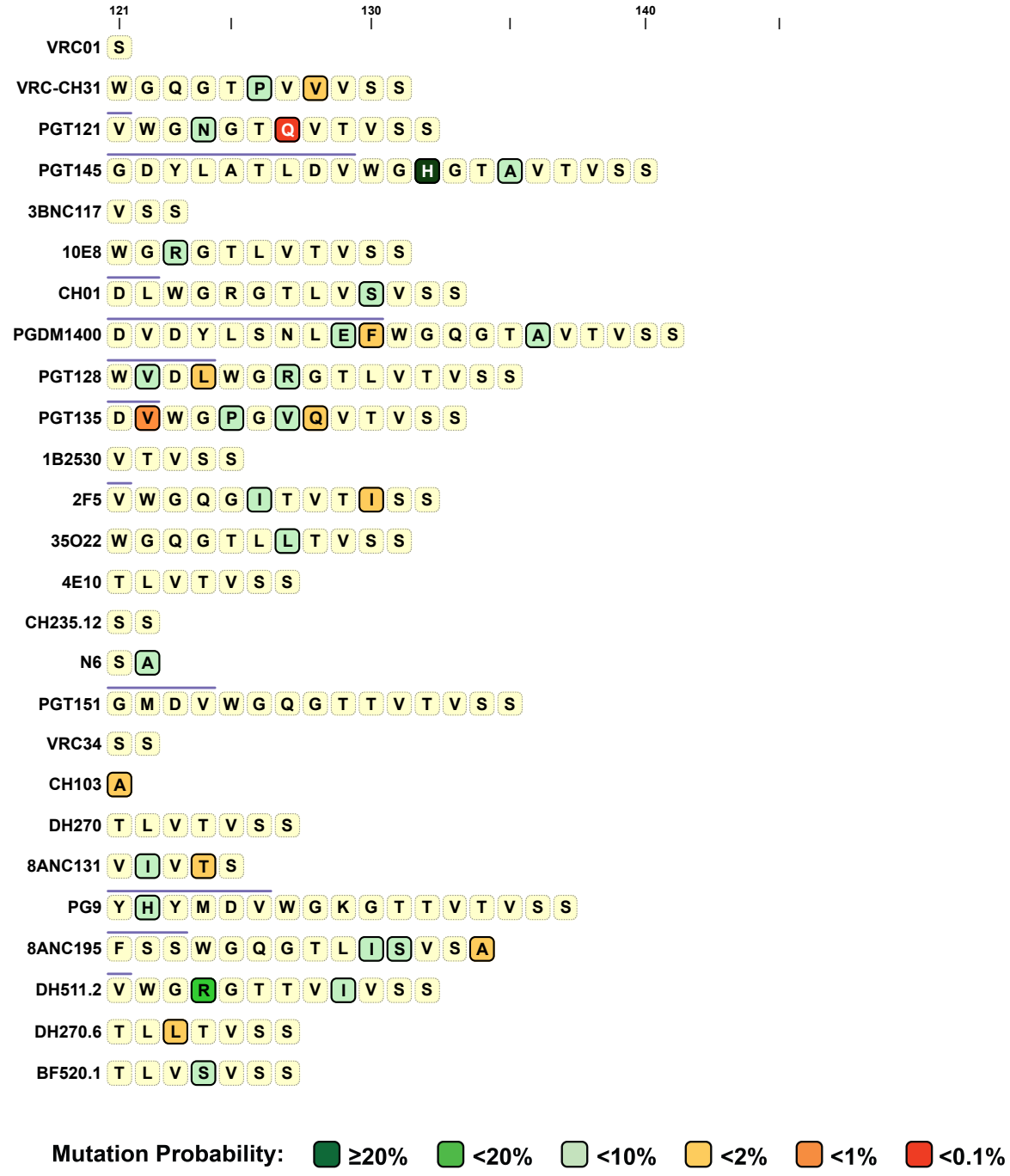
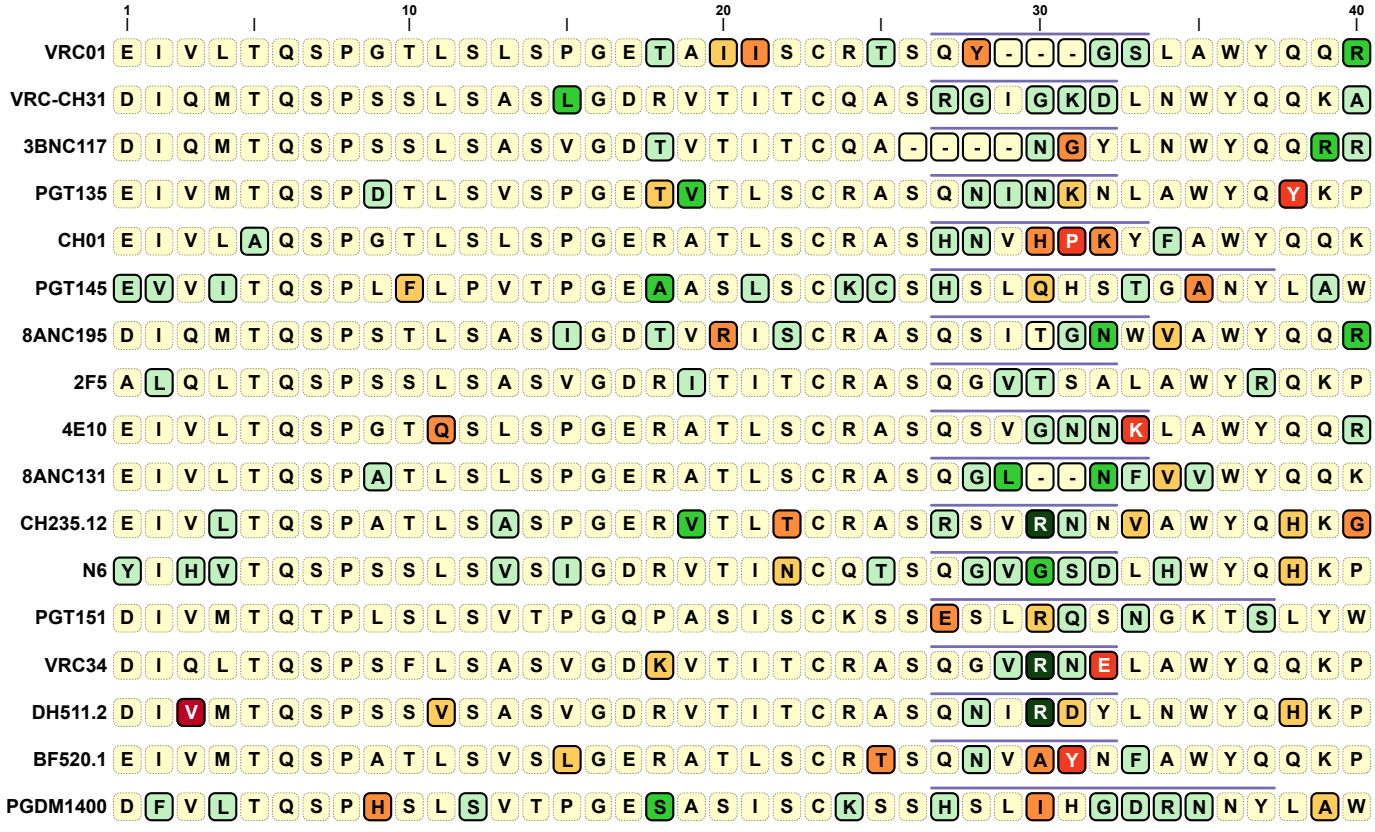


Figure S5. Representative bnAb sequences colored by mutation probability

A) Heavy chain sequences for a representative set of bnAbs are highlighted by their mutation probability as estimated by ARMADiLLO. Mutations that are expected to occur frequently in the absence of selection (high probability mutations) are colored in shades of green (see legend). Mutations that are expected to occur rarely in the absence of selection (improbable mutations) are colored in shades of red (see legend). Amino acids residing in CDRs are denoted with a dark blue line above them. UCA inference was performed with only the observed bnAb sequence as input and as such there may be substantial uncertainty in mutation calls within the CDR3s.

Fig. S5 (cont.)

B.


1 10 20 30 40
VRC01 E I V L T Q S P G T L S L S P G E T A I I S C R T S Q Y - - - G S L A W Y Q Q R
VRC-CH31 D I Q M T Q S P S S L S A S L G D R V T I T C Q A S R G I G K D L N W Y Q Q K A
3BNC117 D I Q M T Q S P S S L S A S V G D T V T I T C Q A - - - - N G Y L N W Y Q Q R R
PGT135 E I V M T Q S P D T L S V S P G E T V T L S C R A S Q N I N K N L A W Y Q Y K P
CH01 E I V L A Q S P G T L S L S P G E R A T L S C R A S H N V H P K Y F A W Y Q Q K
PGT145 E V V I T Q S P L F L P V T P G E A A S L S C K C S H S L Q H S T G A N Y L A W
8ANC195 D I Q M T Q S P S T L S A S I G D T V R I S C R A S Q S I T G N W V A W Y Q Q R
2F5 A L Q L T Q S P S S L S A S V G D R I T I T C R A S Q G V T S A L A W Y R Q K P
4E10 E I V L T Q S P G T Q S L S P G E R A T L S C R A S Q S V G N N K L A W Y Q Q R
8ANC131 E I V L T Q S P A T L S L S P G E R A T L S C R A S Q G L - - N F V V W Y Q Q K
CH235.12 E I V L T Q S P A T L S A S P G E R V T L T C R A S R S V R N N V A W Y Q H K G
N6 Y I H V T Q S P S S L S V S I G D R V T I N C Q T S Q G V G S D L H W Y Q H K P
PGT151 D I V M T Q T P L S L S V T P G Q P A S I S C K S S E S L R Q S N G K T S L Y W
VRC34 D I Q L T Q S P S F L S A S V G D K V T I T C R A S Q G V R N E L A W Y Q Q K P
DH511.2 D I V M T Q S P S S V S A S V G D R V T I T C R A S Q N I R D Y L N W Y Q H K P
BF520.1 E I V M T Q S P A T L S V S L G E R A T L S C R T S Q N V A Y N F A W Y Q Q K P
PGDM1400 D F V L T Q S P H S L S V T P G E S A S I S C K S S H S L I H G D R N N Y L A W

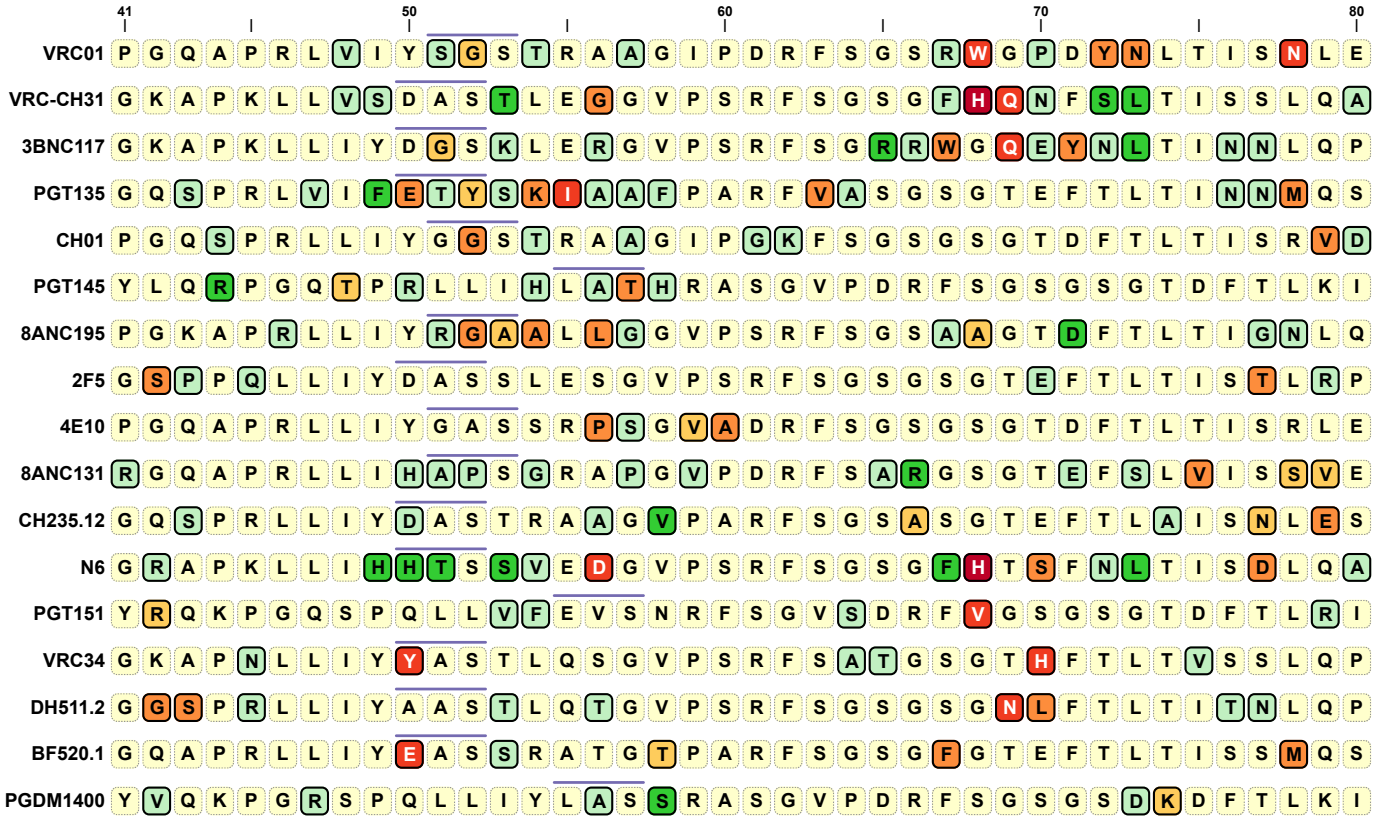

41 50 60 70 80
VRC01 P G Q A P R L V I Y S G S T R A A G I P D R F S G S R W G P D Y N L T I S N L E
VRC-CH31 G K A P K L L V S D A S T L E G G V P S R F S G S G F H Q N F S L T I S S L Q A
3BNC117 G K A P K L L I Y D G S K L E R G V P S R F S G R R W G Q E Y N L T I N N L Q P
PGT135 G Q S P R L V I F E T Y S K I A A F P A R F V A S G S G T E F T L T I N N M Q S
CH01 P G Q S P R L L I Y G G S T R A A G I P G K F S G S G S G T D F T L T I S R V D
PGT145 Y L Q R P G Q T P R L L I H L A T H R A S G V P D R F S G S G S G T D F T L K I
8ANC195 P G K A P R L L I Y R G A A L L G G V P S R F S G S A A G T D F T L T I G N L Q
2F5 G S P P Q L L I Y D A S S L E S G V P S R F S G S G S G T E F T L T I S T L R P
4E10 P G Q A P R L L I Y G A S S R P S G V A D R F S G S G S G T D F T L T I S R L E
8ANC131 R G Q A P R L L I H A P S G R A P G V P D R F S A R G S G T E F S L V I S S V E
CH235.12 G Q S P R L L I Y D A S T R A A G V P A R F S G S A S G T E F T L A I S N L E S
N6 G R A P K L L I H H T S S V E D G V P S R F S G S G F H T S F N L T I S D L Q A
PGT151 Y R Q K P G Q S P Q L L V F E V S N R F S G V S D R F V G S G S G T D F T L R I
VRC34 G K A P N L L I Y Y A S T L Q S G V P S R F S A T G S G T H F T L T V S S L Q P
DH511.2 G G S P R L L I Y A A S T L Q T G V P S R F S G S G S G S G N L F T L T I T N L Q P
BF520.1 G Q A P R L L I Y E A S S R A T G T P A R F S G S G F G T E F T L T I S S M Q S
PGDM1400 Y V Q K P G R S P Q L L I Y L A S S R A S G V P D R F S G S G S G S D K D F T L K I

Fig. S5 (cont.)

B.

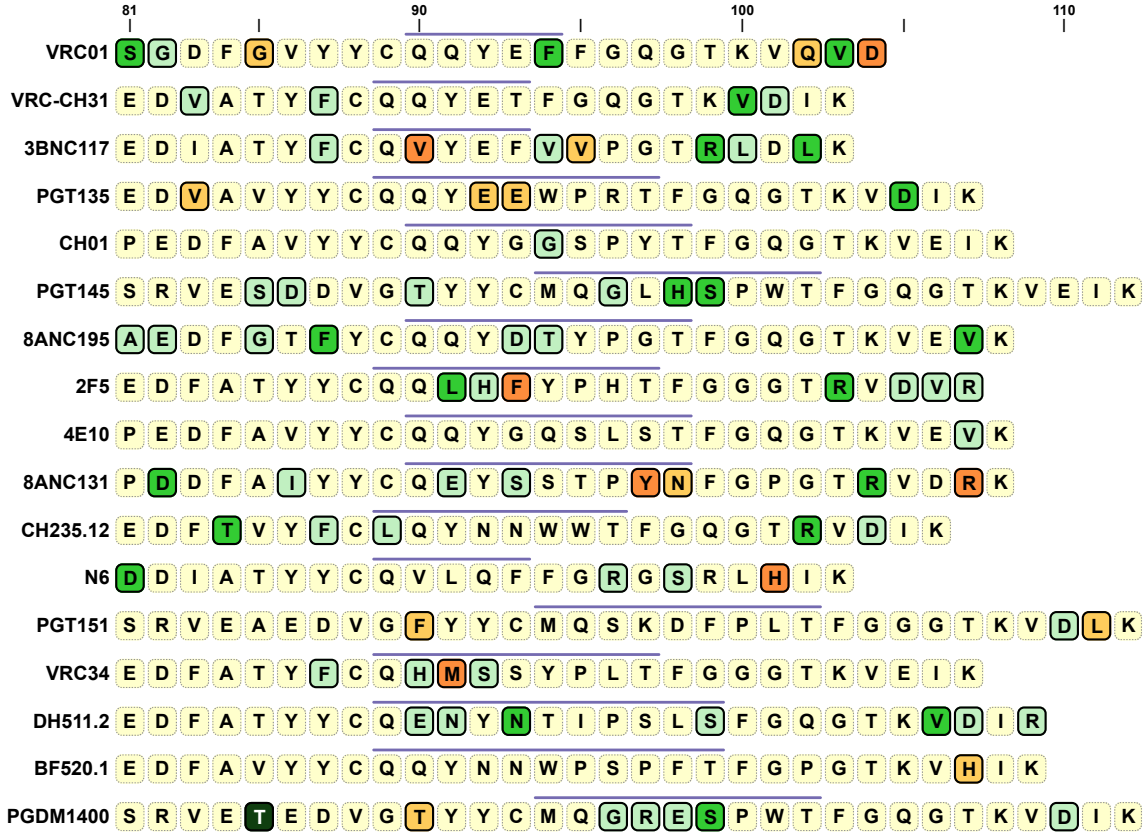


Figure S5 (cont). Representative bnAb sequences colored by mutation probability

B) Kappa chain sequences for a representative set of bnAbs are highlighted by their mutation probability as estimated by ARMADiLLO. Mutations that are expected to occur frequently in the absence of selection (high probability mutations) are colored in shades of green (see legend). Mutations that are expected to occur rarely in the absence of selection (improbable mutations) are colored in shades of red (see legend). Amino acids residing in CDRs are denoted with a dark blue line above them. UCA inference was performed with only the observed bnAb sequence as input and as such there may be substantial uncertainty in mutation calls within the CDR3s.

Fig. S5 (cont.)

C.

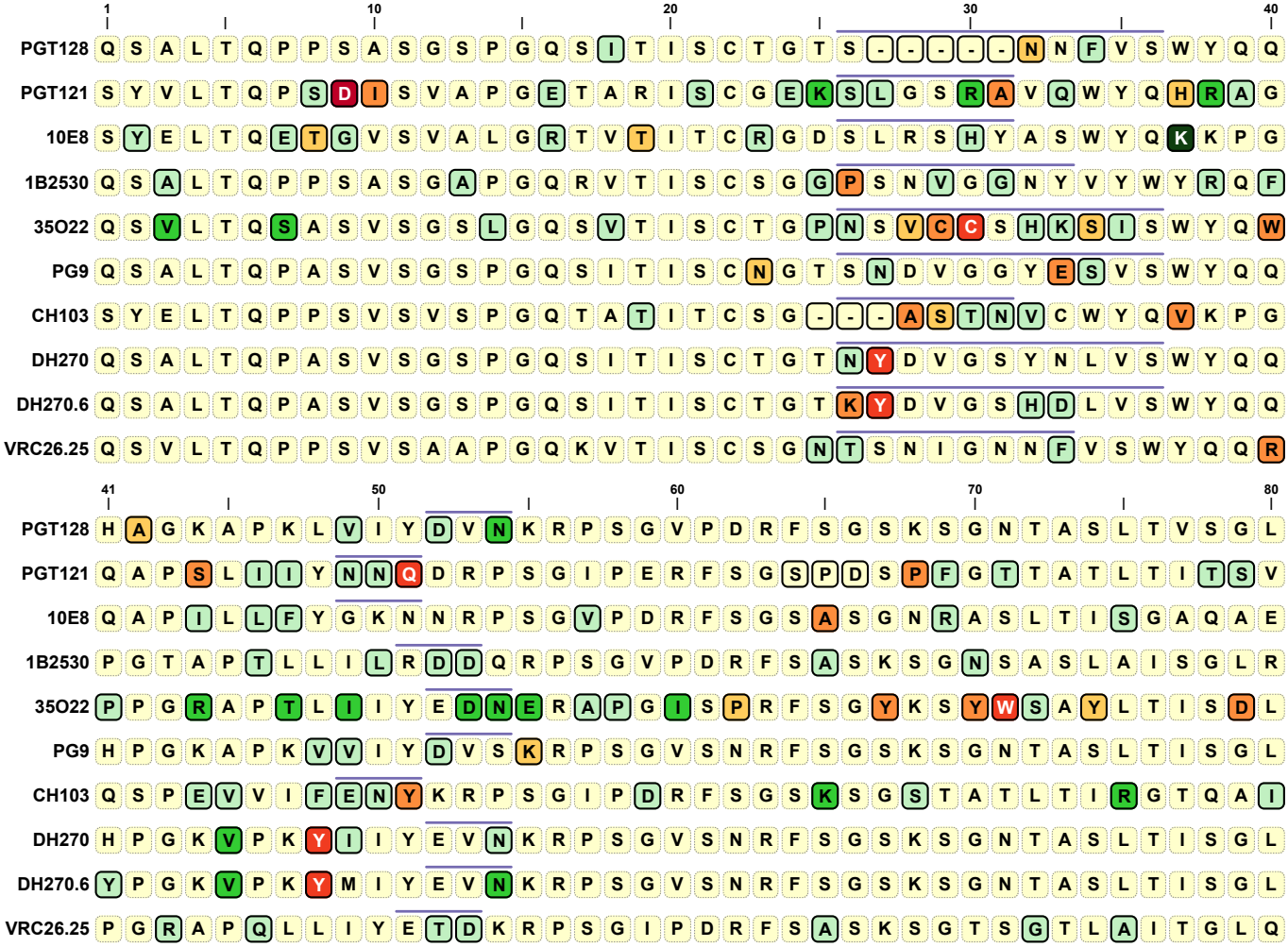

1 10 20 30 40
PGT128 Q S A L T Q P P S A S G S P G Q S I T I S C T G T S - - - - - N N F V S W Y Q Q
PGT121 S Y V L T Q P S D I S V A P G E T A R I S C G E K S L G S R A V Q W Y Q H R A G
10E8 S Y E L T Q E T G V S V A L G R T V T I T C R G D S L R S H Y A S W Y Q K K P G
1B2530 Q S A L T Q P P S A S G A P G Q R V T I S C S G G P S N V G G N Y V Y W Y R Q F
35O22 Q S V L T Q S A S V S G S L G Q S V T I S C T G P N S V C C S H K S I S W Y Q W
PG9 Q S A L T Q P A S V S G S P G Q S I T I S C N G T S N D V G G Y E S V S W Y Q Q
CH103 S Y E L T Q P P S V S V S P G Q T A T I T C S G - - - A S T N V C W Y Q V K P G
DH270 Q S A L T Q P A S V S G S P G Q S I T I S C T G T N Y D V G S Y N L V S W Y Q Q
DH270.6 Q S A L T Q P A S V S G S P G Q S I T I S C T G T K Y D V G S H D L V S W Y Q Q
VRC26.25 Q S V L T Q P P S V S A A P G Q K V T I S C S G N T S N I G N N F V S W Y Q Q R
41 50 60 70 80
PGT128 H A G K A P K L V I Y D V N K R P S G V P D R F S G S K S G N T A S L T V S G L
PGT121 Q A P S L I I Y N N Q D R P S G I P E R F S G S P D S P F G T T A T L T I T S V
10E8 Q A P I L L F Y G K N N R P S G V P D R F S G S A S G N R A S L T I S G A Q A E
1B2530 P G T A P T L L I L R D D Q R P S G V P D R F S A S K S G N S A S L A I S G L R
35O22 P P G R A P T L I I Y E D N E R A P G I S P R F S G Y K S Y W S A Y L T I S D L
PG9 H P G K A P K V V I Y D V S K R P S G V S N R F S G S K S G N T A S L T I S G L
CH103 Q S P E V V I F E N Y K R P S G I P D R F S G S K S G S T A T L T I R G T Q A I
DH270 H P G K V P K Y I I Y E V N K R P S G V S N R F S G S K S G N T A S L T I S G L
DH270.6 Y P G K V P K Y M I Y E V N K R P S G V S N R F S G S K S G N T A S L T I S G L
VRC26.25 P G R A P Q L L I Y E T D K R P S G I P D R F S A S K S G T S G T L A I T G L Q

Fig. S5 (cont.)

C.

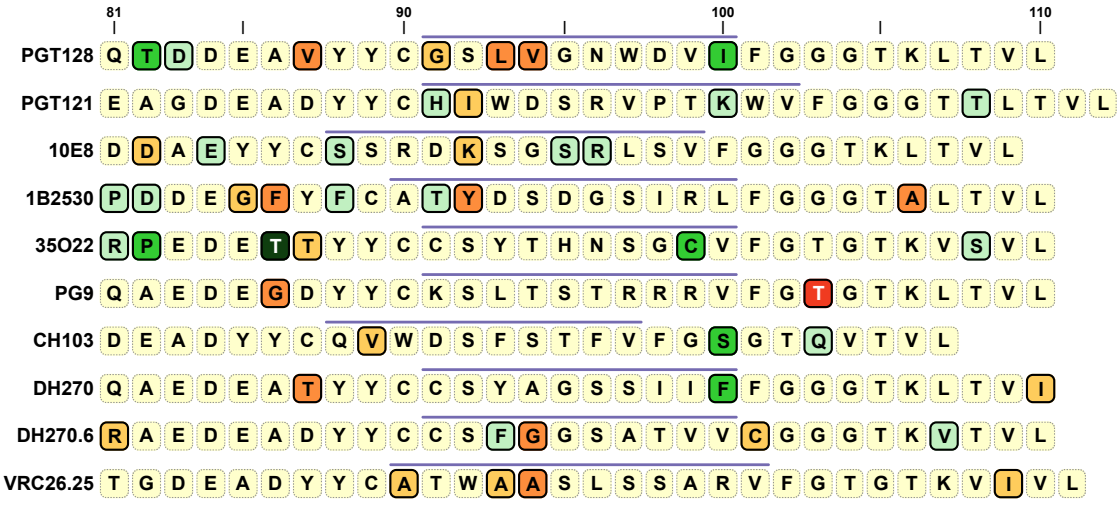


Figure S5. Representative bnAb sequences colored by mutation probability

C) Lambda chain sequences for a representative set of bnAbs are highlighted by their mutation probability as estimated by ARMADiLLO. Mutations that are expected to occur frequently in the absence of selection (high probability mutations) are colored in shades of green (see legend). Mutations that are expected to occur rarely in the absence of selection (improbable mutations) are colored in shades of red (see legend). Amino acids residing in CDRs are denoted with a dark blue line above them. UCA inference was performed with only the observed bnAb sequence as input and as such there may be substantial uncertainty in mutation calls within the CDR3s.

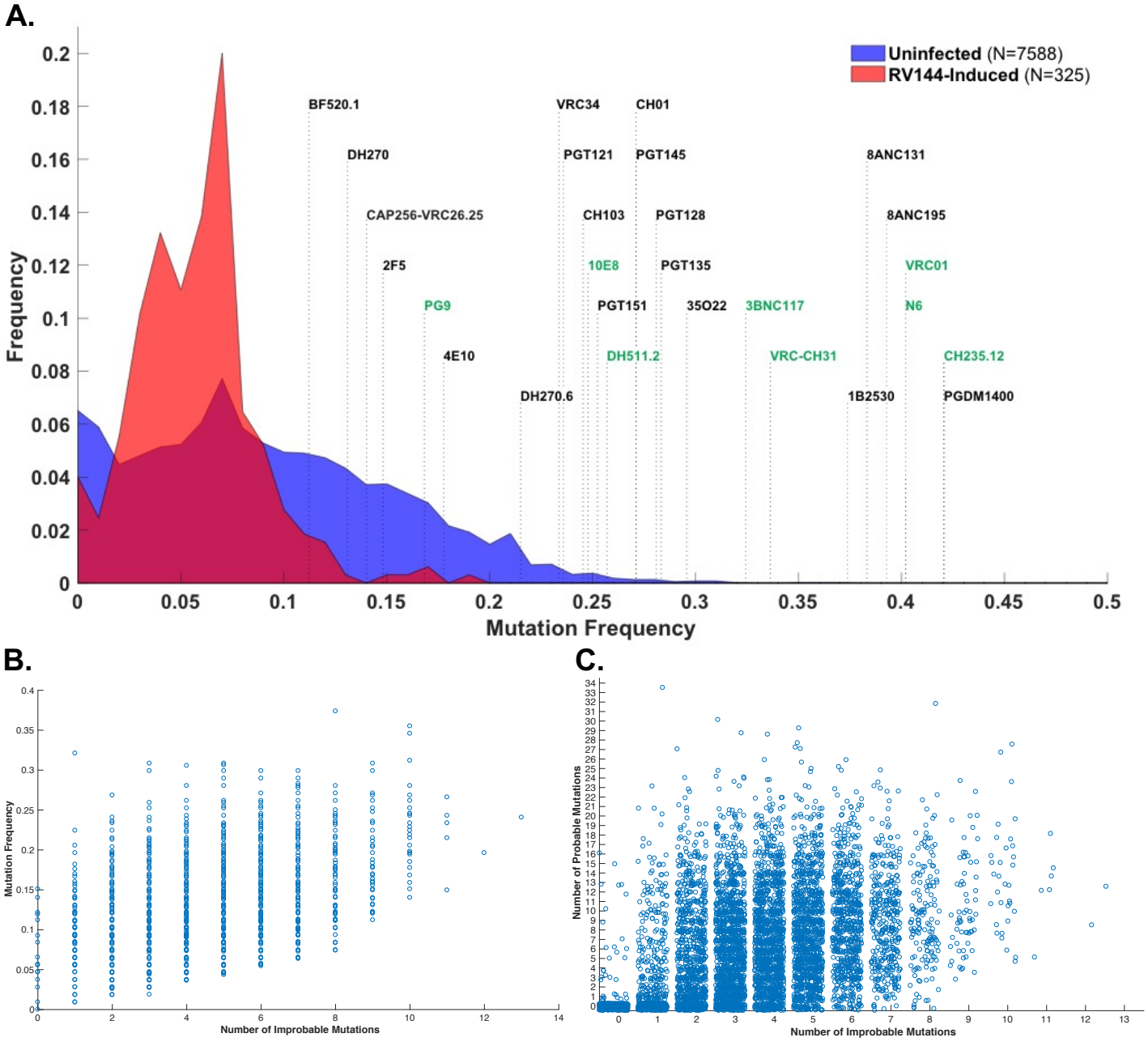


Figure S6. BnAbs have high mutation frequencies and mutation frequency is correlated with improbable mutations

A) Histograms for the distributions of number of improbable mutations (A) and mutation frequency (B) from antibody heavy chain sequences from three groups: "RV144-induced" antibodies were isolated from RV144 vaccinated subjects by antigenically sorting with RV144 immunogens (red shaded area); "Uninfected" antibodies correspond to duplicated NGS reads from IgG antibodies isolated from PBMC samples from 8 HIV-uninfected individuals (blue shaded area; see methods for details on sampling); a representative set of published bnAb antibody sequences are shown labeled above dotted lines that correspond to their mutation frequency (defined as total number of amino acid mutations in non-CDRH3 VDJ sequence divided by non-CDRH3 VDJ sequence length). Scatterplots of **B)** number of improbable mutations versus amino acid mutation frequency for 7588 NGS reads from uninfected IgG antibodies from PBMC samples from 8 HIV-uninfected individuals and **C)** number of improbable mutations versus number of probable mutations (≥2%). Number of improbable mutations was moderately correlated with number of probable mutations (Pearson's r=0.43). A stronger correlation was observed between improbable mutations and mutation frequency (Pearson's r=0.67) as expected because probable mutations are a subset of the total amino acid mutations used to calculate amino acid mutation frequency. Jitter added in order to alleviate over-plotting in panel C.

Table S3. Number of improbable mutations expected by chance given mutation frequency

bnAb	#AA mutations[A]	AA mut freq	# Improbable Mutations (M)[B]	Expected Frequency of >=M[C]	Counts
CH235.12*	45	42.1%	16	0.20%	2/985
8ANC195*	44	39.3%	13	2.34%	23/981
N6*	43	40.2%	12	4.35%	43/989
VRC01*	43	40.2%	12	3.75%	37/986
8ANC131*	41	38.3%	12	3.67%	36/980
VRC-CH31*	39	33.6%	12	1.43%	14/981
CH01*	29	27.1%	11	0.71%	7/985
DH511.2*	28	25.7%	11	0.51%	5/981
PGT151*	27	25.2%	11	0.62%	6/974
35O22*	34	29.6%	11	1.32%	13/984
3BNC117	36	32.4%	10	5.28%	52/985
10E8*	27	24.8%	9	4.54%	45/991
PGDM1400	45	42.1%	9	36.59%	360/984
1B2530	40	37.4%	9	25.43%	249/979
PG9*	18	16.8%	9	1.63%	16/983
PGT135	32	28.3%	8	15.96%	157/984
PGT121	25	23.6%	8	8.61%	85/987
2F5*	16	14.8%	8	4.47%	44/984
PGT145	29	27.1%	7	25.96%	256/986
VRC34	25	23.4%	6	32.21%	315/978
VRC26.25	15	14.0%	6	23.71%	234/987
DH270.6	23	21.5%	6	29.67%	292/984
CH103	26	24.5%	5	56.58%	559/988
4E10	19	17.8%	5	46.97%	465/990
PGT128	32	28.1%	4	82.36%	808/981
BF520.1	12	11.2%	3	84.37%	831/985

A Number of amino acid mutations, excluding CDRH3 region

B The number of improbable mutations M (at <2% cutoff) identified for the BNAb using ARMADiLLO

C The expected frequency of M or greater improbable mutations (at <2% cutoff) from simulations of SHM at the BNAb's AA mutation frequency

* Number of improbable mutations in this bnAb exceeds number expected by chance (cutoff=.05) given the mutation frequency of this bnAb

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Simulating the somatic hypermutation process

Because AID targets hot spots according to their underlying sequence motifs, the probability of mutations is sequence context dependent, making an analytical computation of the probability of a mutation in the absence of selection all but intractable. Instead, we take a numerical approach via simulation. Here, we estimate the probability of an amino acid substitution by simulating the somatic hypermutation (SHM) process and calculating the observed frequency of that substitution in the simulated sequences. The simulation proceeds as follows. Given a matured antibody nucleotide sequence, we first infer its unmutated common ancestor (UCA) sequence (Kepler, 2013; Kepler et al., 2014). Next, the matured antibody nucleotide sequence is aligned to the UCA nucleotide sequence and the number of sites mutated, t, is computed. Starting with the UCA sequence, 1) the mutability score of all consecutive sequence pentamers is computed according to the S5F mutability model (Yaari et al., 2013). 2) The mutability scores for each base position in the sequence are converted into the probability distribution, Q, by:

$$Q_i = \frac{C_i}{\sum_{i=1}^{L} C_i} \qquad [1]$$

where C_i is the mutability score at position i and L is the length of the sequence. 3) A base position, b, is drawn randomly according to Q. 4) The nucleotide n, at b, is substituted according to the S5F substitution model (Yaari et al., 2013), resulting in sequence S_j where j is the number of mutations accrued during the simulation. The procedure then iterates over steps 1-4 until j=t. This results in a simulated sequence, S_t, that has acquired the same number of nucleotide mutations as observed in the matured antibody sequence of interest. If at any iteration during the simulation a mutation results in a stop codon, that sequence is discarded and the process restarts from the UCA sequence. This simulation procedure is then repeated to generate 100,000 simulated matured sequences. These nucleotide sequences are then translated to amino acid sequences.

Estimating the probability of an amino acid substitution

The estimate of the probability of any amino acid substitution $U \rightarrow$Y at site i given the number of mutations t observed in the matured sequence of interest is then calculated as the amino acid frequency observed at site i in the simulated sequences according to:

$$\widehat{\mathrm{P}}(X_{i_{U\rightarrow Y}}|UCA, t) = \frac{1}{N}\sum_{j=0}^{N} 1(X_{ij} = Y) \qquad [2]$$

where X_i is the amino acid at site i which has the amino acid U in the UCA sequence mutating to amino acid Y in the matured sequence of interest, UCA is the UCA sequence, N is the number

of simulated sequences, 1 is an indicator function for observing amino acid Y at site i in the jth simulated sequence. This estimate is for an amino acid substitution *in the absence of selection* and we use this probability as a gauge of how likely it is that a B cell would arise to have this mutation prior to antigenic selection. Amino acid substitutions that are the result of mutations that occur in AID hot spots will have high probabilities, occur frequently and a subset of the reservoir of B cell clonal members would likely have these mutations present prior to antigenic selection. Amino substitutions that are the result of cold spot mutations or require multiple base substitutions will be much less frequent and could represent significant hurdles to lineage development and these substitutions may require strong antigenic selection to be acquired during B cell maturation.

Improbable mutations

The probability of a specific amino substitution at any given position is the product of two components. The first component is due to the bias of the AID enzyme in targeting that specific base position and the DNA repair mechanisms preference for substituting to an alternative base. Practically speaking, substitutions that require mutations at AID cold spots and/or result in disfavored base substitutions by DNA repair mechanisms are infrequent and thus improbable. The second component is the number and length of available paths through codon space to go from an amino acid encoded by the codon in the UCA to that of the codon for the substituted amino acid in the matured sequence. To illustrate this, we turn to a practical example: the TAT codon which encodes the amino acid, Tyr. From the TAT codon, 5 amino acids are achievable by a single nucleotide base substitution (C,D,F,H,N,S), 12 amino acids by two base substitutions (A,E,G,I,K,L,P,Q,R,T,V,W) and 1 amino acid (M) by three base substitutions. Without considering the bias of AID, the Y->M mutation starting from the TAT codon is inherently unlikely to occur because it requires three independent mutational events to occur within the same codon. By simulating the SHM process, ARMADiLLO captures the interplay of these two components and is able to estimate the probability of any amino acid substitution prior to selection by taking both components into account.

For this study, we have selected a cutoff of less than 2% probability to classify mutations as “improbable”. We chose this cutoff to reflect a frequency in which the expected number of mutations in a B cell clone in a single germinal center would be less than or equal to 1. In order to calculate the expected number of mutations, we used estimates of the number of clonally related B cells in a germinal center. The range of these estimates is ~10-100 clonally related B cells for immunizations in mice with various protein antigens (Jacob et al., 1991; Tas et al., 2016). Thus, for a mutation that has 2% probability, the expected number of B cells with this mutation in a germinal center is 1 in 50, reflective of one mutation per clone per germinal center. To demonstrate the effects of different choices of cutoffs, we applied additional cutoffs of 1%, 0.1% and 0.01% when calculating the estimated numbers of improbable mutations in a representative set of bnAbs and these additional data are included in an Excel spreadsheet in the supplementary materials.

Calculating the expected number of improbable mutations

The number of improbable amino acid mutations, M, in an antibody sequence at a given probability cutoff can be estimated by applying [2] and enumerating over the entire amino acid sequence. For example, CH235.12 is estimated to have M=16 improbable mutations in its heavy chain when improbable mutations are defined as amino acid substitutions with <2% estimated probability. We estimate the probability of getting M improbable mutations or greater at a given amino acid mutation frequency, u, from the empirical distribution of the number of improbable mutations observed in sequences simulated to acquire T amino acid mutations, where T=u*L and L is the length of the sequence. To calculate the empirical distribution of improbable mutations for each antibody sequence of interest, we first randomly draw 1000 sequences from an antibody sequence dataset generated from NGS sequencing of 8 HIV-1 negative individuals and infer the UCA of each sequence (Kepler, 2013; Kepler et al., 2014). From these randomly sampled UCAs, we then simulate the SHM process using the same simulation procedure as detailed above and stop the simulation when each sequence acquires T amino acid mutations. This results in a set of 1000 simulated sequences each with an amino acid mutation frequency of u. The probability of observing M or greater improbable mutations in the absence of selection is then:

$$\mathrm{P}(\mathrm{X} \geq \mathrm{M}) = \frac{1}{N}\sum_{j=0}^{N} 1(X_j \geq M) \qquad [3]$$

where N is the number of simulations (here N=1000), X_j is the number of improbable mutations in the jth simulated sequence (calculated from [2] over all amino acid positions in the sequence) and 1 is an indicator function. Here we exclude the CDR3 sequence from our calculations of both M and u as the inference of the UCA has widely varying levels of uncertainty in the CDR3 region depending on the input matured sequence.

Standard methods for determining selection at an amino acid site typically rely on the measure ω which is the ratio of non-synonymous mutations to synonymous mutations at that position in a multiple sequence alignment of related gene sequences. Here, we avoid this measure of selection for two reasons. In many instances in this study we have only two sequences to compare, the UCA and the matured sequence. This does not provide the number of observations needed for ω to reliably indicate selection. In some case, where we do have multiple clonal members to align, the number of mutational events at a site is also not sufficiently large enough for ω to be reliable. Secondly, ω is calculated under the assumption that non-synonymous mutations are of neutral fitness advantage. Clearly, due to the sequence dependence of AID targeting this assumption is violated in B cell evolution. Instead, we employ the heuristic that amino acid mutations that are estimated to be improbable yet occur frequently within a clone are likely to have been selected for. While indicative of selection, this too can be misleading if mutations occur early in a lineage, are neutral and generate a cold spot or *colder* spot, thus making it less likely for the position to mutate again. Thus, it is apparent that much work remains on developing rigorous methods for measuring selection in B cell evolution. Our approach here is to treat improbable amino acid mutations as candidates for

selection and to ultimately confirm the fitness advantage conferred by such mutations through experimentally testing their effect on virus neutralization and antigen binding.

Antibody sequences from HIV-1 negative subjects

We utilized a previously described next generation sequencing dataset generated from 8 HIV-1 negative individuals prior to vaccination (Williams et al., 2015). Briefly, to mitigate error introduced during the PCR amplification, we split the RNA sample into two samples, A and B, and performed PCR amplification on each, independently. Only VDJ sequences that duplicated identically in A and B were then retained. This approach allowed us to be highly confident that nucleotide variations from germline gene segments that occurred in the NGS reads were mutations and not error introduced during PCR. We refer to this dataset as "uninfected".

Antibody sequences from RV144-vaccinated subjects

We utilized a previously described set of antibody sequences (Easterhoff et al., 2017) isolated from subjects enrolled in the RV144 HIV-1 vaccination trial (Rerks-Ngarm et al., 2009). Antibody sequences were isolated from peripheral blood mononuclear cells (PBMC) from 7 RV144-vaccinated subjects that were antigen-specific single-cell sorted with fluorophore-labeled AE.A244 gp120 d11 (Liao et al., 2013). We refer to this dataset as "RV144-immunized".

Analysis of Improbable Mutations in BnAbs

Sequences of HIV-1 bnAbs were obtained either from NCBI GenBank or from the bNAber database(Eroshkin et al., 2014). For the comparison of improbable mutations for the representative set of bnAbs, improbable mutations were calculated using the ARMADiLLO program described above. UCAs were inferred using Cloanalyst (Kepler, 2013; Kepler et al., 2014). While many bnAbs had multiple clonal lineage member sequences available, some bnAbs had no other members isolated. Because of this, only the single sequence of the matured bnAb was used in the UCA inference in order to provide equal treatment of all sequences. Because uncertainty in the UCA inference is highest for the bases in the CDR3 region, precise determination of some mutations in this region is not feasible and we therefore ignored the CDR3 region in our analysis of the representative set of bnAbs. In the simulations, we prohibited any mutations from occurring in the CDR3 region by setting the probability of AID targeting to 0 for each base in the CDR3. Neutralization data for the bnAbs was obtained through the CATNAP database(Yoon et al., 2015) and corresponds to neutralization in the global panel of 12 HIV-1 Env reference strains(deCamp et al., 2014). For the calculation of geometric mean neutralization, undetectable neutralization was set to 100 μg/ml. Breadth was reported for all viruses that were tested and for several bnAbs (8ANC131, 1B2530, N6, CH103, BF520.1, PGT135, PGT145, VRC26.25, PGDM1400) neutralization data was not available for all 12 viruses in the global panel.

Genomic sequencing of the VH1-46 gene segment

To confirm that K19T was a mutation and not an allelic variant in subject CH505 from which CH235 clonal members were isolated, we sequenced CH505 IGHV gene segments according to a previously described experimental protocol for high throughput genomic sequencing of Ig gene segments which is detailed in Scheepers et al (Scheepers et al., 2015). Sequencing was performed using the Illumina MiSeq platform. A custom sequence analysis pipeline was used to analyze the sequencing data for identifying novel alleles. Forward and reverse reads were merged using FLASh (Magoc and Salzberg, 2011) and quality filtered using the FASTx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Primers were trimmed such that only the V region of the merged read was retained and all resulting reads were then de-duplicated. All sequences with fewer than 10 reads were discarded. We then aligned all sequences to all known IGHV gene segments using a custom semi-global pairwise alignment program. Sequences that matched closest to any reference VH1-46 allele (01-03) were retained. Sequences that included stop codons, reflective of potential pseudogenes, were discarded. We then built a multiple sequence alignment of the remaining sequences and produced a sequence logo plot weighted by the number of read copies for each sequence. For the 19th codon, 96% of reads matched the VH1-46*01 reference AAG triplet encoding a lysine, consistent with homozygous lysine at this position (the remaining 4% of reads are consistent with the error level introduced during PCR amplification). The sequencing thus confirms K19T in the CH235 clone was indeed a mutation in the CH505 subject. A consensus of the translated reads was identical to the entire VH1-46*01 reference allele demonstrating no non-synonymous polymorphisms indicating that at the protein level the CH505 subject was homozygous VH1-46*01.

Antibody Site-directed Mutagenesis Primers

Primer	Chain	Sequence
CH235_L47W	Heavy	gatccatcccatccattgaagcccctgtccag
CH235_W55G	Heavy	gtgcgacccccactagggtcgatccatccc
CH235_Q23K	Heavy	agtgacggtttcctgcaaggcatctggatacac
CH235_Q46E	Heavy	gatccatcccatcaactcaagcccctgtccaggg
CH235_R57S	Heavy	cgaccctagttggggtagcacaaactacgca
CH235_T19K	Heavy	gcctggggcctcagtgaaggtttcctgc
CH235_T19R	Heavy	caggaaaccctcactgaggccccagg
CH235_T19N	Heavy	tgcctggcaggaaacattcactgaggccccag
VRC01_E28T	Heavy	gaatccaatttagcgtacaatcaataaacgtatatccagaagcccgacaagaaattctc
VRC01_P63K	Heavy	gccgtcaactacgcacgtaaacttcagggcagagt
VRC01_E16A	Heavy	caagaaattctcatcgacgcgccaggcttcttcatc
VRC01_Y28S	Kappa	accaggctaaggaaccactctgactggtccgacaag
VRC01_Y72F	Kappa	ctgatggtgagattgaagtctggcccccacc
VRC01_W68S	Kappa	tcagcggcagtcggtcggggccag
VRC01_N73T	Kappa	gtgggggccagactacactctcaccatcagc
VRC01_I21L	Kappa	tggtccgacaagagaggatggctgtttcccc

SUPPLEMENTAL REFERENCES

deCamp, A., Hraber, P., Bailer, R.T., Seaman, M.S., Ochsenbauer, C., Kappes, J., Gottardo, R., Edlefsen, P., Self, S., Tang, H.*, et al.* (2014). Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. J Virol *88*, 2489-2507.

Easterhoff, D., Moody, M.A., Fera, D., Cheng, H., Ackerman, M., Wiehe, K., Saunders, K.O., Pollara, J., Vandergrift, N., Parks, R.*, et al.* (2017). Boosting of HIV envelope CD4 binding site antibodies with long variable heavy third complementarity determining region in the randomized double blind RV305 HIV-1 vaccine trial. PLoS Pathog *13*, e1006182.

Eroshkin, A.M., LeBlanc, A., Weekes, D., Post, K., Li, Z., Rajput, A., Butera, S.T., Burton, D.R., and Godzik, A. (2014). bNAber: database of broadly neutralizing HIV antibodies. Nucleic Acids Res *42*, D1133-1139.

Jacob, J., Kelsoe, G., Rajewsky, K., and Weiss, U. (1991). Intraclonal generation of antibody mutants in germinal centres. Nature *354*, 389-392.

Kepler, T.B. (2013). Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. F1000Res *2*, 103.

Kepler, T.B., Munshaw, S., Wiehe, K., Zhang, R., Yu, J.S., Woods, C.W., Denny, T.N., Tomaras, G.D., Alam, S.M., Moody, M.A.*, et al.* (2014). Reconstructing a B-Cell Clonal Lineage. II. Mutation, Selection, and Affinity Maturation. Front Immunol *5*, 170.

Liao, H.X., Bonsignori, M., Alam, S.M., McLellan, J.S., Tomaras, G.D., Moody, M.A., Kozink, D.M., Hwang, K.K., Chen, X., Tsao, C.Y.*, et al.* (2013). Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. Immunity *38*, 176-186.

Magoc, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957-2963.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E.*, et al.* (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med *361*, 2209-2220.

Scheepers, C., Shrestha, R.K., Lambson, B.E., Jackson, K.J., Wright, I.A., Naicker, D., Goosen, M., Berrie, L., Ismail, A., Garrett, N.*, et al.* (2015). Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. J Immunol *194*, 4371-4378.

Tas, J.M., Mesin, L., Pasqual, G., Targ, S., Jacobsen, J.T., Mano, Y.M., Chen, C.S., Weill, J.C., Reynaud, C.A., Browne, E.P.*, et al.* (2016). Visualizing antibody affinity maturation in germinal centers. Science *351*, 1048-1054.

Williams, W.B., Liao, H.X., Moody, M.A., Kepler, T.B., Alam, S.M., Gao, F., Wiehe, K., Trama, A.M., Jones, K., Zhang, R.*, et al.* (2015). HIV-1 VACCINES. Diversion of HIV-1 vaccine-induced immunity by gp41-microbiota cross-reactive antibodies. Science *349*, aab1253.

Yaari, G., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Gupta, N., Stern, J.N., O'Connor, K.C., Hafler, D.A., Laserson, U., Vigneault, F., and Kleinstein, S.H. (2013). Models of somatic hypermutation targeting and substitution based on synonymous mutations from high-throughput immunoglobulin sequencing data. Front Immunol *4*, 358.

Yoon, H., Macke, J., West, A.P., Jr., Foley, B., Bjorkman, P.J., Korber, B., and Yusim, K. (2015). CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. Nucleic Acids Res *43*, W213-219.