Determining probabilities of HIV-1 bNAb development in healthy and chronically infected individuals

Christoph Kreer,[1†] Cosimo Lupo,[2†‡] Meryem S. Ercanoglu,[1†] Lutz Gieselmann,[1,3] Natanael Spisak,[2] Jan Grossbach,[4] Maike Schlotz,[1] Philipp Schommers,[1,5] Henning Gruell,[1,5] Leona Dold,[6,7] Andreas Beyer,[4,8] Armita Nourmohammad,[9,10,11] Thierry Mora,[2§] Aleksandra M. Walczak,[2§] and Florian Klein[1,3,8§*]

[1]Laboratory of Experimental Immunology, Institute of Virology, Faculty of Medicine and University Hospital Cologne, University of Cologne, 50931 Cologne, Germany.

[2]Laboratoire de physique de l'Ecole normale supérieure, CNRS, PSL University, Sorbonne Université, and Université Paris Cité, 75005 Paris, France.

[3]German Center for Infection Research, Partner Site Bonn-Cologne, 50931 Cologne, Germany.

[4]Excellence Cluster on Cellular Stress Responses in Aging Associated Diseases, University of Cologne, 50931 Cologne, Germany.

[5]Department I of Internal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, 50937 Cologne, Germany.

[6]Department of Internal Medicine I, University Hospital of Bonn, Bonn, Germany.

[7]German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Bonn, Germany.

[8]Center for Molecular Medicine Cologne (CMMC), University of Cologne, 50931 Cologne, Germany.

[9]Max Planck Institute for Dynamics and Self-Organization, Am Faßberg 17, 37077 Göttingen, Germany.

[10]Department of Physics, University of Washington, 3910 15th Ave Northeast, Seattle, WA 98195, USA.

[11]Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109, USA.

*Corresponding author. Email: florian.klein@uk-koeln.de

† These authors contributed equally to this work.

‡ Present address: Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Roma I, 00185 Rome, Italy.

§ Shared senior authorship.

Abstract

HIV-1 broadly neutralizing antibodies (bNAbs) are able to suppress viremia and prevent infection. Their induction by vaccination is therefore a major goal. However, in contrast to antibodies that neutralize other pathogens, HIV-1-specific bNAbs frequently carry uncommon molecular characteristics that might prevent their induction. Here, we performed unbiased sequence analyses of B cell receptor repertoires from 57 healthy and 46 chronically HIV-1- or HCV-infected individuals and learned probabilistic models to predict the likelihood of bNAb development. We show that lower probabilities for bNAbs are predictive of higher HIV-1 neutralization activity. Moreover, ranking of bNAbs by their probabilities allowed to identify highly potent antibodies with superior generation probabilities as preferential targets for vaccination approaches. Importantly, we found equal bNAb probabilities across infected and healthy donors. This implies that chronic infection is not a prerequisite for the generation of bNAbs, fostering the hope that HIV-1 vaccines can induce bNAb development in healthy individuals.

INTRODUCTION

Each day, the human body is confronted with numerous pathogens that need to be eliminated to prevent severe diseases. The adaptive immune system is able to cope with the plethora of different antigenic structures by employing a diverse repertoire of lymphocytes, each of which expresses its own unique immune receptor as a consequence of V(D)J recombination during lymphopoiesis *(1)*. B cell receptors (BCRs) further diversify during the process of affinity maturation by somatic hypermutation (SHM) *(2, 3)*. Both processes lead to the generation of antibodies with high affinity that can target and neutralize infectious pathogens.

Despite the diversity and flexibility of the BCR repertoire, some pathogens can evolve quickly and outpace the adaptive immune system. The human immunodeficiency virus-1 (HIV-1), for example, proliferates into antigenically diverse quasispecies within a short timeframe due to its error-prone replication machinery *(4)*. These quasispecies contain viral variants that can escape from autologous circulating antibodies. As a consequence, the immune system adapts to the emerged variants, resembling an ongoing immunological arms race *(5)*. Notably, there is a rare fraction of HIV-1-infected individuals who develop a broad serum neutralization response against numerous viral variants, and from whom monoclonal broadly neutralizing antibodies (bNAbs) could have been isolated (reviewed in *(6–9)*). Although the autologous virus is typically resistant against the hosts bNAbs, they can be used to efficiently prevent and treat heterologous HIV-1 as well as chimeric simian-(S)HIV-1 infections in animal models *(10–15)*, or to suppress viremia in HIV-1-infected individuals without notable adverse events or side effects *(16–25)*. Recent prevention studies with the bNAb VRC01 could not demonstrate a higher overall protection against HIV-1 acquisition in humans *(26)*. However, subgroup analyses of VRC01-sensitive HIV-1 strains revealed that infections with VRC01-sensitive HIV-1 strains were significantly reduced. This translates into an estimated prevention efficacy of about 75% in this subgroup and is the first proof of concept that bNAbs are in principle able to prevent infections in humans *(27)*. Importantly, technical advances and global efforts to screen thousands of HIV-1-infected individuals have recently promoted the isolation of nearly pan-neutralizing bNAbs *(28–30)*. This next generation of bNAbs, engineered bi- or trispecific antibodies *(31–34)*, and combination therapies composed of complementary bNAbs *(20, 21, 24)* have the potential to constrict viral escape as well as resistance and therefore to improve future approaches to treat and prevent HIV-1 infections.

Contrary to the progress in passive antibody administration, all efforts to induce highly potent bNAbs through vaccination have failed so far. One commonality across bNAbs is their tendency to accumulate unusual sequence properties, including a high number of somatic mutations *(28–30, 35–38)*, insertions/deletions *(30, 35, 36, 38, 39)*, distinct V_H gene segment use *(40–42)*, or exceptionally long complementarity determining regions (CDRs) *(37)*. Notably, long CDRH3 regions as well as the usage of the V_H gene segment 4-34 have previously been associated with self-reactive antibodies in autoimmune diseases *(43–45)* and several bNAbs proved to be auto- and polyreactive *(46, 47)*. Since autoreactive B cells are stringently counter-selected during B cell development in healthy individuals *(45)*, it has been speculated that the development of certain bNAbs is normally blocked by immune checkpoints that are

bypassed during chronic infection *(48)*. In line with this, it has been demonstrated that BCRs of chronically Hepatitis C virus (HCV)-infected individuals show increased CDRH3s lengths in comparison to healthy controls *(49)*. In addition, many potent HIV-1 bNAbs have been isolated after several years of chronic infection *(30, 42, 50)*, suggesting that a prolonged virus-antibody co-evolution within the HIV-1-infected host is a requirement for their induction. Guided vaccine design builds on this idea and attempts to mimic virus-antibody co-evolution by serial immunizations with varying immunogens *(51, 52)*. However, it remains to be defined which bNAbs have the highest chance to be elicited and should thus be selected for these strategies. In this regard, precursor frequencies for VRC01-class bNAbs and BG18 have previously been estimated in healthy individuals by CDRH3 similarity searches *(52–54)* and probabilities for mutations at distinct positions in bNAbs have been estimated *(55)*. However, systematic approaches to assess the combined probabilities of V(D)J recombination and overall mutation patterns of bNAbs are lacking. Consequently, there is currently no comprehensive analysis that compares the overall generation probabilities of bNAbs among each other and determines the impact of chronic infection on these probabilities.

Here, we performed unbiased next-generation sequencing (NGS) and learned probabilistic models for somatic point mutations and V(D)J recombinations on B cell receptor repertoire data from 57 healthy individuals. We applied these models to determine and compare the generation probabilities of 75 bNAbs. By correlating probabilities with neutralization efficacies, we identified broad and potent candidate bNAbs that are more likely to be elicited than others and are thus particularly suited as targets for vaccination strategies. Finally, we sequenced 34 HIV-1- and 12 hepatitis C virus (HCV)-infected individuals, to infer repertoire characteristics and learn models to determine bNAb generation probabilities in the presence of chronic infections.

RESULTS

Performing unbiased sequencing of the B cell receptor repertoire

In order to determine and compare the probability of developing an antibody with a specific sequence, we aimed to establish a pipeline for collecting unbiased BCR repertoire data from peripheral B cells of a healthy control cohort to infer antibody sequence statistics with high confidence (**Fig. 1A**). To this end, we set up a sorting strategy to isolate naive or IgG-class-switched (i.e. antigen-experienced) B cells from peripheral blood mononuclear cells (PBMCs, **fig. S1**) and developed a 5'-rapid amplification of cDNA ends (RACE)-based sequencing protocol that includes unique molecular identifiers (UMIs) for computational error correction (**Fig. 1A**).

To test whether this sequencing approach yields sufficient high-quality reads, we analyzed 100,000 IgG^+ B cells in biological duplicates from three different healthy blood donors each (**Fig. 1B**). From 1,354,097 to 2,552,903 raw reads per replicate, we reconstituted on average 6,121 IgG heavy, 18,868 kappa, and 12,000 lambda chains after filtering for error-corrected and productive sequences (see **fig. S2** for UMI copy number distribution, **table S1** for complete sequencing statistics, and Methods section for details on quality filtering). There was substantial overlap in unique CDRH3s within samples from the same individuals (on average 2.4 to 4.1% for biological replicates and

81% for the technical replicate) in comparison to the overlap between different donors (<0.02%, **Fig. 1C**). Moreover, repertoire features such as CDRH3 length distribution (**Fig. 1D**) or V_H gene germline identity (**Fig. 1E**) were indistinguishable between biological replicates, but significantly differed across individuals. To estimate the resolution of the sequencing approach, we spiked in different concentrations (0.01 to 10%) of two monoclonal IgM positive B cell lymphoma (BCL) cell lines into naive B cells from a healthy donor (**fig. S3**). Plotting the pairwise Levenshtein distance between all CDRH3s in the sample and the B cell line-derived CDRH3s reveals a bimodal distribution. Spiked-in B cells, which correspond to the first peak of this distribution, could be detected at all concentrations, including as few as 10 in a total of 100,000 cells. Concluding that the sequencing pipeline yields reproducible and representative statistics, we collected blood samples from a healthy reference cohort of 57 individuals that comprised 54% male and 46% female donors with a mean age of 33 years (**Fig. 1F**). By sorting and sequencing in total 5,700,000 IgG positive B cells we generated 96,825,014 raw reads, yielding high-quality productive sequences for 287,505 heavy, 839,776 kappa, and 476,835 lambda chains, with on average 5,044 heavy, 14,733 kappa, and 8,366 lambda chains per individual (**Fig. 1G**, **table S2**). Although sequence features were predictive of the repertoire's origin (**Fig. 1, D and E**), they are mostly conserved between individuals on a global level (**Fig. 1H**). In accordance with previous studies *(56–58)* we find particular V genes (such as V_H1-69 or V_H3-23) to be highly abundant in our cohort, CDR3 lengths to average around 15-16, 9, and 11 amino acids and average mutational loads to peak around 93, 96, and 96% nucleotide V gene germline identity for heavy, kappa, and lambda chains, respectively.

We conclude that the presented pipeline allows for inferring high-quality repertoire statistics from a starting material of 100,000 B cells, and that this sample is representative of the unique IgG BCR repertoire feature distributions of a single individual. On average, IgG repertoires show conserved sequence feature distributions that reflect different probabilities for specific sequence characteristics to develop during B cell development and maturation.

Predicting bNAb probabilities for V(D)J recombination and somatic hypermutation

To predict the probabilities of developing HIV-1 specific bNAbs, we retrieved sequence and neutralization data for 75 HIV-1 neutralizing antibodies with varying neutralization breadths and potencies against a panel of 56 different HIV-1 strains (**fig. S4** and **table S3**) from the CATNAP (Compile, Analyze and Tally NAb Panels) database *(59)*.

The 75 antibodies target various sites on the envelope spike-protein (**Fig. 2A,** adapted from *(7, 60)*) and cover geometric mean IC_{50} values from 0.005 to 16.991 µg/ml as well as neutralization breadths between 10.7 and 98.2%, both of which we combined into a single neutralization score, to compare and rank bNAbs by their overall neutralization efficacy (**Fig. 2B**, black line, **table S4**; see method section for details). Overall, the 75 bNAbs show broad distributions for V_H gene usage (**Fig. 2C**), CDRH3 lengths (**Fig. 2D**), and V_H gene germline identities (**Fig. 2E**), which differ substantially

from the averaged IgG memory B cell repertoire statistics of the 57 healthy individuals. Notably, separating bNAbs into binding classes demonstrates that individual bNAbs are not necessarily extreme in all features. CD4 binding site antibodies, for example, often incorporate V_H1-2 or 1-46 and are highly mutated, while their CDRH3 lengths are within the range of the memory IgG reference distribution (**Fig. 2, C to E**, blue bars). V2-apex antibodies, on the other hand, typically exhibit exceptionally long CDRH3s, but are also less mutated (**Fig. 2, D and E**, yellow bars). Both classes of antibodies can be found among the top neutralizing antibodies (**Fig. 2B**).

While long CDRH3s and high levels of somatic hypermutations could in part explain the rare occurrence of bNAbs, they do not account for heterogeneity in gene segment selection probabilities, or for any sequence context, which is known to strongly bias V(D)J recombination and somatic hypermutation *(61)*. To provide more comprehensive estimates for the probability of generating a given bNAb CDRH3 (P_{gen}) and accumulating a unique pattern of point mutations (P_{SHM}), we applied the Inference and Generation Of Repertoires (IGoR) tool *(62)*. IGoR infers a probabilistic model that accounts for the statistics of V(D)J usage as well as insertions and deletions in the junctional regions and sums over all generation scenarios consistent with the given BCR sequence to evaluate its overall generation probability. The SHM model learns the identities of mutated 5-mer subsequences. By learning both models on the quality-filtered productive IgG sequences of all 57 healthy individuals, we determined P_{gen} and P_{SHM} for the 75 bNAb heavy chain sequences (**Fig. 2F**, **table S5**). Almost all bNAbs show lower P_{gen} and P_{SHM} values (i.e., are less likely) than the median healthy IgG repertoire distribution, resembling the CDRH3 length and V_H gene germline identity comparison (**Fig. 2D and E)**. Indeed, there is a strong correlation between these sequence features and the probabilities (**fig. S5**), suggesting that CDRH3 length and numbers of mutations are strong determinants of antibody probabilities. However, there are also antibodies with equal CDRH3 lengths (e.g. VRC03/DH511.11P) or similar amounts of SHM (e.g. VRC06b/3BNC117) but substantially different probabilities (over several log steps), highlighting the contribution of additional determinants such as biases in V(D)J recombination and SHM (**fig. S5**).

Given the broad distributions of P_{gen} and P_{SHM} for the different bNAbs, we asked whether neutralization efficacy is correlated with these probabilities, and whether we can identify highly potent bNAbs that are more likely to be generated than other ones. The neutralization score is not correlated with P_{gen} but slightly correlated with P_{SHM} alone for the 75 bNAbs (**Fig. 2G**, left and middle panel). Performing a linear regression that takes into account the logarithms of P_{gen} and P_{SHM} yields a combined probability score S (**Fig. 2B**, right panel, **table S5**) that is highly predictive of the neutralization score ($r = -0.446$ and $p = 6.031 \times 10^{-5}$ for the linear regression). Based on this regression, lower probability scores (i.e. less likely bNAbs) are correlated with better neutralization. Importantly, the correlation allows the identification of bNAbs that are highly potent, but easier to generate than other bNAbs (**Fig. 2G**, right panel, upper right corner). Similar results for the correlation of P_{gen}, P_{SHM}, and the probability score S with the neutralization score were also obtained for bNAb light chains (**fig. S6**, **table S5**).

We conclude that bNAbs are in general unlikely outcomes of the B cell development and that a linear combination of the logarithms of P_{gen} and P_{SHM} is highly

predictive of the neutralization capacity with the overall tendency of less likely bNAbs to be the most potent ones.

Global BCR repertoire features under chronic infections are similar to healthy repertoires with marginal differences

Since chronic infections have been described to interfere with B cell development and functionalities on several levels (reviewed in *(63)*), we wondered whether and to what extent they influence memory IgG B cell receptor sequence characteristics. To answer this question, we sequenced BCRs from the IgG^+ memory compartment of 34 HIV-1 and 12 hepatitis C virus (HCV)-infected individuals and compared them to the 57 healthy control repertoires (**Fig. 3A**).

Processing and filtering of 54,628,577 HIV-1 and 19,767,705 HCV raw reads yielded in total 1,382,301 (HIV-1) as well as 486,488 (HCV) heavy and light chain sequences (**tables S6 and 7**). Overall, sequence characteristics, including V gene segment usage, CDR3 length, and V gene germline identity, were comparable to the healthy control cohort for both chain types (**Fig. 3, B to D**), although HCV-infected individuals showed slightly longer mean CDRH3 lengths (**Fig. 3C**). Similarly, no obvious differences were observed in B cell lineage tree structures (**Fig. 3E**) with respect to skewedness (**Fig. 3F**), size distribution (**Fig. 3G**), or diversity (**Fig. 3H**) across the three cohorts (see methods section for details). Since antiretroviral therapy (ART) is suppressing virus replication and thus may dampen anti-HIV-1 immune responses, we stratified the HIV-1 cohort by treatment (**fig. S7**). Whereas repertoires from treated HIV-1-infected individuals were indistinguishable from healthy individuals, we found a substantially higher fraction of V_H genes 1-69 and 4-34 in untreated HIV-1-infected individuals, which did not reach significance after correcting for multiple testing. In addition, mean CDRH3 lengths were slightly but significantly longer in untreated individuals. Mean V gene germline identities did not differ significantly between the stratified subgroups, although the untreated subgroup contains one individual with noticeably less somatic mutations (i.e. a higher mean V gene germline identity), which is mainly responsible for the visible shift in the V gene germline identity distributions (**Fig. 3D**, **fig. S7**). We also measured serum-derived poly-IgG neutralization breadth against the 12-strain global panel *(64)* for 33 of the 34 HIV-1-infected individuals (**table S8**). Of note, broad serum neutralization is found in ART-naive (untreated) as well as ART-treated individuals of our cohort (**table S8**). Stratifying BCR repertoires by neutralization breadth into high (≥66%), intermediate (<66% and ≥33%), or low (<33% and >0%) breadth and no neutralization (0%) did not yield any evidence that serum neutralization breadth is linked to relevant global changes in BCR repertoire features (**fig. S8**). Neither of the stratifications had any influence on the lineage characteristics (data not shown).

We conclude that V gene usages and CDRH3 lengths are marginally skewed in untreated, chronically infected individuals. However, on average, repertoire features remain almost constant between the cohorts in this study.

Equal probabilities for bNAb development in healthy and chronically infected individuals

To investigate whether chronic infection shifts the probability to develop distinct bNAb sequences, we used the IgG heavy chain repertoire data from HIV-1 and HCV-infected individuals to train IGoR and learn models for P_{gen} and P_{SHM}.

Whole repertoire P_{gen} distributions of both cohorts were almost identical to the healthy control data for heavy, kappa, and lambda chains (**Fig. 4A**). Repertoire P_{SHM} distributions on the other hand were slightly shifted towards less negative values for HCV- and HIV-1-infected individuals (**Fig. 4B**), which closely resembles the increase in V_H gene germline identity (i.e. less somatic mutations) that we already observed in the repertoire data (**Fig. 3D**). We speculated that marginal differences in these distributions could still influence the probability for a certain rare V(D)J recombinations or mutational patterns of distinct bNAb sequences. Therefore, we calculated P_{gen} and P_{SHM} for the individual 75 bNAbs with the models that have been trained on the repertoire data from infected individuals (**table S5**). When comparing those to the values that have been inferred from the healthy cohort, there is only little deviation in P_{gen} or P_{SHM} (**Fig. 4, C and D**). There are a few bNAb heavy and light chain probabilities that deviate by more than 25% of the range of all observed values in comparison to the healthy cohort (**Fig. 4, C and D**, black outlines). Interestingly, P_{gen} for the light chain of 10-1074 is >5 log-fold higher (i.e., more likely) in healthy individuals than in both chronically infected cohorts, suggesting that the induction of this particular light chain sequence is somehow hampered by chronic infections. There is however no tendency for a whole antibody epitope group to be favored in one or the other cohort. To compare the overall probability of the individual bNAbs between the three cohorts, we calculated the combined probability score (S) using the cohort-specific P_{gen} and P_{SHM} values together with the coefficients from the healthy repertoire linear regression model of Figure 2G (**table S5**). As already suggested by the highly similar P_{gen}/P_{SHM} values, combined probability scores for bNAbs were almost identical between healthy and infected individuals (**Fig. 4E**). Despite the slight differences in the repertoire characteristics in untreated individuals (compare **fig. S7**), treatment status had no influence on bNAb probability scores (**Fig. 4F**). Similarly, we found no global change in probability scores in repertoires stratified by serum neutralization breadth, except for three less probable chains in high (CH103 light chain), intermediate (10-1074 heavy chain), and low breadth subgroups (179NC75 heavy chain, **Fig. 4G**). Taken together, these results suggest no substantial differences in the overall probability of bNAb generation in the absence or presence of chronic HIV-1 or HCV infection.

DISCUSSION

Broadly HIV-1 neutralizing antibodies can effectively prevent and treat HIV-1 infections in animal models *(11, 12, 14, 65)* and are considered promising tools for next-generation HIV-1 prevention and therapy in humans *(66, 67)*. However, bNAbs have only been identified in a minor fraction of HIV-1-infected individuals and to elicit highly broad and potent bNAbs by vaccination remains a yet unreached goal *(7, 68, 69)*. The presence of high amounts of somatic mutations, insertions/deletions, or exceptionally

long CDRH3s in bNAbs raised the questions of (i) whether these outstanding features explain their rareness, (ii) whether rareness is correlated with neutralization capacity, and (iii) if chronic infection is a prerequisite for their induction. By learning models from IgG^+ memory B cell receptor sequences of 57 healthy individuals, we estimated the probability for specific V(D)J recombinations (P_{gen}) and for accumulating patterns of somatic point mutations (P_{SHM}) in a representative set of HIV-1-neutralizing antibodies. We also tried to include insertions and deletions (indels). However, it was not possible to build a robust model, since the frequencies of indels were too low in the repertoire data. In line with previous studies, we found that numerous bNAbs accumulate improbable mutation patterns *(55, 70)*. Our analysis illustrates that the degree of improbable mutations is correlated with the binding site and is not a common feature of bNAbs per se. Similarly, we identified a subset of antibodies with highly improbable CDRH3 recombinations. These mainly comprise V2 apex antibodies that require a long CDRH3 to penetrate the envelope glycan shield *(71–73)*, but also certain CD4bs antibodies with average CDR3 lengths that represent unlikely V(D)J recombinations. These results support the hypothesis that bNAbs are rare, because the sum of their sequence features represents unlikely events in B cell development.

Previously, ongoing affinity maturation as well as the mean frequency of somatic mutations have been correlated with neutralization breadth and potency of bNAb lineages in HIV-1-infected individuals *(74, 75)*. In line with this, we detected a weak correlation for the neutralization power of monoclonal bNAbs and SHM, looking not only at the frequency but also at the probability to accumulate a specific pattern of somatic mutations (i.e., P_{SHM}). Notably, we found that a linear combination of the logarithms of P_{gen} and P_{SHM} is predictive of bNAb neutralization efficacy, suggesting that bNAbs tend to be better, the more unlikely they are. Importantly, the correlation allows for identifying bNAbs that are highly potent but on average more likely than other, similarly potent ones. These highly potent and probable antibodies include the above-mentioned V2 apex antibodies, which in contrast to their improbable CDRH3 showed more probable patterns of somatic hypermutations. Interestingly, V2 apex antibodies have been reported to occur relatively early during infection *(76–78)*. This suggests that the selection of rare precursors outperforms the accumulation of distinct SHM in timing and that this bNAb subclass could thus be a preferential target for vaccination strategies.

In addition to healthy individuals, we also sequenced the repertoires of 34 HIV-1- and 12 HCV-infected individuals. In concordance with previous studies *(49, 79)*, we detected marginally longer mean CDRH3 lengths in HCV-infected as well as HIV-1-infected individuals that do not receive antiretroviral therapy. Similar to the findings from Roskin et al. *(79)*, we also detected an increase of the inherently auto-reactive gene segment V_H4-34 in our ART-naive HIV-1-infected subgroup. However, we do not see substantial differences in bNAb sequence probability scores in the IgG^+ memory compartment when averaging over individuals in the cohorts. This suggests that chronic infection itself does not alter the probabilities to develop specific bNAbs substantially in any direction, even if it leaves traces of particular sequence features within the whole IgG repertoire. Of note, we detected high variability in the repertoire composition within each cohort. As a consequence, some individuals might still have better chances of developing certain bNAb classes because of other predispositions, e.g. by carrying

353 specific gene segment alleles that encode for critical contact sites, as it has been
354 previously demonstrated for VRC01-like bNAbs *(53)*.

355 Taken together, we present a framework for assessing the overall probability to
356 develop an antibody with a specific CDR3 motif and pattern of point mutations. Our
357 analysis on HIV-1-neutralizing antibodies suggests that bNAbs are generally unlikely
358 outcomes of the B cell evolution within a host due to individual or a combination of
359 improbable sequence features. Using neutralization data and probability scores, we
360 demonstrate that the more unlikely bNAbs tend to have a higher neutralization efficacy.
361 Moreover, this approach allows us to identify potent antibodies with higher chances to
362 be elicited by vaccination. Finally, the data suggest that chronic infection has no impact
363 on the generation of bNAb sequences within the IgG^+ B cell memory compartment,
364 fostering the hope that a potent vaccine should be able to elicit bNAbs in healthy
365 individuals.

366

367 **MATERIALS AND METHODS**

368 **Blood donors and sample collection**

369 Samples were obtained under a study protocol approved by the Institutional Review
370 Boards of the University of Cologne, University of Bonn, and the respective local IRBs
371 (study protocols 16-054 and 017/16). All participants provided written informed consent
372 and were recruited at hospitals or as outpatients.

373 **Serum IgG isolation**

374 Serum samples from HIV-1-infected individuals were heat-inactivated at 56 °C for 40
375 min and incubated with Protein G Sepharose (GE Life Sciences) overnight at 4 °C. IgGs
376 were eluted from chromatography columns using 0.1 M glycine (pH = 3.0).
377 Subsequently, the eluate was buffered in 0.1 M Tris (pH = 8.0). Buffer was exchanged
378 to PBS using Amicon 30 kDa spin membranes (Millipore). The concentration of purified
379 IgGs was determined by UV/Vis spectroscopy using Nanodrop (A280) and samples
380 were stored at 4°C.

381 **Serum IgG virus neutralization tests for HIV-1-infected individuals**

382 Neutralization assays with serum IgG from HIV-1-infected against the 12-strain "global"
383 virus panel, were performed as previously described *(30, 80)*. In brief, 12 HIV-1
384 pseudovirus strains were mixed with 1:2 serial dilutions of purified IgG (1 mg/ml final
385 starting concentration, 8 dilutions) and incubated for 1 hour at 37°C. After incubation,
386 10^4 TZM-bl cells were added per well on a 96-well plate in 250 µl medium
387 supplemented with DEAE-dextran at a final concentration of 10 µg/ml. This mixture was
388 incubated for 2 days at 37°C and 5% CO_2. Before adding 100 µl luciferase assay
389 reagent to the well 150 µl culture supernatant was removed. After 2 min incubation the
390 mixture was transferred to a microtiter assay plate and luminescence signal was
391 captured using a luminometer (Berthold TriStar^2 LB942) with relative luminescence
392 units (RLUs) as output unit. To assess the 50% inhibitory concentration (IC_{50}), first, the
393 background signal was determined using non-infected TZM-bl cells and subtracted.
394 Finally, IgG concentrations resulting in a 50% RLU reduction compared to untreated
395 virus control wells were determined by using murine leukemia virus (MuLV)-

396 pseudotyped virus as a control for unspecific activity. All samples were tested in
397 duplicates.

398 **Isolation of B cells and RNA Isolation**

399 Collection of peripheral blood samples was performed with syringes pre-filled with
400 Heparin-Natrium. PBMCs were isolated by standard density gradient centrifugation
401 using Histopaque separation medium (Sigma Aldrich) and Leucosep cell tubes (Greiner
402 Bio-one) according to the manufacturer's instructions. Cells were stored at -150 °C in
403 90% (v/v) FBS (Sigma Aldrich) and 10% (v/v) DMSO (Sigma Aldrich). Plasma was
404 collected and stored separately at -80 °C. CD19+ B cells were selected from PBMCs by
405 magnetic-activated cell sorting (MACS) using CD19 microbeads (Miltenyi Biotec). For
406 cell sorting, CD19+ cells were stained with anti-human AF700-CD20 (BD), APC-IgG
407 (BD), PE-Cy-7-IgD (BD), FITC-IgM (BD), PerCP-Cy5.5 (BD) or PE-CD27 (BD) and
408 DAPI (Thermo Fischer). For each individual, 100,000 $CD20^{+}IgG^{+}$ and 100,000
409 $CD20^{+}IgD^{+}IgM^{+}CD27^{-}IgG^{-}$ B cells were sorted into FBS-coated (Sigma-Aldrich) 2 mL
410 tubes using a FACSAria Fusion cell sorter (Becton Dickinson). After the sort the cells
411 were centrifuged for 5 minutes at 400 x g to remove the supernatant. Cell pellets were
412 placed on dry-ice or frozen at -80°C before RNA-Isolation was performed using the
413 RNeasy Micro Kit (Qiagen) on a QiaCube (Qiagen) instrument and RNA was eluted in
414 25 µl RNAse free water.

415 **RT-PCR and next generation sequencing**

416 B cell receptor repertoire sequence data was generated by an unbiased template-
417 switch-based approach. cDNA was generated from 10 µl RNA according to the
418 SMARTer RACE 5'/3' manual using the SMARTScribe Reverse Transcriptase (Takara)
419 and a self-designed template-switch oligo
420 (AGGGCAGTCAGTCGCAGNNNNWSNNNNWSNNNNWSGCrGrGrG) including an 18-
421 nucleotide unique molecular identifier (UMI) in a 20 µl batch. cDNA was diluted with 10
422 µl Tricine-EDTA buffer (Takara) according to the manual. Heavy and light chain variable
423 regions were pre-amplified from 5 µl cDNA each by PCR with 1 µM of a universal
424 forward primer (CTGATACGATTCACGCTAGGGCAGTCAGTCGCAG) and 0.33 µM
425 constant region-specific reverse primers (IgM: ATGGAGTCGGGAAGGAAGTC, IgG:
426 AGGTGTGCACGCCGCTGGTC, IgK: GGTGACTTCGCAGGCGTAG, IgL:
427 GCCGCGTACTTGTTGTTGC) in a 30 µl reaction with Q5 DNA polymerase (New
428 England Biolabs). Cycling conditions were one cycle 98°C/30 s, four cycles 98°C/10 s
429 and 72°C/30 s, four cycles 98°C/10 s, 62°C/30 s (IgG/IgM) or 68°C(IgK/IgL)/30 s, and
430 72°C/30 s, as well as a final extension cycle at 72°C/5 min. PCR products were purified
431 with a PCR clean-up Kit (Macherey Nagel) according to the manual with the cDNA
432 being initially filled upt with a 1/6 dilution of NTI buffer in RNAse-free water to remove
433 primers and low molecular weight products. Samples were eluted in 15 µl pre-warmed
434 elution buffer (Macherey Nagel). Heavy and light chain amplicons were enriched from 5
435 µl purified pre-amplification product by a nested PCR with 0.33 µM forward primer (IgG:
436 NNNNNCACGCTAGGGCAGTCAG; IgM: NNNNNNCACGCTAGGGCAGTCAG;
437 IgK/IgL: NNNNCACGCTAGGGCAGTCAG) and 0.33 µM nested reverse primers (IgG:
438 NNNNNSGATGGGCCCTTGGTGGARGC; IgM:
439 NNNNGGTTGGGGCGGATGCACTCC; IgK:
440 NNNNNNGGGAAGATGAAGACAGATGGT, IgL:

441 NNNNNNGGGYGGGAACAGAGTGACC) with Q5 DNA Polymerase (New England
442 Biolabs) in a 100 µl reactions. For thermocycling, reactions were split into 2x50 µl and
443 run on a thermocycler (Applied Biosystems) with one cycle 98°C/30 s, five cycles
444 98°C/10 s and 72°C/30 s, five cycles 98°C/10 s and 70°C/30 s, 17 cycles 98°C/10 s,
445 68°C/30 s (IgG/IgM) or 62°C(IgK/IgL)/30 s, and 72°C/30 s, as well as a final extension
446 cycle at 72°C/5 min. PCR amplicons separated on a 1% Agarose gel and purified with a
447 PCR and Gel Clean Up Kit (Macherey Nagel), eluted in 23 µl elution buffer (Macherey
448 Nagel) and subjected to library preparation and Illumina MiSeq 2 x 300 bp sequencing
449 (v3) at the Cologne Center for Genomics sequencing core facility.

450 **NGS data processing and sequence annotation**

451 Data pre-processing was performed with a Python(v.3.6)-based pipeline, including
452 python packages BioPython (v.1.78), pandas (v.1.1.5), Numpy (v.1.19.2), Matplotlib
453 (v.3.3.4), and Levensthein (v.0.12.0). Raw NGS reads were filtered for a mean Phred
454 score of 25 and read-lengths of at least 250 bp. Reads were annotated with IgBLAST
455 *(81)* to identify read-orientation and CDR3 sequence. Reads were then grouped by the
456 UMI and wrongly assigned reads (collisions) were identified by comparison of V gene
457 calls and a second 18-nucleotide molecular identifier within the CDR3. Reads were then
458 aligned with Clustal Omega *(82, 83)* and consensus sequences were generated by
459 taking into account the quality score-weighted base-calls for each position. The 2x
460 300pb consensus read pairs were then aligned and combined using the
461 AssemblePairs.py script form the pRESTO toolkit with a minimal overlap of 6
462 nucleotides *(84)*. Annotation of pre-processed reads was performed by using BLAST
463 (v.2.9) and IgBLAST(v.1.13) *(81)*. Reference templates for all functional heavy and light
464 chain V(D)J genes were retrieved from the IMGT database *(85, 86)* (265 IGHV, 30
465 IGHD, 13 IGHJ, 66 IGKV, 9 IGKJ, 70 IGLV, and 7 IGLJ at the time point of this study).
466 After sequence annotation with IgBLAST, sequences were filtered to include complete
467 V/J annotations covering at least 250 nucleotides of the V gene as well as productive
468 sequences (i.e., no STOP codon and in-frame CDR3 recombination, as determined by
469 IgBLAST) only. To minimize the influence of sequencing and PCR errors, annotated
470 sequences were only used for downstream analyses, when their UMIs were initially
471 found in at least three reads (high-quality reads). For learning P_{gen} and P_{SHM} models,
472 sequences were additionally filtered out, when they contained gaps (i.e.
473 insertions/deletions).

474 **Antibody selection**

475 Antibody information was retrieved from the CATNAP database *(59)*. At the time of this
476 study, the available antibody dataset comprised 507 entries, from which 80 were listed
477 as polyclonal antibodies or antibody mixtures and excluded from further analyses. The
478 remaining set of 427 entries required further polishing to remove non-human and
479 engineered antibodies. To this end, antibody entries were filtered to exclude any of the
480 keywords 'llama', 'chimeric', 'mouse', 'rabbit', 'bispecific', 'trispecific', 'IgG3C' in the 'LANL
481 comments' column or any '/', '-LS', or 'mutant' in the entry name. Finally, the remaining
482 mutants/chimeras and non-antibody proteins '10-1074GM', '10-1074V', '10E8v4',
483 '10E8v4-5R-100cF', '45-46m2', '45-46m25', '45-46m28', '45-46m7', 'sCD4', 'CD4-Ig',
484 'gVRC-H1dC38-VRC01L', 'iMab', 'm66-G30K', 'm66-S28H', 'm66-S28H-G30K', 'm66-
485 S28H-G30K-S31K', 'm66-S31K', 'm66-Y100eW', 'NIH45-46G54F', 'NIH45-46G54W',

'NIH45-46G54Y', 'NIH45-46WPY', 'NIH45-46WY', 'PGT121GM', 'VRC34YD.01', 'VRC34YD.02', 'VRC34YD.03', 'VRC34YD.04', 'VRC34YD.05', 'VRC34YD.06', 'VRC34YD.07', 'VRC38.12', 'VRC38.13', 'VRC38.14' were manually removed to yield a final set of 291 antibodies.

Although several standard virus panels such as the 12-strain 'global' panel *(64)* and the 118-strain multi-clade panel *(87)* have already been established, there is currently no consensus panel. As a consequence, the 291 antibodies were tested against different or only partially overlapping viral panels, making it difficult to compare them in terms of breadth and potency *(88)*. Neutralization data against the global and the 118 multi-clade panel were downloaded separately from the CATNAP database. The database outputs antibody information for a given panel, if at least one strain of this panel has been analyzed by a study. Missing values for virus neutralization data (i.e., viruses against which no assay has been performed in any study) are marked with '-'. To identify the number of antibodies that have been tested against the complete panels, we excluded any of the 291 antibodies that had missing values ('-') against one or more viral strain of the respective panel (i.e. 19 antibodies remaining for the 118-strain panel and 50 antibodies remaining for the global panel). As a trade-off between number of antibodies and accuracy of breadth and potency, we selected a subset of 56 strains from the 118 multi-clade panel, which resembles its clade distribution, its tiered categorization, and yields similar values for breadth and potency (**fig. S4**, **table S3**). 89 out of 291 antibodies have been tested against the 56 strain panel. 75 out of the 89 antibodies were published with complete heavy chain nucleotide sequences (covering the full V region, including the CDR3), neutralized at least 1 of the 56 strains, and were therefore used for the downstream analyses. For 69 out of the 75 antibodies, complete light chain nucleotide sequences were published. Heavy and light chains of the 75 bNAbs were annotated with the same IgBLAST/IMGT-based pipeline as the NGS data (64 of the 69 light chains could be properly annotated and were used for analyses).

Determination of antibody breadth, mean potency and neutralization score

IC_{50} values for individual antibody-virus combinations were derived from the CATNAP database *(59)* as the geometric mean of all published IC_{50} values (i.e. from different studies). IC_{50} values below the detection limit were set to an arbitrary threshold of 100 µg/ml. The mean potency of an antibody against a viral panel was determined by the geometric mean of all geometric mean IC_{50} values that were below the arbitrary threshold of 100 µg/ml (also known as the mean potency of all neutralized strains). Antibody breadth was defined as the percentage of neutralized strains (i.e. geometric mean IC_{50} above the arbitrary threshold of 100 µg/ml) to the total number of strains tested and is given as percentage. To determine the combined neutralization score, we took all IC_{50} values that were below the threshold of 100 µg/ml for each antibody and sorted them from lowest to highest. We then divided the maximum coverage (100%) into 56 equally spaced increments (~1.786% coverage per strain), and assigned cumulative increments to the sorted IC_{50} values, i.e. the lowest IC_{50} is assigned to a coverage of 1.786%, the second lowest to 3.572%, and so on. Finally we plotted the $\log_{10}$ of the sorted IC_{50} values against the cumulative coverage and determined the area under the curve. Low IC_{50} values (i.e. highly potent antibodies) will increase the area by shifting the curves to the left, while higher breadth will increase the area by shifting the

plateau of the curve to the top. For the ranking of bNAbs in (**Fig. 2B**), the neutralization score is given as percentage of the highest score among the 75 antibodies. For correlation of probabilities and neutralization (**Fig. 2 and 4**), the $\log_{10}$ of the neutralization score was used.

IGoR inference and prediction of bNAb probabilities

Models for the probability of generation (P_{gen}) and somatic hypermutation (P_{SHM}) were learned by using the Inference and Generation Of Repertoires (IGoR) tool *(62)*. To reduce uncertainty in model inference and focus on comparison between cohorts, we pooled all productive sequences from all patients in a cohort. Productive sequences were chosen to explore the effect of possible selection effects on P_{gen} and P_{SHM} of a given BCR sequence. The somatic hypermutation (SHM) model is 5-mer based, taking into account the mutated position and its two neighbors on both sides. The model takes into account the full composition of these 5-mers and is context dependent. P_{gen} generation models were consistent with models inferred on previously published repertoires *(58, 89)* (data not shown). Indels in bNAbs were reverted to the most likely V and J templates according to the IgBLAST annotation before model building to improve alignment by IGoR. Masking of indels has no influence on P_{gen} or P_{SHM}. To test correlations between the neutralization score and the overall likelihood of seeing a given bNAb, we defined a score S for each bNAb as $S = c_1 \log_{10} P_{gen} + c_2 \log_{10} P_{SHM}$. Coefficients c_1 and c_2 were determined by linear regression of score S versus the $\log_{10}$ of neutralization score.

Lineages and trees

After V, D and J annotation, unique sequences were partitioned into lineages. First, the sequences were grouped into classes of identical V and J genes and equal CDR3 length. Within each class, lineages were identified using single linkage clustering with a fixed threshold of 90% CDR3 nucleotide identity. Tree length was then estimated as the total number of unique mutations found in a given lineage (a lower bound on the true tree length). For productive lineages of more than 50 unique sequences tree topologies and branch lengths were inferred using RAxML with the GTRGAMMA model of nucleotide substitution *(90)*. Germline V and J genes were provided as an outgroup to aid the inference. To quantify the asymmetry of the phylogenies within each cohort we examined the distributions of two indices of imbalance, following *(91)*. We compare the weight of the common ancestor of the lineage w_{anc} with the weights of its immediate descendants, w_D. The weight of a node is the number of leaves that stem from it and the ratio w_D/w_{anc} quantifies the imbalance at the first branching. Additionally, we estimated the distribution of the ratio of mean terminal branch length to the mean length of all branches.

Quantification and statistical analyses

Flow cytometry analysis and quantifications were done with FlowJo10 software. Statistical analyses were performed using GraphPad Prism (v8), Microsoft Excel for Mac (v14.7.3), Python (v3.6.8), and R (v4.0.0). For the identification of overlapping clonotypes in healthy individuals a maximum of one amino acid length difference and three or less differences in absolute amino acid composition of CDR3s were considered as similar. To calculate the CDRH3 sequence overlap between replicates in the control

experiment (**Fig. 1C**), a random sample of n=3623 (i.e. the size of the smallest dataset) was drawn from all unique CDRH3s of each set and the overlap (i.e. identical CDRH3 sequences) between these random samples was determined. Sampling and overlap determination was repeated 100 times and the mean overlap over all iterations was reported. Testing for significant differences between CDR3 length distributions and V gene germline identity distributions (**Fig. 1, D and E**) was performed by global Kruskal wallis tests (stats.kruskal, Scipy v.1.5.2) and Dunn post-hoc tests (posthoc_dunn, scikit_posthocs v. 0.4.0 with 'holm' method for p value adjustment). To estimate the fraction of spiked-in lymphoma B cells among naive B cells from a healthy donor (**fig. S3**), the pairwise Levenshtein distance of all re-constituted CDRH3s was determined (python-Levenshtein v.0.12.2) in comparison to the most frequent cell line CDRH3 (experimentally determined by sequencing 24 single B cells for each cell line, data not shown). The observed distance frequencies revealed a bimodal distribution from which all comparisons with <4 amino acid distance (first peak) were counted as a B cell line CDRH3 and divided by all comparisons to get the fraction. For cluster analysis of the mixed cell sample (**fig. S3**), a network analysis was performed with the networkx python package (v.2.2). Each node represents a unique CDRH3. The node size is proportional to the frequency among all identified CDRH3s and nodes are connected, if they share at least 75% of their CDRH3 amino acid sequence. Nodes are colored according to the cell lines if they share at least 75% of the CDRH3 amino acid sequence with a cell line. Phylogenetic trees of viral panels (**fig. S4**) were generated and illustrated from aligned viral sequences from the CATNAP database *(59)* with Geneious Prime software (v.2020.2.4, Jukes-Cantor genetic distance model and Neighbor-Joining method with 100 bootstrap replicates). Testing for significant differences in mean CDR3 lengths and mean V gene germline identities (**Fig. 3, C** and **D**, **fig. S7 and S8**) was performed by one-way ANOVA (stats.f_oneway, Scipy v.1.5.2) followed by Tukey-HSD post hoc test (stats.multicomp.pairwise_tukeyhsd, statsmodels v.0.12.2). Differences in V gene distributions (**Fig. 3B**, **fig. S7 and S8**) were investigated by individual Kruskal-Wallis tests (stats.kruskal, Scipy v.1.5.2) for each V gene with Bonferroni correction for multiple testing. In the case of significant differences in the global test, a Dunn post hoc test (posthoc_dunn, scikit_posthocs v.0.4.0 with 'holm' method for p value adjustment) was performed for subgroup analysis.

Supplementary Materials

Figs. S1 to S8

Tables S1 to S7

References 92 – 113

References and Notes

1. F. W. Alt, Y. Zhang, F. L. Meng, C. Guo, B. Schwer, Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* **152**, 417–429 (2013).

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

3. N. Matthews, Annual Review of Immunology. *J. Med. Genet.* **23**, 284–285 (1986).

618 4. M. H. Malim, M. Emerman, HIV-1 sequence variation: Drift, shift, and attenuation.
619 *Cell* **104**, 469–472 (2001).

620 5. A. Nourmohammad, J. Otwinowski, J. B. Plotkin, Host-Pathogen Coevolution and the
621 Emergence of Broadly Neutralizing Antibodies in Chronic Infections. *PLoS Genet.* **12**,
622 1–23 (2016).

623 6. L. Stamatatos, L. Morris, D. R. Burton, J. R. Mascola, Neutralizing antibodies
624 generated during natural hiv-1 infection: Good news for an hiv-1 vaccine? *Nat. Med.* **15**,
625 866–870 (2009).

626 7. F. Klein, H. Mouquet, P. Dosenovic, J. F. Scheid, L. Scharf, M. C. Nussenzweig,
627 Antibodies in HIV-1 vaccine development and therapy. *Science (80-.).* **341**, 1199–1204
628 (2013).

629 8. E. Landais, P. L. Moore, Development of broadly neutralizing antibodies in HIV-1
630 infected elite neutralizers. *Retrovirology* **15**, 1–14 (2018).

631 9. I. A. Abela, C. Kadelka, A. Trkola, Correlates of broadly neutralizing antibody
632 development. *Curr. Opin. HIV AIDS* **14**, 279–285 (2019).

633 10. A. J. Hessell, E. G. Rakasz, P. Poignard, L. Hangartner, G. Landucci, D. N. Forthal,
634 W. C. Koff, D. I. Watkins, D. R. Burton, Broadly neutralizing human anti-HIV antibody
635 2G12 is effective in protection against mucosal SHIV challenge even at low serum
636 neutralizing titers. *PLoS Pathog.* **5** (2009), doi:10.1371/journal.ppat.1000433.

637 11. F. Klein, A. Halper-Stromberg, J. A. Horwitz, H. Gruell, J. F. Scheid, S. Bournazos,
638 H. Mouquet, A. Abadir, R. Diskin, A. Abadir, T. Zang, M. Dorner, E. Billerbeck, R. N.
639 Labitt, C. Gaebler, P. M. Marcovecchio, R. B. Incesu, T. R. Eisenreich, P. D. Bieniasz,
640 M. S. Seaman1, P. J. Bjorkman, J. V. Ravetch, A. Ploss, M. C. Nussenzweig, HIV
641 therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature*
642 **492**, 118–122 (2012).

643 12. B. Moldt, E. G. Rakasz, N. Schultz, P. Y. Chan-Hui, K. Swiderek, K. L. Weisgrau, S.
644 M. Piaskowski, Z. Bergman, D. I. Watkins, P. Poignard, D. R. Burtona, Highly potent
645 HIV-specific antibody neutralization in vitro translates into effective protection against
646 mucosal SHIV challenge in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18921–18925
647 (2012).

648 13. M. Shingai, O. K. Donau, R. J. Plishka, A. Buckler-White, J. R. Mascola, G. J.
649 Nabel, M. C. Nason, D. Montefiori, B. Moldt, P. Poignard, R. Diskin, P. J. Bjorkman, M.
650 A. Eckhaus, F. Klein, H. Mouquet, J. C. C. Lorenzi, A. Gazumyan, D. R. Burton, M. C.
651 Nussenzweig, M. A. Martin, Y. Nishimura, Passive transfer of modest titers of potent
652 and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in
653 macaques. *J. Exp. Med.* **211**, 2061–2074 (2014).

654 14. R. Gautam, Y. Nishimura, A. Pegu, M. C. Nason, F. Klein, A. Gazumyan, J.
655 Golijanin, A. Buckler-White, R. Sadjadpour, K. Wang, Z. Mankoff, S. D. Schmidt, J. D.
656 Lifson, J. R. Mascola, M. C. Nussenzweig, M. A. Martin, A single injection of anti-HIV-1
657 antibodies protects against repeated SHIV challenges. *Nature* **533**, 105–109 (2016).

658 15. Y. Nishimura, R. Gautam, T. W. Chun, R. Sadjadpour, K. E. Foulds, M. Shingai, F.
659 Klein, A. Gazumyan, J. Golijanin, M. Donaldson, O. K. Donau, R. J. Plishka, A. Buckler-

660 White, M. S. Seaman, J. D. Lifson, R. A. Koup, A. S. Fauci, M. C. Nussenzweig, M. A.
661 Martin, Early antibody therapy can induce long-lasting immunity to SHIV. *Nature* **543**,
662 559–563 (2017).

663 16. M. Caskey, F. Klein, J. C. C. Lorenzi, M. S. Seaman, A. P. West, N. Buckley, G.
664 Kremer, L. Nogueira, M. Braunschweig, J. F. Scheid, J. A. Horwitz, I. Shimeliovich, S.
665 Ben-Avraham, M. Witmer-Pack, M. Platten, C. Lehmann, L. A. Burke, T. Hawthorne, R.
666 J. Gorelick, B. D. Walker, T. Keler, R. M. Gulick, G. Fätkenheuer, S. J. Schlesinger, M.
667 C. Nussenzweig, Viraemia suppressed in HIV-1-infected humans by broadly
668 neutralizing antibody 3BNC117. *Nature* **522**, 487–491 (2015).

669 17. T. Schoofs, F. Klein, M. Braunschweig, E. F. Kreider, A. Feldmann, L. Nogueira, T.
670 Oliveira, J. C. C. Lorenzi, E. H. Parrish, G. H. Learn, A. P. West, P. J. Bjorkman, S. J.
671 Schlesinger, M. S. Seaman, J. Czartoski, M. J. McElrath, N. Pfeifer, B. H. Hahn, M.
672 Caskey, M. C. Nussenzweig, HIV-1 therapy with monoclonal antibody 3BNC117 elicits
673 host immune responses against HIV-1. *Science (80-.).* **352**, 997–1001 (2016).

674 18. M. Caskey, T. Schoofs, H. Gruell, A. Settler, T. Karagounis, E. F. Kreider, B. Murrell,
675 N. Pfeifer, L. Nogueira, T. Y. Oliveira, G. H. Learn, Y. Z. Cohen, C. Lehmann, D. Gillor,
676 I. Shimeliovich, C. Unson-O'Brien, D. Weiland, A. Robles, T. Kümmerle, C. Wyen, R.
677 Levin, M. Witmer-Pack, K. Eren, C. Ignacio, S. Kiss, A. P. West, H. Mouquet, B. S.
678 Zingman, R. M. Gulick, T. Keler, P. J. Bjorkman, M. S. Seaman, B. H. Hahn, G.
679 Fätkenheuer, S. J. Schlesinger, M. C. Nussenzweig, F. Klein, Antibody 10-1074
680 suppresses viremia in HIV-1-infected individuals. *Nat. Med.* **23**, 185–191 (2017).

681 19. J. F. Scheid, J. A. Horwitz, Y. Bar-On, E. F. Kreider, C. L. Lu, J. C. C. Lorenzi, A.
682 Feldmann, M. Braunschweig, L. Nogueira, T. Oliveira, I. Shimeliovich, R. Patel, L.
683 Burke, Y. Z. Cohen, S. Hadrigan, A. Settler, M. Witmer-Pack, A. P. West, B. Juelg, T.
684 Keler, T. Hawthorne, B. Zingman, R. M. Gulick, N. Pfeifer, G. H. Learn, M. S. Seaman,
685 P. J. Bjorkman, F. Klein, S. J. Schlesinger, B. D. Walker, B. H. Hahn, M. C.
686 Nussenzweig, M. Caskey, HIV-1 antibody 3BNC117 suppresses viral rebound in
687 humans during treatment interruption. *Nature* **535**, 556–560 (2016).

688 20. Y. Bar-On, H. Gruell, T. Schoofs, J. A. Pai, L. Nogueira, A. L. Butler, K. Millard, C.
689 Lehmann, I. Suárez, T. Y. Oliveira, T. Karagounis, Y. Z. Cohen, C. Wyen, S. Scholten,
690 L. Handl, S. Belblidia, J. P. Dizon, J. J. Vehreschild, M. Witmer-Pack, I. Shimeliovich, K.
691 Jain, K. Fiddike, K. E. Seaton, N. L. Yates, J. Horowitz, R. M. Gulick, N. Pfeifer, G. D.
692 Tomaras, M. S. Seaman, G. Fätkenheuer, M. Caskey, F. Klein, M. C. Nussenzweig,
693 Safety and antiviral activity of combination HIV-1 broadly neutralizing antibodies in
694 viremic individuals. *Nat. Med.* **24**, 1701–1707 (2018).

695 21. P. Mendoza, H. Gruell, L. Nogueira, J. A. Pai, A. L. Butler, K. Millard, C. Lehmann, I.
696 Suárez, T. Y. Oliveira, J. C. C. Lorenzi, Y. Z. Cohen, C. Wyen, T. Kümmerle, T.
697 Karagounis, C. L. Lu, L. Handl, C. Unson-O'Brien, R. Patel, C. Ruping, M. Schlotz, M.
698 Witmer-Pack, I. Shimeliovich, G. Kremer, E. Thomas, K. E. Seaton, J. Horowitz, A. P.
699 West, P. J. Bjorkman, G. D. Tomaras, R. M. Gulick, N. Pfeifer, G. Fätkenheuer, M. S.
700 Seaman, F. Klein, M. Caskey, M. C. Nussenzweig, Combination therapy with anti-HIV-1
701 antibodies maintains viral suppression. *Nature* **561**, 479–484 (2018).

702 22. C. Gaebler, L. Nogueira, E. Stoffel, T. Y. Oliveira, G. Breton, K. G. Millard, M.

703 Turroja, A. Butler, V. Ramos, M. S. Seaman, J. D. Reeves, C. J. Petroupoulos, I.
704 Shimeliovich, A. Gazumyan, C. S. Jiang, N. Jilg, J. F. Scheid, R. Gandhi, B. D. Walker,
705 M. C. Sneller, A. Fauci, T.-W. Chun, M. Caskey, M. C. Nussenzweig, Prolonged viral
706 suppression with anti-HIV-1 antibody therapy. *Nature* (2022), doi:10.1038/s41586-022-
707 04597-1.

708 23. R. M. Lynch, E. Boritz, E. E. Coates, A. DeZure, P. Madden, P. Costner, M. E.
709 Enama, S. Plummer, L. Holman, C. S. Hendel, I. Gordon, J. Casazza, M. Conan-Cibotti,
710 S. A. Migueles, R. Tressler, R. T. Bailer, A. McDermott, S. Narpala, S. O'Dell, G. Wolf,
711 J. D. Lifson, B. A. Freemire, R. J. Gorelick, J. P. Pandey, S. Mohan, N. Chomont, R.
712 Fromentin, T. W. Chun, A. S. Fauci, R. M. Schwartz, R. A. Koup, D. C. Douek, Z. Hu, E.
713 Capparelli, B. S. Graham, J. R. Mascola, J. E. Ledgerwood, Virologic effects of broadly
714 neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Sci. Transl.*
715 *Med.* **7**, 1–15 (2015).

716 24. B. Julg, K. E. Stephenson, K. Wagh, S. C. Tan, R. Zash, S. Walsh, J. Ansel, D.
717 Kanjilal, J. Nkolola, V. E. K. Walker-Sperling, J. Ophel, K. Yanosick, E. N. Borducchi, L.
718 Maxfield, P. Abbink, L. Peter, N. L. Yates, M. S. Wesley, T. Hassell, H. C. Gelderblom,
719 A. DeCamp, B. T. Mayer, A. Sato, M. W. Gerber, E. E. Giorgi, L. Gama, R. A. Koup, J.
720 R. Mascola, A. Monczor, S. Lupo, C.-P. Rolle, R. Arduino, E. DeJesus, G. D. Tomaras,
721 M. S. Seaman, B. Korber, D. H. Barouch, Safety and antiviral activity of triple
722 combination broadly neutralizing monoclonal antibody therapy against HIV-1: a phase 1
723 clinical trial. *Nat. Med.* **27**, 1718–1724 (2022).

724 25. K. E. Stephenson, B. Julg, C. S. Tan, R. Zash, S. R. Walsh, C. P. Rolle, A. N.
725 Monczor, S. Lupo, H. C. Gelderblom, J. L. Ansel, D. G. Kanjilal, L. F. Maxfield, J.
726 Nkolola, E. N. Borducchi, P. Abbink, J. Liu, L. Peter, A. Chandrashekar, R.
727 Nityanandam, Z. Lin, A. Setaro, J. Sapiente, Z. Chen, L. Sunner, T. Cassidy, C.
728 Bennett, A. Sato, B. Mayer, A. S. Perelson, A. deCamp, F. H. Priddy, K. Wagh, E. E.
729 Giorgi, N. L. Yates, R. C. Arduino, E. DeJesus, G. D. Tomaras, M. S. Seaman, B.
730 Korber, D. H. Barouch, Safety, pharmacokinetics and antiviral activity of PGT121, a
731 broadly neutralizing monoclonal antibody against HIV-1: a randomized, placebo-
732 controlled, phase 1 clinical trial. *Nat. Med.* **27**, 1718–1724 (2021).

733 26. L. Corey, P. B. Gilbert, M. Juraska, D. C. Montefiori, L. Morris, S. T. Karuna, S.
734 Edupuganti, N. M. Mgodi, A. C. deCamp, E. Rudnicki, Y. Huang, P. Gonzales, R.
735 Cabello, C. Orrell, J. R. Lama, F. Laher, E. M. Lazarus, J. Sanchez, I. Frank, J.
736 Hinojosa, M. E. Sobieszczyk, K. E. Marshall, P. G. Mukwekwerere, J. Makhema, L. R.
737 Baden, J. I. Mullins, C. Williamson, J. Hural, M. J. McElrath, C. Bentley, S. Takuva, M.
738 M. Gomez Lorenzo, D. N. Burns, N. Espy, A. K. Randhawa, N. Kochar, E. Piwowar-
739 Manning, D. J. Donnell, N. Sista, P. Andrew, J. G. Kublin, G. Gray, J. E. Ledgerwood, J.
740 R. Mascola, M. S. Cohen, Two Randomized Trials of Neutralizing Antibodies to Prevent
741 HIV-1 Acquisition. *N. Engl. J. Med.* **384**, 1003–1014 (2021).

742 27. D. R. Burton, Amping up HIV antibodies. *Science (80-.).* **372**, 1397–1398 (2021).

743 28. J. Huang, B. H. Kang, E. Ishida, T. Zhou, T. Griesman, Z. Sheng, F. Wu, N. A.
744 Doria-Rose, B. Zhang, K. McKee, S. O'Dell, G. Y. Chuang, A. Druz, I. S. Georgiev, C.
745 A. Schramm, A. Zheng, M. G. Joyce, M. Asokan, A. Ransier, S. Darko, S. A. Migueles,
746 R. T. Bailer, M. K. Louder, S. M. Alam, R. Parks, G. Kelsoe, T. Von Holle, B. F. Haynes,

747 D. C. Douek, V. Hirsch, M. S. Seaman, L. Shapiro, J. R. Mascola, P. D. Kwong, M.
748 Connors, Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan
749 Neutralization Breadth. *Immunity* **45**, 1108–1121 (2016).

750 29. M. M. Sajadi, A. Dashti, Z. Rikhtegaran Tehrani, W. D. Tolbert, M. S. Seaman, X.
751 Ouyang, N. Gohain, M. Pazgier, D. Kim, G. Cavet, J. Yared, R. R. Redfield, G. K. Lewis,
752 A. L. DeVico, Identification of Near-Pan-neutralizing Antibodies against HIV-1 by
753 Deconvolution of Plasma Humoral Responses. *Cell* **173**, 1783-1795.e14 (2018).

754 30. P. Schommers, H. Gruell, M. E. Abernathy, M. K. Tran, A. S. Dingens, H. B.
755 Gristick, C. O. Barnes, T. Schoofs, M. Schlotz, K. Vanshylla, C. Kreer, D. Weiland, U.
756 Holtick, C. Scheid, M. M. Valter, M. J. van Gils, R. W. Sanders, J. J. Vehreschild, O. A.
757 Cornely, C. Lehmann, G. Fätkenheuer, M. S. Seaman, J. D. Bloom, P. J. Bjorkman, F.
758 Klein, Restriction of HIV-1 Escape by a Highly Broad and Potent Neutralizing Antibody.
759 *Cell* **180**, 471-489.e22 (2020).

760 31. M. Sun, C. S. Pace, X. Yao, F. Yu, N. N. Padte, Y. Huang, M. S. Seaman, Q. Li, D.
761 D. Ho, Rational design and characterization of the novel, broad and potent bispecific
762 HIV-1 neutralizing antibody iMabm36. *J. Acquir. Immune Defic. Syndr.* **66**, 473–483
763 (2014).

764 32. J. J. Steinhardt, J. Guenaga, H. L. Turner, K. McKee, M. K. Louder, S. O'Dell, C. I.
765 Chiang, L. Lei, A. Galkin, A. K. Andrianov, N. A. Doria-Rose, R. T. Bailer, A. B. Ward, J.
766 R. Mascola, Y. Li, Rational design of a trispecific antibody targeting the HIV-1 Env with
767 elevated anti-viral activity. *Nat. Commun.* **9**, 1–12 (2018).

768 33. E. Rujas, H. Cui, J. Burnie, C. B. Aschner, T. Zhao, S. Insausti, K. Muthuraman, A.
769 Semesi, J. Ophel, J. L. Nieva, M. S. Seaman, C. Guzzo, B. Treanor, J.-P. Julien,
770 Engineering pan–HIV-1 neutralization potency through multispecific antibody avidity.
771 *Proc. Natl. Acad. Sci.* **119**, 1–11 (2022).

772 34. L. Xu, A. Pegu, E. Rao, N. Doria-Rose, J. Beninga, K. McKee, D. M. Lord, R. R.
773 Wei, G. Deng, M. Louder, S. D. Schmidt, Z. Mankoff, L. Wu, M. Asokan, C. Beil, C.
774 Lange, W. D. Leuschner, J. Kruip, R. Sendak, Y. Do Kwon, T. Zhou, X. Chen, R. T.
775 Bailer, K. Wang, M. Choe, L. J. Tartaglia, D. H. Barouch, S. O'Dell, J. P. Todd, D. R.
776 Burton, M. Roederer, M. Connors, R. A. Koup, P. D. Kwong, Z. Y. Yang, J. R. Mascola,
777 G. J. Nabel, Trispecific broadly neutralizing HIV antibodies mediate potent SHIV
778 protection in macaques. *Science (80-.).* **358**, 85–90 (2017).

779 35. J. F. Scheid, H. Mouquet, B. Ueberheide, R. Diskin, F. Klein, T. Y. K. Oliveira, J.
780 Pietzsch, D. Fenyo, A. Abadir, K. Velinzon, A. Hurley, S. Myung, F. Boulad, P.
781 Poignard, D. R. Burton, F. Pereyra, D. D. Ho, B. D. Walker, M. S. Seaman, P. J.
782 Bjorkman, B. T. Chait, M. C. Nussenzweig, Sequence and Structural Convergence.
783 *Science* **333**, 1633–1637 (2011).

784 36. L. M. Walker, M. Huber, K. J. Doores, E. Falkowska, R. Pejchal, J. P. Julien, S. K.
785 Wang, A. Ramos, P. Y. Chan-Hui, M. Moyle, J. L. Mitcham, P. W. Hammond, O. A.
786 Olsen, P. Phung, S. Fling, C. H. Wong, S. Phogat, T. Wrin, M. D. Simek, W. C. Koff, I.
787 A. Wilson, D. R. Burton, P. Poignard, Broad neutralization coverage of HIV by multiple
788 highly potent antibodies. *Nature* **477**, 466–470 (2011).

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

790 Wrin, M. D. Simek, S. Fling, J. L. Mitcham, J. K. Lehrman, F. H. Priddy, O. A. Olsen, S.
791 M. Frey, P. W. Hammond, S. Kaminsky, T. Zamb, M. Moyle, W. C. Koff, P. Poignard, D.
792 R. Burton, Broad and Potent Neutralizing Antibodies from an African Donor Reveal a
793 New HIV-1 Vaccine Target. *Science (80-.).* **326**, 285–289 (2009).

794 38. X. Wu, Z.-Y. Yang, Y. Li, C.-M. Hogerkorp, W. R. Schief, M. S. Seaman, T. Zhou, S.
795 D. Schmidt, L. Wu, L. Xu, N. S. Longo, K. McKee, S. O'Dell, M. K. Louder, D. L. Wycuff,
796 Y. Feng, M. Nason, N. Doria-Rose, M. Connors, P. D. Kwong, M. Roederer, R. T.
797 Wyatt, G. J. Nabel, J. R. Mascola, Rational Design of Envelope Identifies Broadly
798 Neutralizing Human Monoclonal Antibodies to HIV-1. *Science (80-.).* **329**, 856–861
799 (2010).

800 39. T. B. Kepler, H. X. Liao, S. M. Alam, R. Bhaskarabhatla, R. Zhang, C. Yandava, S.
801 Stewart, K. Anasti, G. Kelsoe, R. Parks, K. E. Lloyd, C. Stolarchuk, J. Pritchett, E.
802 Solomon, E. Friberg, L. Morris, S. S. A. Karim, M. S. Cohen, E. Walter, M. A. Moody, X.
803 Wu, H. R. Altae-Tran, I. S. Georgiev, P. D. Kwong, S. D. Boyd, A. Z. Fire, J. R. Mascola,
804 B. F. Haynes, Immunoglobulin gene insertions and deletions in the affinity maturation of
805 HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe* **16**, 304–313 (2014).

806 40. H. B. Gristick, L. Von Boehmer, A. P. West, M. Schamber, A. Gazumyan, J.
807 Golijanin, M. S. Seaman, G. Fätkenheuer, F. Klein, M. C. Nussenzweig, P. J. Bjorkman,
808 Natively glycosylated HIV-1 Env structure reveals new mode for antibody recognition of
809 the CD4-binding site. *Nat. Struct. Mol. Biol.* **23**, 906–915 (2016).

810 41. T. Zhou, I. Georgiev, X. Wu, Z. Y. Yang, K. Dai, A. Finzi, Y. Do Kwon, J. F. Scheid,
811 W. Shi, L. Xu, Y. Yang, J. Zhu, M. C. Nussenzweig, J. Sodroski, L. Shapiro, G. J. Nabel,
812 J. R. Mascola, P. D. Kwong, Structural basis for broad and potent neutralization of HIV-
813 1 by antibody VRC01. *Science (80-.).* **329**, 811–817 (2010).

814 42. D. T. MacLeod, N. M. Choi, B. Briney, F. Garces, L. S. Ver, E. Landais, B. Murrell,
815 T. Wrin, W. Kilembe, C. H. Liang, A. Ramos, C. B. Bian, L. Wickramasinghe, L. Kong,
816 K. Eren, C. Y. Wu, C. H. Wong, S. L. Kosakovsky Pond, I. A. Wilson, D. R. Burton, P.
817 Poignard, The IAVI Protocol C Investigators, Early Antibody Lineage Diversification and
818 Independent Limb Maturation Lead to Broad HIV-1 Neutralization Targeting the Env
819 High-Mannose Patch. *Immunity* **44**, 1215–1226 (2016).

820 43. V. Pascual, K. Victor, D. Lelsz, M. B. Spellerberg, T. J. Hamblin, K. M. Thompson, I.
821 Randen, J. Natvig, J. D. Capra, F. K. Stevenson, Nucleotide sequence analysis of the V
822 regions of two IgM cold agglutinins: Evidence that the V(H)4-21 gene segment is
823 responsible for the major cross-reactive idiotype. *J. Immunol.* **146**, 4385–43891 (1991).

824 44. R. F. van Vollenhoven, M. M. Bieber, M. J. Powell, P. K. Gupta, N. M. Bhat, K. L.
825 Richards, S. A. Albano, N. N. Teng, VH4-34 encoded antibodies in systemic lupus
826 erythematosus: a specific diagnostic marker that correlates with clinical disease
827 characteristics. *J. Rheumatol.* **26**, 1727–1733 (1999).

828 45. H. Wardemann, S. Yurasov, A. Schaefer, J. W. Young, E. Meffre, M. C.
829 Nussenzweig, Predominant autoantibody production by early human B cell precursors.
830 *Science (80-.).* **301**, 1374–1377 (2003).

831 46. B. F. Haynes, J. Fleming, E. W. St. Clair, H. Katinger, G. Stiegler, R. Kunert, J.
832 Robinson, R. M. Scearce, K. Plonk, H. F. Staats, T. L. Ortel, H. X. Liao, S. M. Alam,

833 Immunology: Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1
834 antibodies. *Science (80-.).* **308**, 1906–1908 (2005).

835 47. M. Liu, G. Yang, K. Wiehe, N. I. Nicely, N. A. Vandergrift, W. Rountree, M.
836 Bonsignori, S. M. Alam, J. Gao, B. F. Haynes, G. Kelsoe, Polyreactivity and
837 Autoreactivity among HIV-1 Antibodies. *J. Virol.* **89**, 784–798 (2015).

838 48. B. F. Haynes, G. Kelsoe, S. C. Harrison, T. B. Kepler, B-cell-lineage immunogen
839 design in vaccine development with HIV-1 as a case study. *Nat. Biotechnol.* **30**, 423–
840 433 (2012).

841 49. F. A. Tucci, S. Kitanovski, P. Johansson, L. Klein-Hitpass, A. Kahraman, J. Dürig, D.
842 Hoffmann, R. Küppers, Biased IGH VDJ gene repertoire and clonal expansions in B
843 cells of chronically hepatitis C virus–infected individuals. *Blood* **131**, 546–557 (2018).

844 50. H. X. Liao, R. Lynch, T. Zhou, F. Gao, S. Munir Alam, S. D. Boyd, A. Z. Fire, K. M.
845 Roskin, C. A. Schramm, Z. Zhang, J. Zhu, L. Shapiro, J. C. Mullikin, S. Gnanakaran, P.
846 Hraber, K. Wiehe, G. Kelsoe, G. Yang, S. M. Xia, D. C. Montefiori, R. Parks, K. E.
847 Lloyd, R. M. Scearce, K. A. Soderberg, M. Cohen, G. Kamanga, M. K. Louder, L. M.
848 Tran, Y. Chen, F. Cai, S. Chen, S. Moquin, X. Du, M. Gordon Joyce, S. Srivatsan, B.
849 Zhang, A. Zheng, G. M. Shaw, B. H. Hahn, T. B. Kepler, B. T. M. Korber, P. D. Kwong,
850 J. R. Mascola, B. F. Haynes, J. Becker, B. Benjamin, R. Blakesley, G. Bouffard, S.
851 Brooks, H. Coleman, M. Dekhtyar, M. Gregory, X. Guan, J. Gupta, J. Han, A. Hargrove,
852 S. L. Ho, T. Johnson, R. Legaspi, S. Lovett, Q. Maduro, C. Masiello, B. Maskeri, J.
853 McDowell, C. Montemayor, J. Mulliki, M. Park, N. Riebow, K. Schandler, B. Schmidt, C.
854 Sison, M. Stantripop, J. Thomas, P. Thomas, M. Vemulapalli, A. Young, Co-evolution of
855 a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469–476 (2013).

856 51. B. Briney, D. Sok, J. G. Jardine, D. W. Kulp, P. Skog, S. Menis, R. Jacak, O.
857 Kalyuzhniy, N. de Val, F. Sesterhenn, K. M. Le, A. Ramos, M. Jones, K. L. Saye-
858 Francisco, T. R. Blane, S. Spencer, E. Georgeson, X. Hu, G. Ozorowski, Y. Adachi, M.
859 Kubitz, A. Sarkar, I. A. Wilson, A. B. Ward, D. Nemazee, D. R. Burton, W. R. Schief,
860 Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies. *Cell*
861 **166**, 1459-1470.e11 (2016).

862 52. J. M. Steichen, Y. C. Lin, C. Havenar-Daughton, S. Pecetta, G. Ozorowski, J. R.
863 Willis, L. Toy, D. Sok, A. Liguori, S. Kratochvil, J. L. Torres, O. Kalyuzhniy, E. Melzi, D.
864 W. Kulp, S. Raemisch, X. Hu, S. M. Bernard, E. Georgeson, N. Phelps, Y. Adachi, M.
865 Kubitz, E. Landais, J. Umotoy, A. Robinson, B. Briney, I. A. Wilson, D. R. Burton, A. B.
866 Ward, S. Crotty, F. D. Batista, W. R. Schief, A generalized HIV vaccine design strategy
867 for priming of broadly neutralizing antibody responses. *Science (80-.).* **366** (2019),
868 doi:10.1126/science.aax4380.

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

873 54. C. Havenar-Daughton, A. Sarkar, D. W. Kulp, L. Toy, X. Hu, I. Deresa, O.
874 Kalyuzhniy, K. Kaushik, A. A. Upadhyay, S. Menis, E. Landais, L. Cao, J. K. Diedrich, S.
875 Kumar, T. Schiffner, S. M. Reiss, G. Seumois, J. R. Yates, J. C. Paulson, S. E.

876 Bosinger, I. A. Wilson, W. R. Schief, S. Crotty, The human naive B cell repertoire
877 contains distinct subclasses for a germline-targeting HIV-1 vaccine immunogen. *Sci.*
878 *Transl. Med.* **10**, 1–16 (2018).

879 55. K. Wiehe, T. Bradley, R. R. Meyerhoff, C. Hart, W. B. Williams, D. Easterhoff, W. J.
880 Faison, T. B. Kepler, K. O. Saunders, S. M. Alam, M. Bonsignori, B. F. Haynes,
881 Functional Relevance of Improbable Antibody Mutations for HIV Broadly Neutralizing
882 Antibody Development. *Cell Host Microbe* **23**, 759-765.e6 (2018).

883 56. S. D. Boyd, B. A. Gaëta, K. J. Jackson, A. Z. Fire, E. L. Marshall, J. D. Merker, J. M.
884 Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen,
885 K. C. Nadeau, M. Egholm, D. B. Miklos, J. L. Zehnder, A. M. Collins, Individual Variation
886 in the Germline Ig Gene Repertoire Inferred from Variable Region Gene
887 Rearrangements. *J. Immunol.* **184**, 6986–6992 (2010).

888 57. L. D. Goldstein, Y. J. J. Chen, J. Wu, S. Chaudhuri, Y. C. Hsiao, K. Schneider, K. H.
889 Hoi, Z. Lin, S. Guerrero, B. S. Jaiswal, J. Stinson, A. Antony, K. B. Pahuja, D.
890 Seshasayee, Z. Modrusan, I. Hötzel, S. Seshagiri, Massively parallel single-cell B-cell
891 receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies.
892 *Commun. Biol.* **2**, 1–10 (2019).

893 58. B. Briney, A. Inderbitzin, C. Joyce, D. R. Burton, Commonality despite exceptional
894 diversity in the baseline human antibody repertoire. *Nature* **566**, 393–397 (2019).

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

898 60. D. Sok, D. R. Burton, Recent progress in broadly neutralizing antibodies to HIV. *Nat.*
899 *Immunol.* **19**, 1179–1188 (2018).

900 61. Y. Elhanati, Z. Sethna, Q. Marcou, C. G. Callan, T. Mora, A. M. Walczak, Inferring
901 processes underlying B-cell repertoire diversity. *Philos. Trans. R. Soc. B Biol. Sci.* **370**
902 (2015), doi:10.1098/rstb.2014.0243.

903 62. Q. Marcou, T. Mora, A. M. Walczak, High-throughput immune repertoire analysis
904 with IGoR. *Nat. Commun.* **9** (2018), doi:10.1038/s41467-018-02832-w.

905 63. L. Cooper, K. L. Good-Jacobson, Dysregulation of humoral immunity in chronic
906 infection. *Immunol. Cell Biol.* **98**, 456–466 (2020).

907 64. A. deCamp, P. Hraber, R. T. Bailer, M. S. Seaman, C. Ochsenbauer, J. Kappes, R.
908 Gottardo, P. Edlefsen, S. Self, H. Tang, K. Greene, H. Gao, X. Daniell, M. Sarzotti-
909 Kelsoe, M. K. Gorny, S. Zolla-Pazner, C. C. LaBranche, J. R. Mascola, B. T. Korber, D.
910 C. Montefiori, Global Panel of HIV-1 Env Reference Strains for Standardized
911 Assessments of Vaccine-Elicited Neutralizing Antibodies. *J. Virol.* **88**, 2489–2507
912 (2014).

913 65. D. H. Barouch, J. B. Whitney, B. Moldt, F. Klein, T. Y. Oliveira, J. Liu, K. E.
914 Stephenson, H. W. Chang, K. Shekhar, S. Gupta, J. P. Nkolola, M. S. Seaman, K. M.
915 Smith, E. N. Borducchi, C. Cabral, J. Y. Smith, S. Blackmore, S. Sanisetty, J. R. Perry,
916 M. Beck, M. G. Lewis, W. Rinaldi, A. K. Chakraborty, P. Poignard, M. C. Nussenzweig,
917 D. R. Burton, Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal

918 antibodies in SHIV-infected rhesus monkeys. *Nature* **503**, 224–228 (2013).

919 66. L. M. Walker, D. R. Burton, Passive immunotherapy of viral infections: “super-
920 antibodies” enter the fray. *Nat. Rev. Immunol.* **18**, 297–308 (2018).

921 67. H. Gruell, F. Klein, Antibody-mediated prevention and treatment of HIV-1 infection.
922 *Retrovirology* **15**, 1–11 (2018).

923 68. P. D. Kwong, J. R. Mascola, HIV-1 Vaccines Based on Antibody Identification, B
924 Cell Ontogeny, and Epitope Structure. *Immunity* **48**, 855–871 (2018).

925 69. K. O. Saunders, K. Wiehe, M. Tian, P. Acharya, T. Bradley, S. Munir Alam, E. P.
926 Go, R. Scearce, L. Sutherland, R. Henderson, A. L. Hsu, M. J. Borgnia, H. Chen, X. Lu,
927 N. R. Wu, B. Watts, C. Jiang, D. Easterhoff, H. L. Cheng, K. McGovern, P. Waddicor, A.
928 Chapdelaine-Williams, A. Eaton, J. Zhang, W. Rountree, L. Verkoczy, M. Tomai, M. G.
929 Lewis, H. R. Desaire, R. J. Edwards, D. W. Cain, M. Bonsignori, D. Montefiori, F. W. Alt,
930 B. F. Haynes, Targeted selection of HIV-specific antibody mutations by engineering B
931 cell maturation. *Science (80-.).* **366** (2019), doi:10.1126/science.aay7199.

932 70. F. Klein, R. Diskin, J. F. Scheid, C. Gaebler, H. Mouquet, I. S. Georgiev, M.
933 Pancera, T. Zhou, R. B. Incesu, B. Z. Fu, P. N. P. Gnanapragasam, T. Y. Oliveira, M. S.
934 Seaman, P. D. Kwong, P. J. Bjorkman, M. C. Nussenzweig, Somatic mutations of the
935 immunoglobulin framework are generally required for broad and potent HIV-1
936 neutralization. *Cell* **153**, 126–138 (2013).

937 71. J. P. Julien, J. H. Lee, A. Cupo, C. D. Murin, R. Derking, S. Hoffenberg, M. J.
938 Caulfield, C. R. King, A. J. Marozsan, P. J. Klasse, R. W. Sanders, J. P. Moore, I. A.
939 Wilson, A. B. Ward, Asymmetric recognition of the HIV-1 trimer by broadly neutralizing
940 antibody PG9. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 4351–4356 (2013).

941 72. J. S. McLellan, M. Pancera, C. Carrico, J. Gorman, J. P. Julien, R. Khayat, R.
942 Louder, R. Pejchal, M. Sastry, K. Dai, S. O’Dell, N. Patel, S. Shahzad-Ul-Hussan, Y.
943 Yang, B. Zhang, T. Zhou, J. Zhu, J. C. Boyington, G. Y. Chuang, D. Diwanji, I.
944 Georgiev, Y. Do Kwon, D. Lee, M. K. Louder, S. Moquin, S. D. Schmidt, Z. Y. Yang, M.
945 Bonsignori, J. A. Crump, S. H. Kapiga, N. E. Sam, B. F. Haynes, D. R. Burton, W. C.
946 Koff, L. M. Walker, S. Phogat, R. Wyatt, J. Orwenyo, L. X. Wang, J. Arthos, C. A.
947 Bewley, J. R. Mascola, G. J. Nabel, W. R. Schief, A. B. Ward, I. A. Wilson, P. D. Kwong,
948 Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*
949 **480**, 336–343 (2011).

950 73. R. Pejchal, L. M. Walker, R. L. Stanfield, S. K. Phogat, W. C. Koff, P. Poignard, D.
951 R. Burton, I. A. Wilson, Structure and function of broadly reactive antibody PG16 reveal
952 an H3 subdomain that mediates potent neutralization of HIV-1. *Proc. Natl. Acad. Sci. U.*
953 *S. A.* **107**, 11483–11488 (2010).

954 74. M. Bonsignori, E. F. Kreider, D. Fera, R. R. Meyerhoff, T. Bradley, K. Wiehe, S. M.
955 Alam, B. Aussedat, W. E. Walkowicz, K. K. Hwang, K. O. Saunders, R. Zhang, M. A.
956 Gladden, A. Monroe, A. Kumar, S. M. Xia, M. Cooper, M. K. Louder, K. McKee, R. T.
957 Bailer, B. W. Pier, C. A. Jette, G. Kelsoe, W. B. Williams, L. Morris, J. Kappes, K. Wagh,
958 G. Kamanga, M. S. Cohen, P. T. Hraber, D. C. Montefiori, A. Trama, H. X. Liao, T. B.
959 Kepler, M. A. Moody, F. Gao, S. J. Danishefsky, J. R. Mascola, G. M. Shaw, B. H.
960 Hahn, S. C. Harrison, B. T. Korber, B. F. Haynes, Staged induction of HIV-1 glycan-

961 dependent broadly neutralizing antibodies. *Sci. Transl. Med.* **9** (2017),
962 doi:10.1126/scitranslmed.aai7514.

963 75. D. Cizmeci, G. Lofano, E. Rossignol, A. S. Dugast, D. Kim, G. Cavet, N. Nguyen, Y.
964 C. Tan, M. S. Seaman, G. Alter, B. Julg, Distinct clonal evolution of b-cells in hiv
965 controllers with neutralizing antibody breadth. *Elife* **10**, 1–15 (2021).

966 76. N. A. Doria-Rose, C. A. Schramm, J. Gorman, P. L. Moore, J. N. Bhiman, B. J.
967 DeKosky, M. J. Ernandes, I. S. Georgiev, H. J. Kim, M. Pancera, R. P. Staupe, H. R.
968 Altae-Tran, R. T. Bailer, E. T. Crooks, A. Cupo, A. Druz, N. J. Garrett, K. H. Hoi, R.
969 Kong, M. K. Louder, N. S. Longo, K. McKee, M. Nonyane, S. O'Dell, R. S. Roark, R. S.
970 Rudicell, S. D. Schmidt, D. J. Sheward, C. Soto, C. K. Wibmer, Y. Yang, Z. Zhang, J. C.
971 Mullikin, J. M. Binley, R. W. Sanders, I. A. Wilson, J. P. Moore, A. B. Ward, G.
972 Georgiou, C. Williamson, S. S. A. Karim, L. Morris, P. D. Kwong, L. Shapiro, J. R.
973 Mascola, Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies.
974 *Nature* **508**, 55–62 (2014).

975 77. P. L. Moore, E. S. Gray, D. Sheward, M. Madiga, N. Ranchobe, Z. Lai, W. J.
976 Honnen, M. Nonyane, N. Tumba, T. Hermanus, S. Sibeko, K. Mlisana, S. S. Abdool
977 Karim, C. Williamson, A. Pinter, L. Morris, Potent and Broad Neutralization of HIV-1
978 Subtype C by Plasma Antibodies Targeting a Quaternary Epitope Including Residues in
979 the V2 Loop. *J. Virol.* **85**, 3128–3141 (2011).

980 78. C. K. Wibmer, J. N. Bhiman, E. S. Gray, N. Tumba, S. S. Abdool Karim, C.
981 Williamson, L. Morris, P. L. Moore, Viral Escape from HIV-1 Neutralizing Antibodies
982 Drives Increased Plasma Neutralization Breadth through Sequential Recognition of
983 Multiple Epitopes and Immunotypes. *PLoS Pathog.* **9** (2013),
984 doi:10.1371/journal.ppat.1003738.

985 79. K. M. Roskin, K. J. L. Jackson, J. Y. Lee, R. A. Hoh, S. A. Joshi, K. K. Hwang, M.
986 Bonsignori, I. Pedroza-Pacheco, H. X. Liao, M. A. Moody, A. Z. Fire, P. Borrow, B. F.
987 Haynes, S. D. Boyd, Aberrant B cell repertoire selection associated with HIV
988 neutralizing antibody breadth. *Nat. Immunol.* **21**, 199–209 (2020).

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

993 81. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable
994 domain sequence analysis tool. *Nucleic Acids Res.* (2013), doi:10.1093/nar/gkt382.

995 82. M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, R. Lopez, A
996 new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* **38**, 695–
997 699 (2010).

998 83. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H.
999 McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, Fast, scalable
1000 generation of high-quality protein multiple sequence alignments using Clustal Omega.
1001 *Mol. Syst. Biol.* (2011), doi:10.1038/msb.2011.75.

1002 84. J. A. Vander Heiden, G. Yaari, M. Uduman, J. N. H. Stern, K. C. O'connor, D. A.

1003 Hafler, F. Vigneault, S. H. Kleinstein, PRESTO: A toolkit for processing high-throughput
1004 sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–
1005 1932 (2014).

1006 85. M.-P. Lefranc, V. Giudicelli, C. Ginestoux, J. Bodmer, W. Muller, R. Bontrop, M.
1007 Lemaitre, A. Malik, V. Barbie, D. Chaume, IMGT, the international ImMunoGeneTics
1008 database. *Nucleic Acids Res.* **27**, 209–212 (1999).

1009 86. M.-P. Lefranc, From IMGT-ONTOLOGY IDENTIFICATION axiom to IMGT
1010 standardized keywords: for immunoglobulins (IG), T cell receptors (TR), and
1011 conventional genes. *Cold Spring Harb. Protoc.* **2011**, 604–613 (2011).

1012 87. M. S. Seaman, H. Janes, N. Hawkins, L. E. Grandpre, C. Devoy, A. Giri, R. T.
1013 Coffey, L. Harris, B. Wood, M. G. Daniels, T. Bhattacharya, A. Lapedes, V. R. Polonis,
1014 F. E. McCutchan, P. B. Gilbert, S. G. Self, B. T. Korber, D. C. Montefiori, J. R. Mascola,
1015 Tiered Categorization of a Diverse Panel of HIV-1 Env Pseudoviruses for Assessment
1016 of Neutralizing Antibodies. *J. Virol.* **84**, 1439–1452 (2010).

1017 88. S. R. Walsh, M. S. Seaman, Broadly Neutralizing Antibodies for HIV-1 Prevention.
1018 *Front. Immunol.* **12**, 1–14 (2021).

1019 89. W. S. DeWitt, P. Lindau, T. M. Snyder, A. M. Sherwood, M. Vignali, C. S. Carlson,
1020 P. D. Greenberg, N. Duerkopp, R. O. Emerson, H. S. Robins, A public database of
1021 memory and naive B-cell receptor sequences. *PLoS One* **11**, 1–18 (2016).

1022 90. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis
1023 of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

1024 91. A. Nourmohammad, J. Otwinowski, M. Łuksza, T. Mora, A. M. Walczak, Fierce
1025 Selection and Interference in B-Cell Repertoire Response to Chronic HIV-1. *Mol. Biol.*
1026 *Evol.* **36**, 2184–2194 (2019).

1027 92. M. Bonsignori, K.-K. Hwang, X. Chen, C.-Y. Tsao, L. Morris, E. Gray, D. J. Marshall,
1028 J. A. Crump, S. H. Kapiga, N. E. Sam, F. Sinangil, M. Pancera, Y. Yongping, B. Zhang,
1029 J. Zhu, P. D. Kwong, S. O'Dell, J. R. Mascola, L. Wu, G. J. Nabel, S. Phogat, M. S.
1030 Seaman, J. F. Whitesides, M. A. Moody, G. Kelsoe, X. Yang, J. Sodroski, G. M. Shaw,
1031 D. C. Montefiori, T. B. Kepler, G. D. Tomaras, S. M. Alam, H.-X. Liao, B. F. Haynes,
1032 Analysis of a Clonal Lineage of HIV-1 Envelope V2/V3 Conformational Epitope-Specific
1033 Broadly Neutralizing Antibodies and Their Inferred Unmutated Common Ancestors. *J.*
1034 *Virol.* **85**, 9998–10009 (2011).

1035 93. M. Bonsignori, T. Zhou, Z. Sheng, L. Chen, F. Gao, M. G. Joyce, G. Ozorowski, G.
1036 Y. Chuang, C. A. Schramm, K. Wiehe, S. M. Alam, T. Bradley, M. A. Gladden, K. K.
1037 Hwang, S. Iyengar, A. Kumar, X. Lu, K. Luo, M. C. Mangiapani, R. J. Parks, H. Song, P.
1038 Acharya, R. T. Bailer, A. Cao, A. Druz, I. S. Georgiev, Y. D. Kwon, M. K. Louder, B.
1039 Zhang, A. Zheng, B. J. Hill, R. Kong, C. Soto, J. C. Mullikin, D. C. Douek, D. C.
1040 Montefiori, M. A. Moody, G. M. Shaw, B. H. Hahn, G. Kelsoe, P. T. Hraber, B. T. Korber,
1041 S. D. Boyd, A. Z. Fire, T. B. Kepler, L. Shapiro, A. B. Ward, J. R. Mascola, H. X. Liao, P.
1042 D. Kwong, B. F. Haynes, Maturation Pathway from Germline to Broad HIV-1 Neutralizer
1043 of a CD4-Mimic Antibody. *Cell* **165**, 449–463 (2016).

1044 94. A. Buchacher, R. Predl, K. Strutzenberger, W. Steinfellner, A. Trkola, M. Purtscher,

1045 G. Gruber, C. Tauer, F. Steindl, A. Jungbauer, H. Katinger, Generation of Human
1046 Monoclonal Antibodies against HIV-1 Proteins; Electrofusion and Epstein-Barr Virus
1047 Transformation for Peripheral Blood Lymphocyte Immortalization. *AIDS Res. Hum.*
1048 *Retroviruses* **10**, 359–369 (1994).

1049 95. D. R. Burton, C. F. Barbas, M. A. A. Persson, S. Koenig, R. M. Changcr, R. A.
1050 Lerner, A large array of human monoclonal antibodies to type 1 human
1051 immunodeficiency virus from combinatorial libraries of asymptomatic seropositive
1052 individuals. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 10134–10137 (1991).

1053 96. E. M. Cale, J. Gorman, N. A. Radakovich, E. T. Crooks, K. Osawa, T. Tong, J. Li, R.
1054 Nagarajan, G. Ozorowski, D. R. Ambrozak, M. Asokan, R. T. Bailer, A. K. Bennici, X.
1055 Chen, N. A. Doria-Rose, A. Druz, Y. Feng, M. G. Joyce, M. K. Louder, S. O'Dell, C.
1056 Oliver, M. Pancera, M. Connors, T. J. Hope, T. B. Kepler, R. T. Wyatt, A. B. Ward, I. S.
1057 Georgiev, P. D. Kwong, J. R. Mascola, J. M. Binley, Virus-like Particles Identify an HIV
1058 V1V2 Apex-Binding Neutralizing Antibody that Lacks a Protruding Loop. *Immunity* **46**,
1059 777-791.e10 (2017).

1060 97. D. Corti, J. P. M. Langedijk, A. Hinz, M. S. Seaman, F. Vanzetta, B. M. Fernandez-
1061 Rodriguez, C. Silacci, D. Pinna, D. Jarrossay, S. Balla-Jhagjhoorsingh, B. Willems, M. J.
1062 Zekveld, H. Dreja, E. O'Sullivan, C. Pade, C. Orkin, S. A. Jeffs, D. C. Montefiori, D.
1063 Davis, W. Weissenhorn, Á. McKnight, J. L. Heeney, F. Sallusto, Q. J. Sattentau, R. A.
1064 Weiss, A. Lanzavecchia, Analysis of memory B cell responses and isolation of novel
1065 monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS*
1066 *One* **5** (2010), doi:10.1371/journal.pone.0008805.

1067 98. N. A. Doria-Rose, J. N. Bhiman, R. S. Roark, C. A. Schramm, J. Gorman, G.-Y.
1068 Chuang, M. Pancera, E. M. Cale, M. J. Ernandes, M. K. Louder, M. Asokan, R. T.
1069 Bailer, A. Druz, I. R. Fraschilla, N. J. Garrett, M. Jarosinski, R. M. Lynch, K. McKee, S.
1070 O'Dell, A. Pegu, S. D. Schmidt, R. P. Staupe, M. S. Sutton, K. Wang, C. K. Wibmer, B.
1071 F. Haynes, S. Abdool-Karim, L. Shapiro, P. D. Kwong, P. L. Moore, L. Morris, J. R.
1072 Mascola, New Member of the V1V2-Directed CAP256-VRC26 Lineage That Shows
1073 Increased Breadth and Exceptional Potency. *J. Virol.* **90**, 76–91 (2016).

1074 99. E. Falkowska, K. M. Le, A. Ramos, K. J. Doores, J. H. Lee, C. Blattner, A. Ramirez,
1075 R. Derking, M. J. vanGils, C. H. Liang, R. Mcbride, B. von Bredow, S. S. Shivatare, C.
1076 Y. Wu, P. Y. Chan-Hui, Y. Liu, T. Feizi, M. B. Zwick, W. C. Koff, M. S. Seaman, K.
1077 Swiderek, J. P. Moore, D. Evans, J. C. Paulson, C. H. Wong, A. B. Ward, I. A. Wilson,
1078 R. W. Sanders, P. Poignard, D. R. Burton, Broadly neutralizing HIV antibodies define a
1079 glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope
1080 trimers. *Immunity* **40**, 657–668 (2014).

1081 100. N. T. Freund, J. A. Horwitz, L. Nogueira, S. A. Sievers, L. Scharf, J. F. Scheid, A.
1082 Gazumyan, C. Liu, K. Velinzon, A. Goldenthal, R. W. Sanders, J. P. Moore, P. J.
1083 Bjorkman, M. S. Seaman, B. D. Walker, F. Klein, M. C. Nussenzweig, A New Glycan-
1084 Dependent CD4-Binding Site Neutralizing Antibody Exerts Pressure on HIV-1 In Vivo.
1085 *PLoS Pathog.* **11**, 1–19 (2015).

1086 101. N. T. Freund, H. Wang, L. Scharf, L. Nogueira, J. A. Horwitz, Y. Bar-On, J.
1087 Golijanin, S. A. Sievers, D. Sok, H. Cai, J. C. C. Lorenzi, A. Halper-Stromberg, I. Toth,

1088 A. Piechocka-Trocha, H. B. Gristick, M. J. Van Gils, R. W. Sanders, L. X. Wang, M. S.
1089 Seaman, D. R. Burton, A. Gazumyan, B. D. Walker, A. P. West, P. J. Bjorkman, M. C.
1090 Nussenzweig, Coexistence of potent HIV-1 broadly neutralizing antibodies and
1091 antibody-sensitive viruses in a viremic controller. *Sci. Transl. Med.* **9**, 1–14 (2017).

1092 102. F. Gao, M. Bonsignori, H. X. Liao, A. Kumar, S. M. Xia, X. Lu, F. Cai, K. K. Hwang,
1093 H. Song, T. Zhou, R. M. Lynch, S. M. Alam, M. A. Moody, G. Ferrari, M. Berrong, G.
1094 Kelsoe, G. M. Shaw, B. H. Hahn, D. C. Montefiori, G. Kamanga, M. S. Cohen, P.
1095 Hraber, P. D. Kwong, B. T. Korber, J. R. Mascola, T. B. Kepler, B. F. Haynes,
1096 Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell*
1097 **158**, 481–491 (2014).

1098 103. M. K. Gorny, C. Williams, B. Volsky, K. Revesz, X.-H. Wang, S. Burda, T. Kimura,
1099 F. A. J. Konings, A. Nádas, C. A. Anyangwe, P. Nyambi, C. Krachmarov, A. Pinter, S.
1100 Zolla-Pazner, Cross-Clade Neutralizing Activity of Human Anti-V3 Monoclonal
1101 Antibodies Derived from the Cells of Individuals Infected with Non-B Clades of Human
1102 Immunodeficiency Virus Type 1. *J. Virol.* **80**, 6865–6872 (2006).

1103 104. J. Huang, G. Ofek, L. Laub, M. K. Louder, N. A. Doria-Rose, N. S. Longo, H.
1104 Imamichi, R. T. Bailer, B. Chakrabarti, S. K. Sharma, S. M. Alam, T. Wang, Y. Yang, B.
1105 Zhang, S. A. Migueles, R. Wyatt, B. F. Haynes, P. D. Kwong, J. R. Mascola, M.
1106 Connors, Broad and potent neutralization of HIV-1 by a gp41-specific human antibody.
1107 *Nature* **491**, 406–412 (2012).

1108 105. J. Huang, B. H. Kang, M. Pancera, J. H. Lee, T. Tong, Y. Feng, H. Imamichi, I. S.
1109 Georgiev, G. Y. Chuang, A. Druz, N. A. Doria-Rose, L. Laub, K. Sliepen, M. J. Van Gils,
1110 A. T. De La Peña, R. Derking, P. J. Klasse, S. A. Migueles, R. T. Bailer, M. Alam, P.
1111 Pugach, B. F. Haynes, R. T. Wyatt, R. W. Sanders, J. M. Binley, A. B. Ward, J. R.
1112 Mascola, P. D. Kwong, M. Connors, Broad and potent HIV-1 neutralization by a human
1113 antibody that binds the gp41-gp120 interface. *Nature* **515**, 138–142 (2014).

1114 106. R. Kong, K. Xu, T. Zhou, P. Acharya, T. Lemmin, K. Liu, G. Ozorowski, C. Soto, J.
1115 D. Taft, R. T. Bailer, E. M. Cale, L. Chen, C. W. Choi, G.-Y. Chuang, N. A. Doria-Rose,
1116 A. Druz, I. S. Georgiev, J. Gorman, J. Huang, M. G. Joyce, M. K. Louder, X. Ma, K.
1117 McKee, S. O'Dell, M. Pancera, Y. Yang, S. C. Blanchard, W. Mothes, D. R. Burton, W.
1118 C. Koff, M. Connors, A. B. Ward, P. D. Kwong, J. R. Mascola, Fusion peptide of HIV-1
1119 as a site of vulnerability to neutralizing antibody. *Science* **352**, 828–33 (2016).

1120 107. Y. Li, S. O'Dell, R. Wilson, X. Wu, S. D. Schmidt, C.-M. Hogerkorp, M. K. Louder,
1121 N. S. Longo, C. Poulsen, J. Guenaga, B. K. Chakrabarti, N. Doria-Rose, M. Roederer,
1122 M. Connors, J. R. Mascola, R. T. Wyatt, HIV-1 Neutralizing Antibodies Display Dual
1123 Recognition of the Primary and Coreceptor Binding Sites and Preferential Binding to
1124 Fully Cleaved Envelope Glycoproteins. *J. Virol.* **86**, 11231–11241 (2012).

1125 108. H. Mouquet, L. Scharf, Z. Euler, Y. Liu, C. Eden, J. F. Scheid, A. Halper-
1126 Stromberg, P. N. P. Gnanapragasam, D. I. R. Spencer, M. S. Seaman, H.
1127 Schuitemaker, T. Feizi, M. C. Nussenzweig, P. J. Bjorkman, Complex-type N-glycan
1128 recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. U. S. A.*
1129 (2012), doi:10.1073/pnas.1217207109.

1130 109. R. S. Rudicell, Y. Do Kwon, S.-Y. Ko, A. Pegu, M. K. Louder, I. S. Georgiev, X.

1131 Wu, J. Zhu, J. C. Boyington, X. Chen, W. Shi, Z.-Y. Yang, N. A. Doria-Rose, K. McKee,
1132 S. O'Dell, S. D. Schmidt, G.-Y. Chuang, A. Druz, C. Soto, Y. Yang, B. Zhang, T. Zhou,
1133 J.-P. Todd, K. E. Lloyd, J. Eudailey, K. E. Roberts, B. R. Donald, R. T. Bailer, J.
1134 Ledgerwood, NISC Comparative Sequencing Program, J. C. Mullikin, L. Shapiro, R. A.
1135 Koup, B. S. Graham, M. C. Nason, M. Connors, B. F. Haynes, S. S. Rao, M. Roederer,
1136 P. D. Kwong, J. R. Mascola, G. J. Nabel, Enhanced potency of a broadly neutralizing
1137 HIV-1 antibody in vitro improves protection against lentiviral infection in vivo. *J. Virol.* **88**,
1138 12669–82 (2014).

1139 110. D. Sok, M. J. V. Gils, M. Pauthner, J. P. Julien, K. L. Saye-Francisco, J. Hsueh, B.
1140 Briney, J. H. Lee, K. M. Le, P. S. Lee, Y. Hua, M. S. Seaman, J. P. Moore, A. B. Ward,
1141 I. A. Wilson, R. W. Sanders, D. R. Burton, Recombinant HIV envelope trimer selects for
1142 quaternary-dependent antibodies targeting the trimer apex. *Proc. Natl. Acad. Sci. U. S.*
1143 *A.* **111**, 17624–17629 (2014).

1144 111. L. D. Williams, G. Ofek, S. Schätzle, J. R. McDaniel, X. Lu, N. I. Nicely, L. Wu, C.
1145 S. Lougheed, T. Bradley, M. K. Louder, K. McKee, R. T. Bailer, S. O'Dell, I. S. Georgiev,
1146 M. S. Seaman, R. J. Parks, D. J. Marshall, K. Anasti, G. Yang, X. Nie, N. L. Tumba, K.
1147 Wiehe, K. Wagh, B. Korber, T. B. Kepler, S. Munir Alam, L. Morris, G. Kamanga, M. S.
1148 Cohen, M. Bonsignori, S.-M. Xia, D. C. Montefiori, G. Kelsoe, F. Gao, J. R. Mascola, M.
1149 A. Moody, K. O. Saunders, H.-X. Liao, G. D. Tomaras, G. Georgiou, B. F. Haynes,
1150 Potent and broad HIV-neutralizing antibodies in memory B cells and plasma. *Sci.*
1151 *Immunol.* **2** (2017), doi:10.1126/sciimmunol.aal2200.

1152 112. T. Zhou, J. Zhu, X. Wu, S. Moquin, B. Zhang, P. Acharya, I. S. Georgiev, H. R.
1153 Altae-Tran, G.-Y. Chuang, M. G. Joyce, Y. Do Kwon, N. S. Longo, M. K. Louder, T.
1154 Luongo, K. McKee, C. A. Schramm, J. Skinner, Y. Yang, Z. Yang, Z. Zhang, A. Zheng,
1155 M. Bonsignori, B. F. Haynes, J. F. Scheid, M. C. Nussenzweig, M. Simek, D. R. Burton,
1156 W. C. Koff, NISC Comparative Sequencing Program, J. C. Mullikin, M. Connors, L.
1157 Shapiro, G. J. Nabel, J. R. Mascola, P. D. Kwong, Multidonor analysis reveals structural
1158 elements, genetic determinants, and maturation pathway for HIV-1 neutralization by
1159 VRC01-class antibodies. *Immunity* **39**, 245–58 (2013).

1160 113. T. Zhou, R. M. Lynch, L. Chen, P. Acharya, X. Wu, N. A. Doria-Rose, M. G. Joyce,
1161 D. Lingwood, C. Soto, R. T. Bailer, M. J. Ernandes, R. Kong, N. S. Longo, M. K. Louder,
1162 K. McKee, S. O'Dell, S. D. Schmidt, L. Tran, Z. Yang, A. Druz, T. S. Luongo, S. Moquin,
1163 S. Srivatsan, Y. Yang, B. Zhang, A. Zheng, M. Pancera, T. Kirys, I. S. Georgiev, T.
1164 Gindin, H.-P. Peng, A.-S. Yang, NISC Comparative Sequencing Program, J. C. Mullikin,
1165 M. D. Gray, L. Stamatatos, D. R. Burton, W. C. Koff, M. S. Cohen, B. F. Haynes, J. P.
1166 Casazza, M. Connors, D. Corti, A. Lanzavecchia, Q. J. Sattentau, R. A. Weiss, A. P.
1167 West, P. J. Bjorkman, J. F. Scheid, M. C. Nussenzweig, L. Shapiro, J. R. Mascola, P. D.
1168 Kwong, Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting the CD4
1169 Supersite in 14 Donors. *Cell* **161**, 1280–92 (2015).

1170

1171 **Acknowledgments:** We thank all members of the Klein Laboratory for continuous
1172 support and helpful discussion, Viera Kovacova, Antonios Papadakis, Lukas
1173 Maas, and Milos Nikolic for valuable advice on data evaluation and NGS pipeline
1174 programming, Peter Nürnberg, Janine Altmüller, and Christian Becker from the

Cologne Center for Genomics (CCG) for sequencing support, and Michael Lässig and Christa Stitz for valuable discussions and support within the CRC1310.

Funding:

German Research Foundation (DFG) CRC 1279 (FK)

German Research Foundation (DFG) CRC 1310 (FK, CK, AMW, AN, JG, AB)

German Center for Infection Research (DZIF) (FK)

European Research Council ERC-StG639961 (FK)

European Research Council COG 724208 (AMW)

Agence Nationale de la Recherche ANR-19-CE45-0018 “RESP-REP” (TM)

CAREER Award from the National Science Foundation grant 2045054 (AN)

MIRA award from the National Institutes of Health grant 1R35GM142795-01 (AN)

Author contributions:

Conceptualization: CK, TM, AMW, FK

Methodology: CK, CL, MSE, NS, JG, AB, LD, AN, TM, AMW, FK

Investigation: CK, LG, MSE;

Resources: LG, MS, LD, PS, HG

Formal analysis: CK, CL, NS, JG

Writing - original draft: CK, TM, AMW, FK

Writing - reviewing & editing: all authors

Visualization: CK

Supervision: AB, TM, AMW, FK

Funding acquisition: CK, AN, TM, AMW, FK

Competing interests: The authors have nothing to declare.

Data and materials availability: FASTQ files for all repertoires have been deposited at the Sequence Read Archive (SRA) with accession numbers SAMN29624595-713 (BioProject Accession Number PRJNA857338). IGoR is freely available from a github repository (https://github.com/statbiophys/IGoR). Upon request, the corresponding author will share any additional code that has been used to process the data. Requests for materials should be directed to the corresponding author and may be subject to restrictions based on data and privacy protection regulations and/or may require a Material Transfer Agreement (MTA).

Figures:

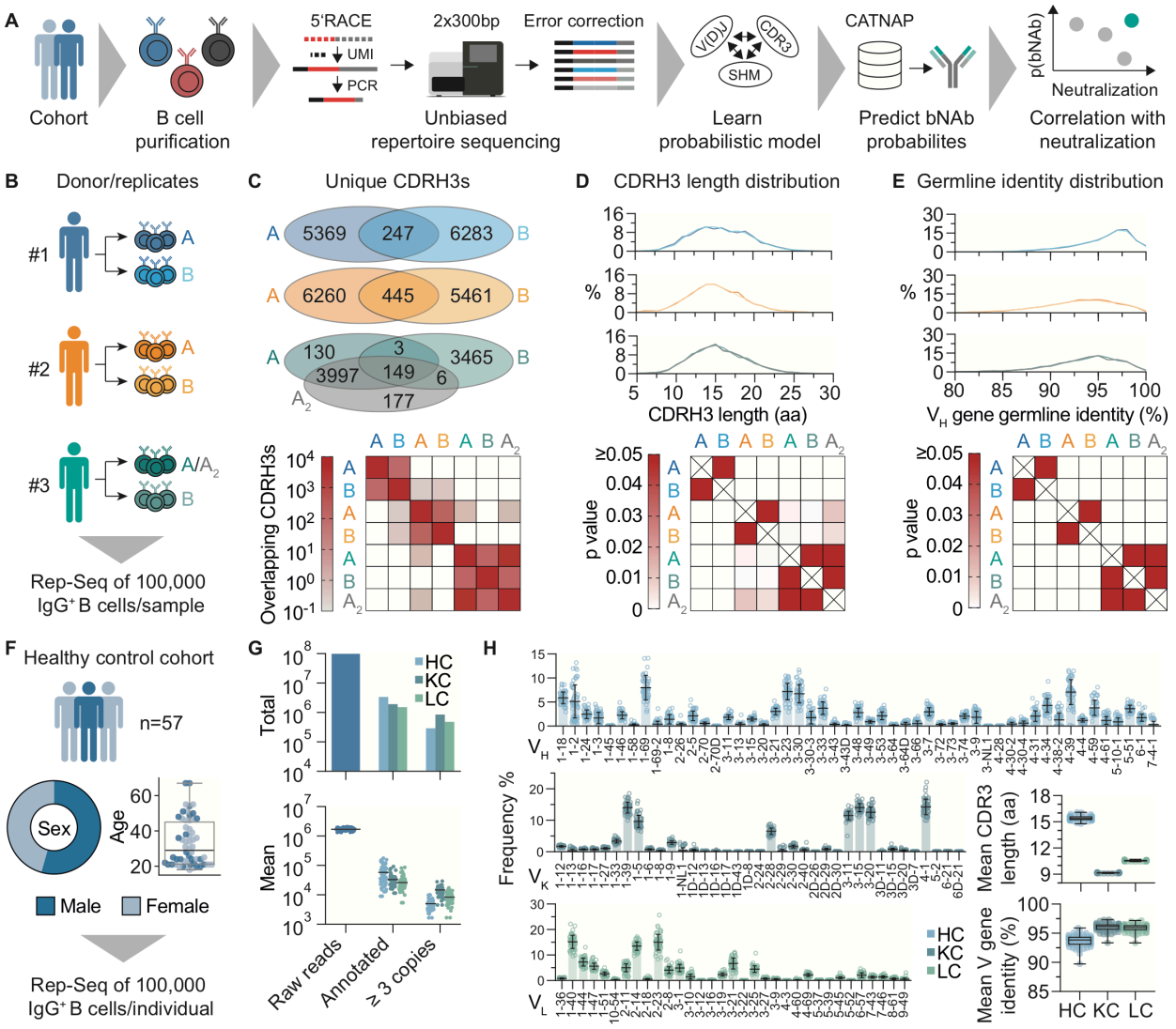

Fig. 1. Overall approach and unbiased repertoire sequencing. (**A**) B cells are purified from peripheral blood and total mRNA is isolated. 5' rapid amplification of cDNA ends (5'RACE) is performed to add unique molecular identifiers (UMI) and amplify variable regions. Amplicons are sequenced by 2x300 bp Illumina sequencing and UMIs are exploited for error correction. High-quality sequences are used to learn probabilistic models and predict probabilities of bNAb sequences, which are derived from the CATNAP database. Correlation of probabilities and neutralization allows for identifying highly probable and potent bNAbs. (**B**) Biological replicate samples of 100,000 IgG$^+$ B cells were isolated from three healthy donors for sequencing pipeline validation. A_2 represents a technical replicate of A (i.e., the same PCR product, for which library preparation and sequencing have been performed independently). (**C**) Overlap of unique CDRH3s between replicates from the samples in (B). Upper panel shows the total number of overlapping CDRH3s between biological and technical replicates.

1224 Lower panel shows mean CDRH3 overlap across all samples after 100 iterations
1225 of subsampling n=3623 CDRH3s from each dataset. (**D**) CDRH3 length
1226 distributions from samples in (B). Upper panel shows overlayed distributions for
1227 biological replicates. Lower panel shows p-values from Dunn post-hoc test after
1228 global Kruskal Wallis test. (**E**) V_H gene germline identity distributions from
1229 samples in (B). As for (D), upper panel shows overlayed distributions for
1230 biological replicates and lower panel p-values from Dunn post-hoc test after
1231 global Kruskal Wallis test. (**F**) Cohort sex and age distributions of n=57 healthy
1232 individuals for IgG^+ repertoire sequencing. (**G**) Total and mean number of raw
1233 reads, annotated reads, and identical sequences (identified by the same UMI)
1234 that were found at least three times. (**H**) V gene distributions, mean CDR3
1235 lengths and mean V gene germline identities for heavy, kappa, and lambda
1236 chains of all 57 individuals. CDR: Complementarity determining region; SHM:
1237 Somatic hypermutation; HC: heavy chain; KC: kappa chain; LC: lambda chain.
1238

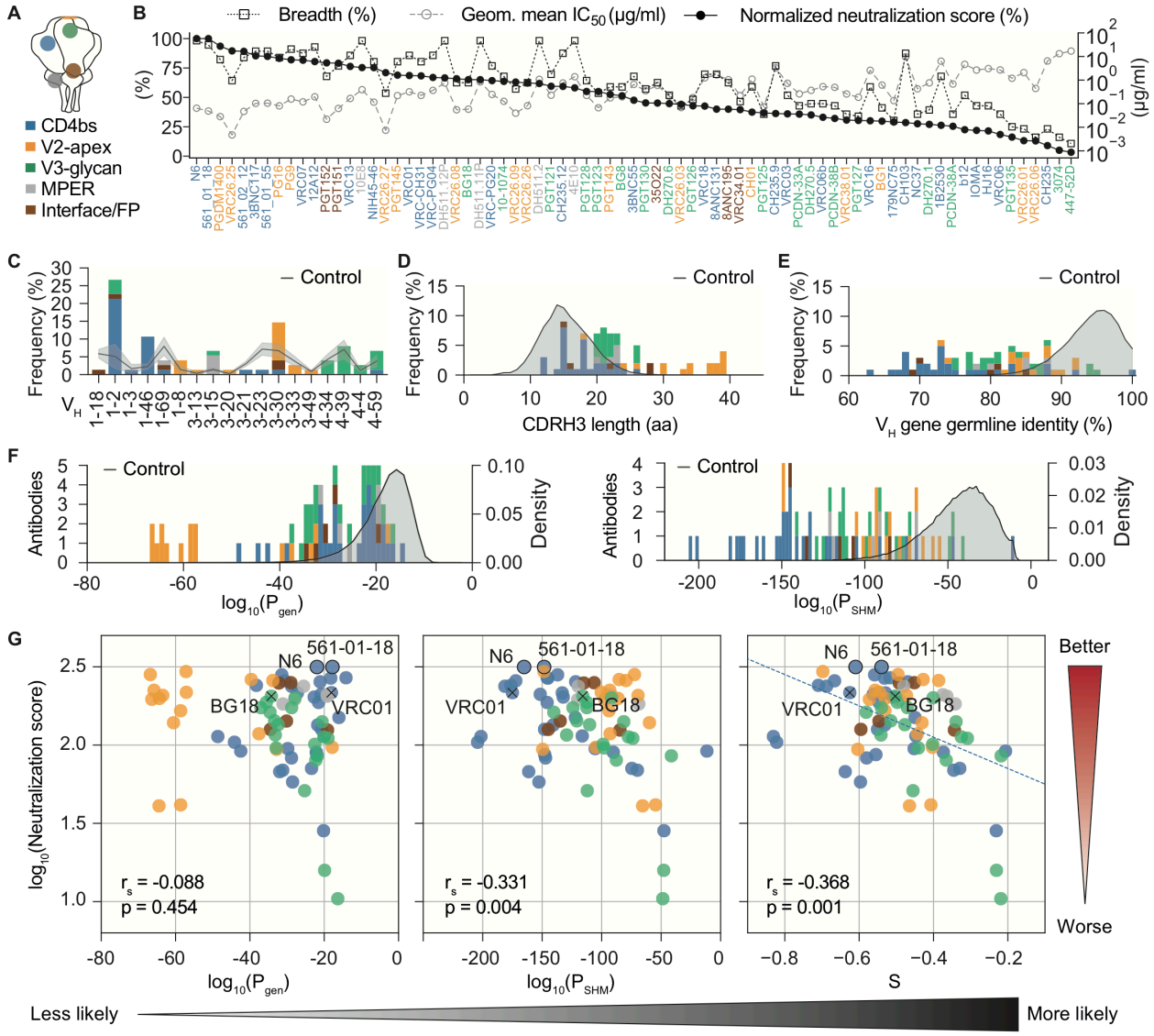

Fig. 2. Neutralization efficacy and sequence characteristics of broadly neutralizing antibodies targeting HIV-1. (**A**) bNAb epitopes on the HIV-1 envelope spike. MPER: membrane proximal external region; CD4bs: CD4 binding site; FP: fusion peptide. (**B**) Breadth, geometric mean half inhibitory concentration (IC_{50}) of neutralized strains, and normalized neutralization score for 75 bNAbs that have been tested against the same 56 HIV-1 strains. (**C**) Heavy chain V gene frequencies for the selected bNAbs. (**D**) Heavy chain CDR3 length distribution for the selected bNAbs in amino acids (aa). (**E**) Heavy chain V gene germline identity distribution for selected bNAbs. Controls in (C), (D) and (E) refer to sequence features from productive sequences of the 57 healthy individuals (Figure 1 F-H). (**F**) Heavy chain P_{gen} and P_{SHM} distribution for selected bNAbs derived by a model learned on productive sequences from 57 healthy individuals. Controls represent P_{gen} and P_{SHM} distributions for productive sequences from the 57 healthy individuals. (**G**) Correlation plots of bNAb neutralization scores against heavy chain P_{gen}, P_{SHM} and a combined probability

1255 score $S = c_1\log_{10}(P_{gen}) + c_2\log_{10}(P_{SHM})$, which was derived by a linear regression
1256 (dashed line) with $c_1 = 4.0\text{x}10^{-03}$ and $c_2 = 3.2\text{x}10^{-03}$. Spearman correlation
1257 coefficients r_s and p values are given in the figure. Correlation coefficient and p
1258 value from linear regression for S are $r = -0.446$ and $p = 6.031\text{x}10^{-5}$. Two near
1259 pan-neutralizing antibodies (N6 and 561-01-18) are highlighted by black outlines,
1260 two antibodies that have been used for structure-guided vaccine approaches
1261 (VRC01 and BG18) are highlighted by 'x'. Gradients on the right and bottom
1262 show directions for increasing neutralization activity and generation probability,
1263 respectively.
1264

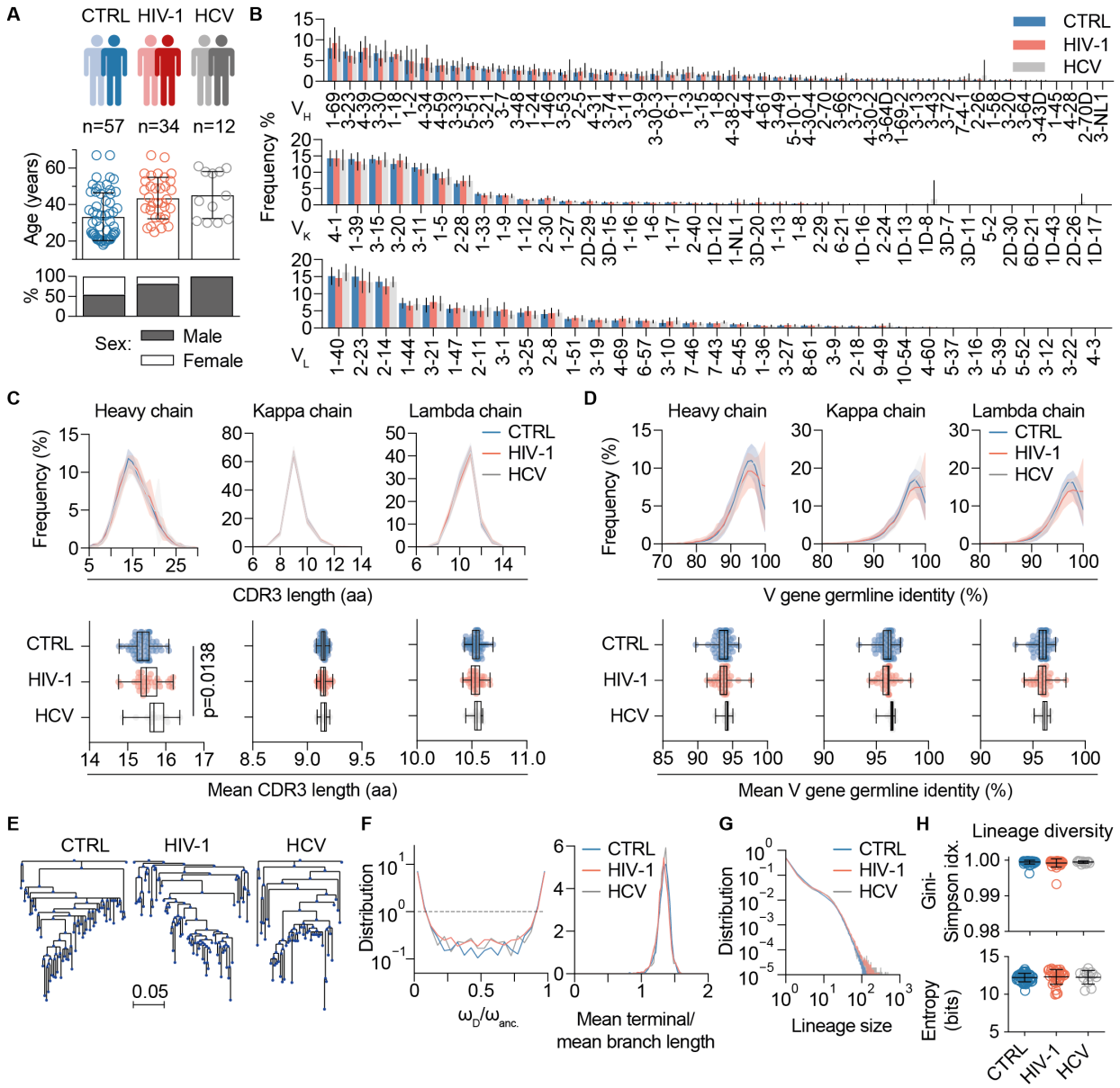

Fig. 3. IgG heavy chain repertoire characteristics of healthy and chronically infected individuals. (**A**) Cohort overview with age and sex distributions. Single dots in age distributions represent individuals, bar graphs and error bars depict mean cohort age with standard deviations. (**B**) V gene distributions for heavy, kappa, and lambda chains. V genes were ordered by descending frequencies according to the healthy control group. Bars depict mean frequencies across individuals for each cohort, error bars show standard deviations. (**C**) Mean CDR3 length distributions for heavy, kappa, and lambda chains in amino acids (aa). Upper panel shows the mean CDR3 length distributions across individuals for each cohort as solid lines and standard deviations as shaded areas. Lower panel shows mean CDR3 amino acid length as dots for each individual and cohort statistics as box plots on top. (**D**) Mean V gene germline nucleotide identity

distributions for heavy, kappa, and lambda chains. Representation of mean distributions (upper panel) and individual means (lower panel) as in C. One-way ANOVA with Tukey-HSD post-hoc test was performed on the means in (C) and (D). (**E**) Representative examples of lineages for the different cohorts based on heavy chain sequences. (**F**) Skewedness distribution of lineage trees (left panel, dashed line = neutral) and mean terminal/mean branch length distribution for lineage trees. (**G**) Lineage size distribution for the three cohorts. (**H**) Lineage diversity determined by Gini-Simpson index and entropy. CTRL = healthy control.

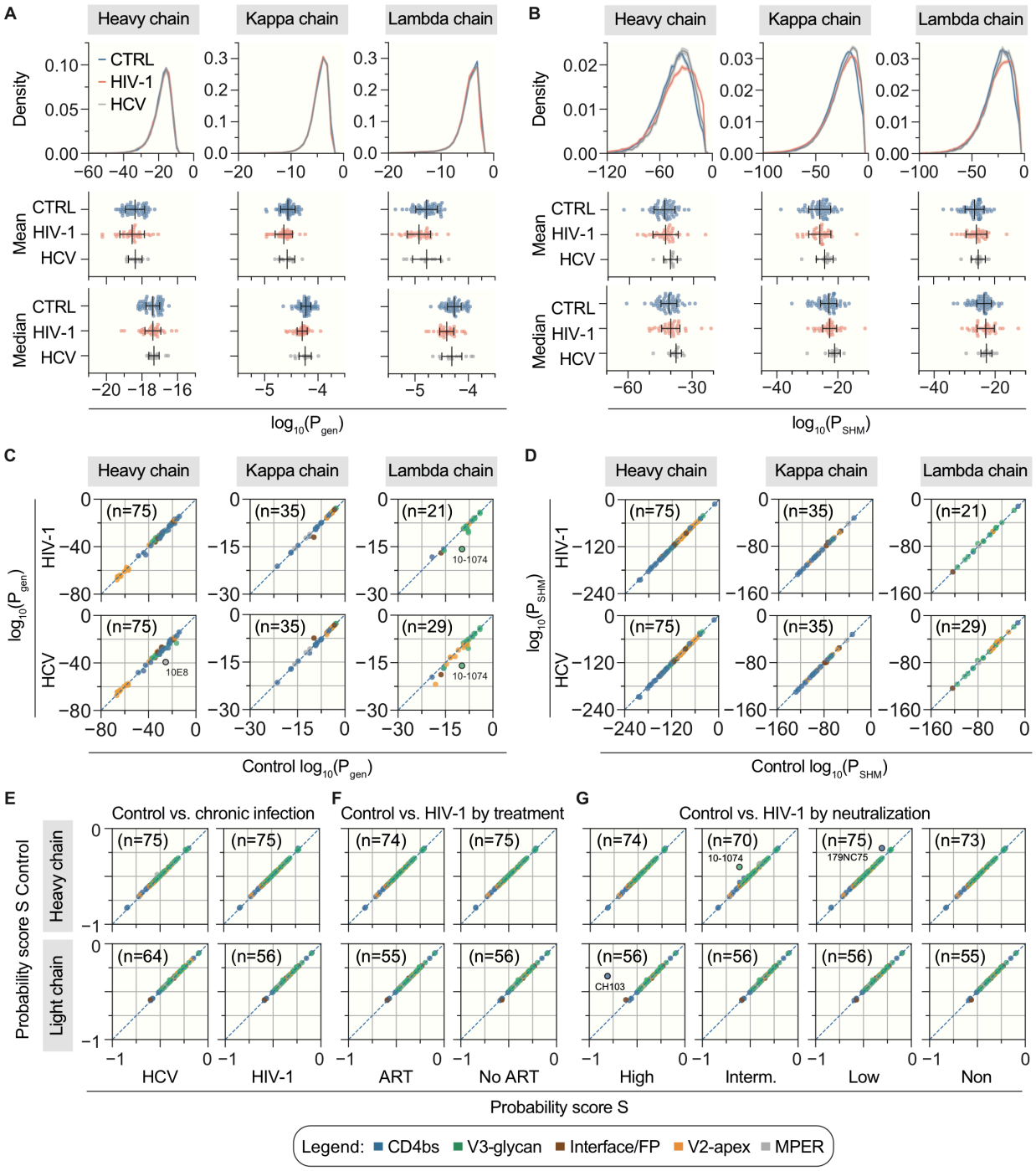

Fig. 4. **Probability distributions and bNAb probability scores across cohorts.** (**A**) P_{gen} distribution as well as mean and median P_{gen} for heavy and light chains derived from healthy control (CTRL), HIV-1-infected, and HCV-infected cohorts. (**B**) P_{SHM} distributions for heavy and light chains as in (A). (**C**) Correlation plots of P_{gen} for neutralizing antibody heavy, kappa, and lambda chains derived from either healthy control (x-axis) or chronically infected cohorts (HIV-1, HCV, y-axis).

(**D**) Correlation plots of P_{SHM} as in (C). Highlighted antibodies in (C) and (D) deviate >25% of the range of all observed values in comparison to the healthy cohort. (**E**) Correlation plot for bNAb heavy and light chain probability scores derived from HCV or HIV-1 cohorts (x-axis) in comparison to healthy control (CTRL, y-axis). Cohort-specific P_{gen} and P_{SHM} of bNAbs were used to calculate the bNAb probability scores S using coefficients from the linear regression on healthy individuals (Figure 3G). The dashed line represents identity. (**F**) Comparison as in (E) with the HIV-1 cohort stratified by antiretroviral therapy (on ART, n=22; off ART, n=12). (**G**) Comparison as in (E) with the HIV-1 cohort stratified by serum neutralization breadth against the global HIV-1 panel into high (≥66%, n=8), intermediate (≥33%, n=5), or low (>0%, n=16) breadth and no neutralization (0%, n=4) to recalculate P_{gen} and P_{SHM} of bNAbs for these cohorts. Serum neutralization data for one individual was not available. ART: anti-retroviral therapy. Numbers in brackets in (C - G) indicate bNAbs for which P_{gen}/P_{SHM} and scores could be inferred.