HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2019 May 15.

Published in final edited form as:

Cell Rep. 2019 April 30; 27(5): 1422–1433.e4. doi:10.1016/j.celrep.2019.03.097.

Comprehensive Profiling of HIV Antibody Evolution

Susan H. Eshleman[1,*], **Oliver Laeyendecker**[2,3], **Kai Kammers**[4], **Athena Chen**[5], **Mariya V. Sivay**[1], **Sanjay Kottapalli**[1], **Brandon M. Sie**[1], **Tiezheng Yuan**[1], **Daniel R. Monaco**[1], **Divya Mohan**[1], **Daniel Wansley**[1], **Tomasz Kula**[6], **Charles Morrison**[7], **Stephen J. Elledge**[6], **Ron Brookmeyer**[8], **Ingo Ruczinski**[5], and **H. Benjamin Larman**[1,9,*]

[1]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[2]Laboratory of Immunoregulation, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, NIH, Baltimore, MD, USA

[3]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[4]Division of Biostatistics and Bioinformatics, Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[5]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[6]Division of Genetics, Department of Medicine, Howard Hughes Medical Institute, Brigham and Women's Hospital, Department of Genetics, Harvard University Medical School, Boston, MA 02115, USA

[7]FHI 360, Clinical and Epidemiologic Sciences, Durham, NC, USA

[8]Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA, USA

[9]Lead Contact

SUMMARY

[*]Correspondence: seshlem@jhmi.edu (S.H.E.), hlarman1@jhmi.edu (H.B.L.) .

AUTHOR CONTRIBUTIONS

S.H.E. conceived of the study, assisted with study design, analyzed study data, and co-wrote the manuscript. O.L. assisted with study design, analyzed study data, and co-wrote the manuscript. B.M.S. performed VirScan testing and prepared the public epitope library. A.C. analyzed baseline variable data and carried out time-to-event analyses. S.K. determined the breadth of HIV responses. T.Y. determined peptide read counts and enrichment scores from the raw sequencing data. M.V.S. analyzed genomic location of HIV peptides and assisted with data presentation. D.M. performed the VirScan screening of the validation set samples. D.R.M. assisted with data analysis. D.W. prepared the VirScan library for screening. C.M. provided samples and data from the Genital Shedding Study. S.J.E. supported development of the public epitope library. T.K. designed the public epitope library. R.B. provided input for statistical analysis. K.K. analyzed VirScan data and generated figures. I.R. analyzed the study data, generated figures, and co-wrote the manuscript. H.B.L. oversaw all aspects of VirScan screening and data analysis and co-wrote the manuscript. All authors contributed to manuscript preparation and critically reviewed the final version of the manuscript.

SUPPLEMENTAL INFORMATION

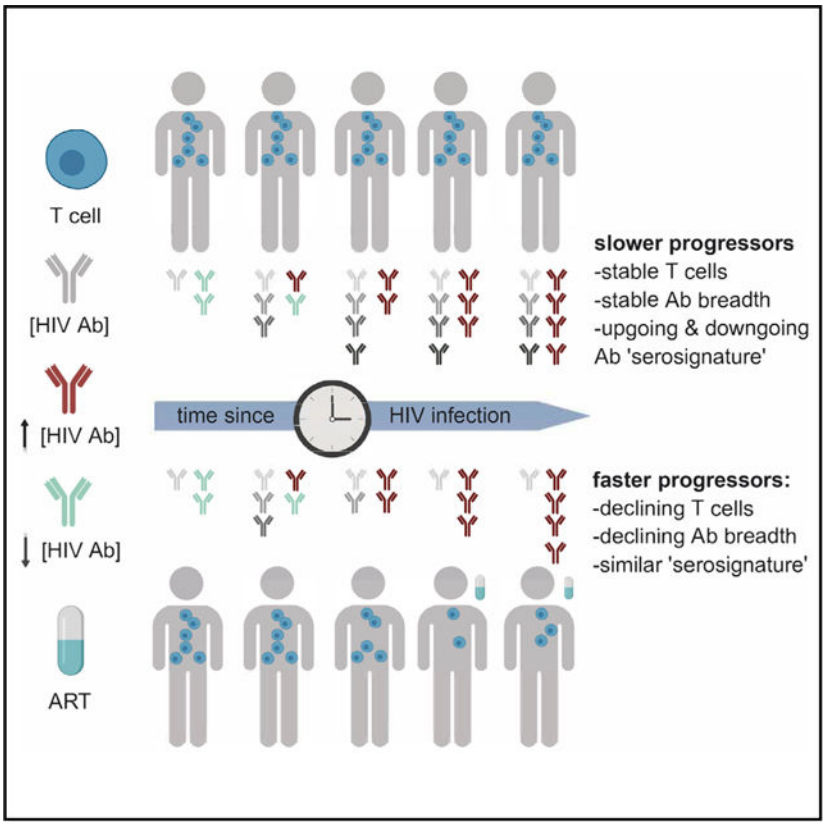
Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.03.097.

DECLARATION OF INTERESTS

None of the authors has a financial or personal relationship with other people or organizations that could inappropriately influence their work. S.H.E., O.L., K.K., and H.B.L. are listed as inventors on a patent application related to development of a serosignature-based assay for HIV incidence testing. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

This study evaluates HIV antibody responses and their evolution during the course of HIV infection. A phage display system is used to characterize antibody binding to >3,300 HIV peptides in 57 adults with early- to late-stage infection. We find that the number of unique epitopes targeted ("antibody breadth") increases early in infection and then stabilizes or declines. A decline in antibody breadth 9 months to 2 years after infection is associated with subsequent antiretroviral treatment (ART) initiation, and a faster decline in antibody breadth is associated with a shorter time to ART initiation. We identify 266 peptides with increasing antibody reactivity over time and 43 peptides with decreasing reactivity over time. These data are used to design a prototype four-peptide "serosignature" to predict duration of HIV infection. We also demonstrate that epitope engineering can be used to optimize peptide binding properties for applications such as cross-sectional HIV incidence estimation.

Graphical Abstract



In Brief

Eshleman et al. quantify antibody binding to >3,300 HIV peptides from early- to late-stage infection using a phage display system (VirScan). Binding diversity (breadth) reaches individual-specific set points; breadth decline is associated with CD4 cell loss. Time-dependent binding specificities are identified, optimized, and used to predict duration of HIV infection.

INTRODUCTION

Antibodies to HIV appear shortly after infection. The titer and avidity of anti-HIV antibodies generally increase over time, but may be impacted by antiretroviral treatment (ART), CD4 T cell decline, and other factors (Koenig et al., 2013; Fiebig et al., 2003). The breadth and specificity of anti-HIV antibodies also evolve during the course of infection (Geiß and Dietrich, 2015). A detailed understanding of the serologic response to HIV infection is helpful for understanding HIV immune containment and for vaccine development. Multiplexed immunoassays have been used to analyze the specificity of anti-HIV antibodies. These include a microarray assay composed of 15 recombinant HIV env protein targets and five gp41 peptide targets (Dotsey et al., 2015), and an assay based on the Luminex platform that includes six recombinant HIV protein targets (Curtis et al., 2012). Phage display technology has also been used to screen HIV peptides for binding to immobilized antibodies (Delhalle et al., 2012).

In this report, we used a massively multiplexed antibody profiling system to analyze the fine specificity of the antibody response to HIV infection. This system is based on phage immunoprecipitation sequencing (PhIP-Seq) (Larman et al., 2011). Testing is performed by incubating samples with a bacteriophage library that expresses peptides encoded by oligonucleotides generated by high-throughput DNA synthesis. The abundance and specificity of antibodies in test samples are assessed by immunoprecipitating phage-antibody complexes, and then amplifying and sequencing the DNA in the captured phage particles. The “VirScan” phage library includes >96,000 peptides that span the genomes of >200 viruses that infect humans (the human “virome”) (Xu et al., 2015). We performed PhIP-Seq using the VirScan library to analyze HIV antibodies from individuals with known duration of HIV infection, ranging from <1 month to 8.7 years. This allowed us to examine dynamic changes in antibody diversity and the fine specificity of HIV antibodies from individuals with early to late stage infection, including individuals on ART and individuals with advanced HIV disease.

HIV incidence is often determined by following cohorts of HIV-uninfected individuals and quantifying the rate of new HIV infections. HIV incidence can also be estimated using a cross-sectional study design, using laboratory assays to identify individuals who are likely to have recent HIV infection. Most serologic assays used for cross-sectional HIV incidence estimation measure general characteristics of the antibody response to HIV infection (e.g., antibody titer, antibody avidity) (Murphy and Parry, 2008; Guy et al., 2009; Busch et al., 2010), which may be impacted by viral suppression, loss of CD4 T cells, and other factors (Laeyendecker et al., 2012b, 2012a; Kassanjee et al., 2014; Brookmeyer et al., 2013). We used the VirScan assay to identify peptide biomarkers associated with the duration of HIV infection, and demonstrated that peptide engineering can be used to enhance the properties of peptides for discriminating between early- and late-stage infection. This information could be used to develop improved methods for estimating HIV incidence from cross-sectional surveys, for surveillance of the HIV/AIDS epidemic (Justman et al., 2018), and for evaluating the impact of interventions for HIV prevention in clinical trials (Coates et al., 2014).

RESULTS

Antibody Reactivity to HIV Peptides

We used the VirScan assay to characterize anti-HIV antibodies in 403 plasma samples from 57 women with subtype C HIV infection (discovery sample set; Table S1). The time from seroconversion to sample collection ranged from 14 days to 8.7 years. The density of peptides in the library varied across the open reading frames for HIV proteins (the HIV proteome, Figures 1A and 1B). The level and frequency of antibody binding were highly variable (Figures 1C and 1D); the strongest and most frequent antibody binding was observed for peptides in the gag and env regions. Some peptides were consistently targeted over the course of the infection; in contrast, the level and frequency of antibody binding to other peptides changed over the course of HIV infection (Figure 1E). Samples are diluted prior to VirScan analysis to normalize the amount of immunoglobulin G (IgG) in each sample. We analyzed the IgG concentrations in our sample set to assess whether this normalization procedure might lead to bias (e.g., in participants with hypergammaglobulinemia). While we did observe differences in IgG concentration as a function of duration of infection for study participants who did versus did not initiate ART during study follow-up, further analysis suggested that diluting samples to adjust for differences in IgG concentration did not confound our data (Figure S1).

Breadth of Antibody Reactivity

We next analyzed the diversity of each individual's antibody response to HIV over time. Network graphs were used to determine antibody breadth at each time point; antibody breadth was defined as the number of non-overlapping peptides with high levels of antibody binding. Figure 2A shows the network graph for peptides that reacted with antibodies from a representative study sample (one immunoprecipitation reaction). This analysis identified 45 non-overlapping peptides; these peptides were located in the gag, pol, env, vpu, and rev regions. We next analyzed the change in antibody breadth over the course of HIV infection. Since ART is known to influence HIV antibody production, we compared antibody breadth data from participants who did versus did not start ART during the genital shedding (GS) study (Figure 2B). ART also serves as a surrogate for disease progression; in the GS study, ART was recommended when the CD4 cell count fell below 250 cells/mm^3. Overall, 32 participants started ART during the GS study.

In both groups (with and without ART initiation), antibody breadth increased during the first 6 months of infection. In the group that did not start ART, a relatively stable value for antibody breadth (termed "antibody breadth set point") was established in most individuals approximately 9 months to 1 year after infection; the antibody breadth set point varied considerably among study participants. In contrast, in the group that ultimately started ART, a decline in antibody breadth was observed approximately 1 year after infection. In all cases, the fall in antibody breadth occurred at least a year before ART initiation (Figure 2B). To further explore the relationship between the drop in antibody breadth and disease progression, we compared the timing of the drop in antibody breadth in the ART group (~9 months to 2 years after infection) to the temporal distribution of samples with CD4 cell counts of <350 cells/mm^3 or viral loads of <1,000 copies/mL; there was no apparent

relationship between antibody breadth and the timing of these two factors (Figure S2). After participants started ART, HIV antibody breadth appeared to stabilize at levels similar to those seen in early HIV infection.

We next evaluated the relationship between HIV infection and the antibody response to a different, chronic infection that was expected to have a high prevalence in the study setting (Epstein-Barr virus [EBV]). Data used to calculate the breadth of the antibody response to EBV infection were obtained from the same VirScan datasets used for HIV analysis (Figure 2C). In most participants, EBV antibody breadth was relatively stable in the first 6 months of HIV infection, and then declined. EBV antibody breadth then appeared to stabilize in participants who did not start ART for HIV infection. In contrast, in most participants who started ART, EBV antibody breadth increased after ART initiation; in 16 (66.7%) of these cases, EBV antibody breadth after ART initiation exceeded the highest level observed prior to ART.

In this study, antibody reactivity to rabies virus and Ebola virus was used to normalize data for HIV prior to analysis (to adjust for sample-to-sample differences in sequencing depth). Antibody reactivity to these viruses most likely reflected non-specific antibody binding, since participants in the GS study did not have these infections. In contrast to the high antibody breadth values observed for HIV and EBV, antibody breadth values for these “control” viruses were very low (maximum antibody breadth value: 5).

Factors Associated with Changes in Antibody Breadth over Time

To explore the relationship between the decline in HIV antibody breadth and subsequent ART initiation, we calculated the rate of change of antibody breadth over the period ~9 months to ~2 years after HIV infection (based on sample availability); none of the participants included in the analysis was on ART during this time window. For this time-to-event analysis (the outcome being time to ART initiation), participants were divided into two groups: those with declining breadth and those with stable or increasing breadth. We found that participants who had stable or increasing antibody breadth ~9 months to ~2 years after infection were less likely to start ART earlier in infection (log-rank test, $p = 0.009$; hazards ratio, 0.29; 95% confidence interval: 0.11, 0.78; $p = 0.014$; Figure 3). The average time between the study visits used to evaluate the change in antibody breadth (~9 months and ~2 years after infection) was similar in the two groups ($p = 0.28$), so this was not likely to have biased the analysis.

We next evaluated the relationship between the rate of decline in antibody breadth and other factors, including age at infection, baseline CD4 cell count, rate of decline in CD4 cell count, and viral load set point (Figure 4). A faster decline in antibody breadth was strongly associated with lower baseline CD4 cell count ($R = 0.42$; 95% confidence interval: 0.17,0.62; $p = 0.002$) and higher viral load set point ($R = -0.43$; 95% confidence interval: −0.62, −0.18; $p = 0.001$), and was also associated with earlier ART initiation ($R = 0.28$; 95% confidence interval: 0.01, 0.51; $p = 0.043$).

Dynamic Changes in Antibody Binding

We next explored the relationship between HIV antibody specificity and the duration of HIV infection. First, we used a linear model to quantify the association between antibody binding and the duration of infection for the 3,384 HIV peptides in the VirScan library. This analysis was performed using all 403 samples in the discovery sample set. The model identified 309 peptides that had a significant association between these two factors ($p < 0.05$ after adjusting for multiple comparisons using the Bonferroni method; Figure 5); 266 peptides had increasing antibody binding over time (positive association) and 43 peptides had decreasing antibody binding over time (negative association). The position of peaks representing increased versus decreased antibody binding were observed at different positions in the HIV genome. Peptides that had a strong positive association with duration of infection tended to cluster in the N-terminal gag region, the C-terminal pol region, and defined domains within the env region. In contrast, peptides that had a strong negative association with duration of infection clustered in the C-terminal gag region, and the middle of the pol region, with others scattered across the env region or located in non-structural (accessory) proteins, such as nef.

We then selected the four peptides that had the strongest independent association between antibody binding and the duration of HIV infection (Figure S3; Table S2). This included two peptides that had increased antibody binding over time (one in gp41; one in gp120) and two peptides that had decreased antibody binding over time (one in gag; one in pol). Antibody binding measures from each of the four peptides were combined in a simple linear model to generate a single, unweighted, four-peptide composite measure. The duration of infection predicted by this model was highly correlated with the observed (true) duration of infection (generalized estimating equations [GEE], $p < 1 \times 10^{-100}$; Figure 6A).

Most serologic assays developed for HIV incidence estimation measure general characteristics of the antibody response to HIV infection (e.g., antibody titer or avidity). Factors such as low viral load, low CD4 cell count, and ART can lead to misclassification of long-term infections as “recent” using those assays. Figure 6 shows the true and predicted duration of infection for samples with these characteristics. To evaluate the impact of these factors on “false-recent” misclassification by the four-peptide model, we calculated the proportion of long-term infections (defined here as >1-year duration) that were misclassified as recent infections (defined here as <6 months) in different groups. None of the “challenge” samples (on ART, viral load <400 copies/mL, or CD4 cell count <350 cells/mm^3) was misclassified as “recent” by the four-peptide model (Table S3). We also compared data from the four-peptide model with data obtained using the limiting antigen avidity (LAg-avidity) assay, which is widely used for cross-sectional HIV incidence estimation (Figure S4) (Wei et al., 2010). That analysis shows that data from the four-peptide model is more strongly correlated with duration of HIV infection than data from the LAg-avidity assay.

We next evaluated the performance of the four-peptide model using an independent validation sample set (Table S1). This set consisted of samples from individuals in the GS study who were not included in the discovery set that was used to identify the model peptides. This sample set also included “challenge samples” that have characteristics known to complicate cross-sectional HIV incidence estimation using other serologic assays: 28

(38.9%) of the samples were HIV subtype D; 37 (51.4%) had CD4 cell counts <350 cells/mm^3; 16 (22.2%) had viral loads <1,000 copies/mL, and 12 (16.7%) were from individuals on ART. The duration of infection predicted by the four-peptide model was also correlated with the observed (true) duration of infection using this independent sample set (GEE, $p < 3 \times 10^{-36}$; Figure 6D). Figure 6D shows the true and predicted duration of infection for subtype D samples; the distribution of those data was similar to the distribution of data for subtype C samples, suggesting that the performance of the four-peptide model was similar in these two subtypes.

Epitope Engineering

Next, we explored whether peptide epitopes could be modified to improve the association between antibody binding and the duration of HIV infection. We first selected 11 non-overlapping peptides that were targeted by the majority of HIV-infected individuals ("public epitope peptides"). We then generated variant peptides by substituting each set of three consecutive amino acids with alanine residues. Figure 7 shows the impact of alanine substitutions on antibody binding for 2 of the 11 parent peptides; these peptides were targeted by >98% of the study participants. In the first case (parent peptide A), antibody binding to the parent peptide and most of the variant peptides decreased with increasing duration of infection (Figure 7A). Alanine substitutions at amino acid positions 26–34 appeared to disrupt antibody binding at all time points. In the second case (parent peptide B), antibody binding to the parent peptide and most of the variant peptides increased with increasing duration of infection (Figure 7B). In this case, alanine substitutions at amino acid positions 13–21 preferentially disrupted antibody binding early in infection. Figure 7C shows the level of antibody binding as a function of duration of infection for parent peptide B and variant peptides that had alanine substitutions in the region most impacted by mutagenesis (nine peptides, with substitutions at positions 13–21). Over the course of HIV infections assessed in our study, antibody binding to the parent peptide increased by 57%; in contrast, antibody binding to one of the variant peptides increased by approximately 479% over the same time period. These data provide proof-of-principle that epitope engineering can be used to improve the capacity of peptides to serve as quantitative biomarkers of disease processes, such as the duration of HIV infection.

DISCUSSION

This study provides the most comprehensive analysis of HIV antibody specificities to date, including their characterization from early- to late-stage infection. We found that changes in antibody diversity early in infection were associated with differences in clinical outcome (measured as time to ART initiation). This study also provides proof-of-principle that an "HIV serosignature," reactivity to a panel of HIV peptides, may be useful for cross-sectional HIV incidence estimation.

We used the measure, "antibody breadth," to quantify HIV antibody diversity and found that this measure reaches a plateau ("antibody breadth set point") early in infection in individuals who do not start ART. In the GS study cohort, a decline in antibody breadth between 9 months and 2 years after infection was associated with a shorter time to ART initiation,

which was prompted in the GS study cohort by a decline in CD4 cell count to <250 cells/mm^3. The decline in antibody breadth among those who subsequently started ART likely reflected declining B cell support due to loss of T helper cells. HIV antibody breadth appeared to stabilize at a low level after ART initiation. In contrast, the breadth of the EBV antibody response increased sharply after ART initiation, which may have reflected immune reconstitution (Sharma and Soneja, 2011). While these observations provide important insights into the immune response to HIV infection, antibody breadth measurements generated with the VirScan assay are unlikely to be useful for monitoring HIV infection in clinical settings. Use of CD4 cell counts to monitor HIV disease progression is well established, and CD4 cell count data were more strongly correlated with time to ART initiation than antibody breadth in this study.

Previous studies have identified several factors associated with HIV disease progression, including virologic factors (e.g., HIV viral load [Touloumi et al., 2013], replication capacity [Ng et al., 2014], and subtype [Baeten et al., 2007]), immunologic factors (e.g., inversion of the CD4/CD8 ratio [Margolick et al., 2006], polyclonality of the anti-HIV T cell response [Pantaleo et al., 1997], and degree of early immune activation [Fahey et al., 1990]), and host factors (e.g., human leukocyte antigen [HLA] type B57 [Costello et al., 1999] and CCR5 delta 32 mutations [Huang et al., 1996]). It is not clear whether the decline in antibody breadth that we observed caused disease progression leading to ART initiation, or whether it was a surrogate for other changes, such as a decline in T cell number or function. If the decline in antibody breadth has a causative role in disease progression, then use of therapeutic vaccines to boost antibody diversity may in theory provide clinical benefit.

Generalized antibody responses to HIV infection, such as antibody titer and avidity, tend to plateau approximately 1 year after HIV infection (Busch et al., 2010). These characteristics of the antibody response are impacted by a variety of factors, including natural and drug-induced viral suppression (Koenig et al., 2013; Wendel et al., 2017; Kassanjee et al., 2014), disease progression (Laeyendecker et al., 2012b), and HIV subtype (Longosz et al., 2014, 2015). Previous studies evaluating the banding pattern in western blots demonstrate that HIV antibody specificity evolves early in infection (Fiebig et al., 2003). Recent studies have explored whether assays that include a small number of protein or peptide targets could be used to identify recent HIV infections (Dotsey et al., 2015; Curtis et al., 2012). Using the VirScan assay to analyze 403 plasma samples, we were able to quantify antibody binding to >3,300 HIV peptides from early- to late-stage HIV infection. These data were used to generate a simple, unweighted, four-peptide model that predicted duration of HIV infection. Data from the prototype four-peptide model were more strongly correlated with duration than data from the LAg-avidity assay that is in wide use for cross-sectional HIV incidence estimation (Figure S4) (Wei et al., 2010).

The peptides included in this prototype model were from four different HIV proteins (gp41, gp120, gag, and pol). Two of these peptides had increasing antibody reactivity over time, and two had decreasing antibody reactivity over time. It is noteworthy that the gp41 peptide, which showed the strongest association with duration of infection, included a sequence shared by the HIV subtype B target peptide in the LAg assay. Our analysis also

demonstrated that epitope engineering can be used to enhance the capacity of individual peptides to discriminate between early and late HIV infection.

The VirScan assay has several advantages over alternative multiplex serological assays for peptide discovery. These advantages include quantitative assessment of antibody binding to peptides that span all open reading frames in the HIV genome, including both structural and regulatory proteins; representation of a wide range of HIV subtypes and strains, including groups M, N, and O, and HIV-2 (Table S4); and fine resolution for epitope identification, which can be further refined with alanine scanning mutagenesis. The assay also provides information about antibody binding to >200 other human viruses. In this report, data from other viral peptides were used to normalize peptide binding measures and allowed us to compare the impact of ART on the antibody response to a prevalent non-HIV viral infection (EBV). Data from the same assay runs could be used to examine the evolution and fine specificity of antibodies to other viruses, and the impact of viral co-infections on the anti-HIV antibody response. Future studies could also explore use of the VirScan assay to identify serosignatures for estimating incidence of other viral infections, such as hepatitis C virus. Finally, future phage libraries composed of additional protein products, such as those from the gut microbiome, may be used to explore the impact of immune system pre-conditioning on the response to HIV infection.

Data obtained in this study provide proof-of-principle that the VirScan assay can be used to identify peptides for applications such as cross-sectional HIV incidence estimation. Considerable work is needed to translate findings from this study into a laboratory test that can be used for improved cross-sectional HIV incidence testing. We are now evaluating antibody reactivity to peptides identified with the VirScan assay using a quantitative, multi-peptide electrochemiluminescent assay that has a wide dynamic range for antibody reactivity. Our preliminary data indicate that this platform provides measures of antibody reactivity that are strongly correlated with antibody reactivity data from the VirScan assay. This system will facilitate analysis of large sample sets to obtain the additional data needed to identify optimal serosignatures for cross-sectional HIV incidence estimation and related applications. This simpler assay system could ultimately be used as the testing platform for a serosignature-based HIV incidence assay.

Further evaluation of serosignatures will be performed using sophisticated statistical and machine learning approaches that we used in previous studies to evaluate multi-assay algorithms for cross-sectional incidence estimation (e.g., algorithms included the LAg-Avidity assay, other serologic assays, CD4 cell count and HIV viral load) (Konikoff et al., 2013; Laeyendecker et al., 2018). Those studies used large datasets from individuals with HIV infection duration ranging from 1 month to >8 years (e.g., data from 2,442 samples from 278 adults with subtype C infection [Laeyendecker et al., 2018]; data from 1,782 samples from 709 adults with subtype B infection [Laeyendecker et al., 2013]). We will use the electrochemiluminescent assay described above to obtain peptide-binding data for the same sample sets. We will use those data to compare serosignatures that include different sets of peptides, weighting for individual peptides, and different cutoffs for antibody binding to each peptide included in the models; the analyses will also determine key performance characteristics of different serosignatures (e.g., the mean window period for recent

infection). As a final step, performance of serosignatures will be validated by comparing cross-sectional HIV incidence estimates to HIV incidence observed in longitudinal trials and cohort studies.

This study has several limitations. One limitation of the VirScan assay is that it does not allow evaluation of discontinuous or highly conformational epitopes, or epitopes that include post-translational modification, such as glycosylation. In this study, we used protein A/G-coated magnetic beads for immunoprecipitation of all IgG subclasses. Analyses of IgA or specific IgG subclasses may provide complementary information (Kadelka et al., 2018). Another limitation of this report is that the sample cohort included only women. The samples in the discovery set all had subtype C HIV; the validation set included subtype C and D samples. Further studies are needed to compare these data to data from samples from other geographic regions where other HIV subtypes and strains circulate. The preliminary studies described in this report also did not include viremic controllers who naturally suppress HIV infection, or individuals who initiated ART early in infection. It would be worthwhile to evaluate the serologic responses to HIV infection in those populations, since viremic control and early suppression of viral replication can impact the production and evolution of HIV antibodies (Koenig et al., 2013).

This study provides detailed information on the humoral response to HIV infection over time and demonstrates the utility of the VirScan assay for identifying peptide biomarkers for applications such as cross-sectional HIV incidence estimation. This technology could also be used to evaluate serologic responses to other infectious diseases, as well as the impact of viral co-infections on immune responses. This may improve understanding of the complex relationships between viral infections and the immune responses that they elicit.

STAR★METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author H. Benjamin Larman (hlarman1@jhmi.edu). Use of the VirScan library is subject to MTA with Brigham & Women’s Hospital.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ethics statement—This study was approved by the Institutional Review Board of the Johns Hopkins University. The study involved analysis of data and samples collected in the Hormonal Contraception and HIV Genital Shedding (GS) Study. The study was conducted according to the ethical standards set forth by the Institutional Review Board of Johns Hopkins University and the Helsinki Declaration of the World Medical Association. All participants were adults who provided written informed consent.

Samples used for analysis—Plasma samples (Table S1) were obtained from the GS Study (Uganda and Zimbabwe; 2001–2009), which evaluated the relationship between hormonal contraceptive use, genital shedding of HIV, and HIV disease progression among women with known dates of HIV seroconversion (Morrison et al., 2011). ART was recommended for study participants with CD4 cell counts below 250 cells/mm^3, consistent

with local treatment guidelines at the time the GS Study was performed. Data for CD4 cell count and viral load were collected in the GS Study (Morrison et al., 2011); data on the timing of ART initiation was obtained by review of clinic records.

We analyzed samples from participants who acquired HIV infection, where the maximum time between collection of the last HIV-negative sample and the first HIV-positive sample was four months. For each individual, the estimated date of infection was defined as either the midpoint between visits with the last negative HIV antibody test and the first positive HIV antibody test, or fifteen days before documentation of acute infection (HIV RNA positive/HIV antibody negative status). HIV-1 LAg-Avidity EIA (SEDIA Biosciences Corporation Portland, OR) was performed on all samples according to the manufacturer's instructions. Two sets of samples were analyzed in this report: a discovery sample set (403 samples from 57 individuals; Table S1) and a validation sample set (72 samples from 32 individuals; Table S1). The discovery sample set included participants who had at least one year of follow-up after seroconversion, with samples collected at three or more study visits during that period. The independent validation sample set included samples from participants from the GS Study who were not included in the discovery sample set. HIV subtype assignments were based on phylogenetic analysis of the HIV env C2V3 region (Morrison et al., 2010). All of the samples in the discovery sample set were HIV subtype C; the validation sample set also included "challenge" samples with HIV subtype D, which are often misclassified using currently available serologic HIV incidence assays (Longosz et al., 2014, 2015).

Sample size estimation was not performed. All samples meeting study inclusion criteria were analyzed (Table S1). Only one group allocation was performed: individuals who did or did not initiate anti-retroviral therapy. These data were obtained from the original clinical study.

METHOD DETAILS

Phage library used for analysis—The VirScan library includes 3,384 HIV peptides spanning all HIV proteins (Xu et al., 2015). The protein sequences used to design peptide tiles were selected from the UniProtKB database, balancing sequence diversity and library size (Xu et al., 2015). The peptides are 56-amino acids long with 28-amino acid overlaps and represent diverse HIV subtypes and strains (Table S4). In this study, the VirScan library was augmented with a public epitope library that included peptides previously found to be targeted by a high proportion of HIV-infected individuals (Xu et al., 2015). Eleven "parent" peptides in the public epitope library were modified by introducing triple alanine substitutions centered at each amino acid position; the resulting public epitope library included 594 genetically-engineered variant HIV peptides. Silent nucleotide substitutions were encoded in the first 50 nucleotides of each DNA tile, so that variant peptides could be uniquely identified using 50-nucleotide single-end Illumina sequencing.

The VirScan library also includes 2,263 Epstein-Barr virus (EBV) peptides, 718 Ebola virus peptides, and 518 rabies virus peptides; the public epitope library includes an additional 227 Ebola virus peptides. In this report, EBV data were used to evaluate the impact of antiretroviral therapy for HIV infection on the breadth of the anti-EBV antibody response.

Ebola and rabies virus data were used to normalize antibody binding data to account for differences in sequencing depth between samples.

Phage immunoprecipitation and DNA sequencing—Detailed procedures for the VirScan assay were described previously (Mohan et al., 2018; Xu et al., 2015). In this study, the concentration of IgG in plasma samples was determined using an in-house enzyme-linked immunosorbent assay (capture and detection antibodies 2040–01 and 2042–05, respectively Southern Biotech, Birmingham, AL). Approximately 2 mg of IgG from each sample were added to the combined T7 bacteriophage VirScan and public epitope libraries (1×10^5 plaque forming units for each phage clone in each library), diluted in phosphate-buffered saline to a final reaction volume of 1 mL in a deep 96-well plate, and incubated overnight at 4 °C. Eight mock immunoprecipitation reactions (no plasma) were included on each plate; these reactions served as negative controls for data normalization. After rotating the plates overnight at 4°C, 20 μL of protein A-coated magnetic beads and 20 μL of protein G-coated beads (catalog numbers 10002D and 10004D, Invitrogen, Carlsbad, CA) were added to each reaction; the plates were rotated for another 4 hours at 4°C. Immunoprecipitation reactions were processed using the Agilent Bravo liquid handling system (Agilent Technologies, Santa Clara, CA). Beads were washed twice with Tris-buffered saline (50 mM Tris-HCl with 150 mM NaCl, pH 7.5) containing 0.1% NP-40 and then resuspended in 20 μL of a polymerase chain reaction (PCR) mix containing Herculase II Polymerase (catalog number 600679, Agilent Technologies). After 20 cycles of PCR, 2 μL of the PCR products was added to a second 20-cycle PCR reaction, which added sample-specific barcodes and P5/P7 Illumina sequencing adapters to the amplified DNA. DNA sequencing of the pooled PCR products was performed using an Illumina HiSeq 2500 instrument (Illumina, San Diego, CA) in rapid mode (50 cycles, single end reads).

QUANTIFICATION AND STATISTICAL ANALYSIS

Each plasma sample was profiled using VirScan without technical replication (“n=1”).

Analysis of DNA sequencing data—Fastq files from DNA sequencing were demultiplexed using exact matching of 8-nucleotide sample-specific i5 and i7 DNA barcodes (Illumina). For each sample, a read count (the number of times each sequence was detected) was obtained for each peptide using Bowtie alignment (Langmead et al., 2009), without allowing any mismatches. The level of antibody-dependent enrichment of each peptide in each sample was determined by comparing the read count for the sample to the read counts obtained for 40 mock immunoprecipitation reactions (8 mock reactions per plate). Two different measures were used to quantify the degree of antibody binding: “z-scores” were used to reduce false positivity in cases of low sequencing depth (this approach was used to generate data for Figure 1 and for calculation of antibody breath); “relative fold-change” was used to normalize data for highly-enriched peptides (this approach was used to generate data for Figures 5, 6, 7, and S2). Z-scores were calculated by subtracting the expected normalized read count (determined by regression against the mock immunoprecipitation reactions) from the observed normalized read count; the resulting value was then divided by an estimate of the standard deviation of the normalized read counts, based on the mock immunoprecipitation reactions (Yuan et al., 2018). Relative fold change values were

quantified as peptide logarithmic relative fold changes. Since effects and errors for proteomic data and technologies such as RNA-seq data are typically believed to be on a multiplicative scale, we $\log_{10}$ transformed the read counts after adding one read count for each peptide in each sample, as is customary in these settings. This results in effects and errors acting on an additive scale, which for example is required in the linear models employed. In order to normalize the read count data, peptides from Ebola virus and rabies virus (under the assumption the participants had not been infected with these viruses) were used after trimming by removal of outlier values (the lowest 5% and highest 5%). The $\log_{10}$ transformed read count for each HIV peptide (after adding one read count) was then normalized to the average read count for all Ebola virus and rabies virus peptides of the respective sample. To generate a fold change value for each HIV peptide, while accounting for biased peptide representation in the library, the normalized value of the peptide was divided by the average of the normalized values for the same peptide observed across the mock immunoprecipitation reactions that were run on the same plate (a subtraction on the logarithmic scale).

Determination of antibody breadth—The term, “antibody breadth,” was used to indicate the number of unique non-overlapping epitopes that had high levels of antibody binding (z-scores > 10). Antibody breadth was determined for HIV and EBV peptides using network graphs as follows. The amino acid sequences of all peptides in the VirScan library (HIV or EBV) were first analyzed to identify sequence overlaps (linkages, defined as two peptides sharing an identical sequence at least 7 amino acids long). The linkages were used to construct an undirected network graph, where each node represented a peptide with high-level antibody binding, and each linkage between two nodes represented a sequence overlap between the two peptides. The number of linkages for each peptide defined its degree of connectivity. Peptides were then removed from the graph one at a time using the following approach. At each iteration, the peptide(s) with maximum connectivity was removed, and the degree of connectivity was recalculated for each of the remaining peptides. If multiple linked peptides had equivalent connectivity, the peptide with the lowest z-score was removed first. This process was repeated until the only remaining structures in the network were simple paths and cycles. For cycles (simple paths without end peptides), the peptide with the lowest z-score was removed first; this resulted in a simple path. Peptides were iteratively removed from simple paths in order to retain the greatest number of unlinked peptides. The number of remaining unlinked peptides was defined as the antibody breadth (Monaco et al., 2018).

Rate of change in antibody breadth—For each participant, we estimated the rate of change in antibody breadth over the time period from 9 months to 2 years after HIV infection. This was calculated by determining the difference in antibody breadth for samples collected closest to time points 9 months and 2 years after HIV infection, and dividing this value by the length of time between the two visits. The rate of change in CD4 cell count was derived in the same way, using samples that had associated CD4 cell count data. Since study participants were recommended to start ART when CD4 counts fell below 250 cells/mm3, time to ART initiation was used as a surrogate for disease progression. As described in the text, we examined differences in time to ART initiation between two groups: individuals

with increasing or stable antibody breadth and individuals with decreasing antibody breadth. Inspection of scaled Schoenfeld residual plots across time and tests for independence between the Schoenfeld residuals and time indicated no significant violation of the proportional hazards assumption of the model. The relationship between the rate of change in antibody breadth (and other factors) with time to ART initiation was determined using Cox proportional hazards models. The following factors were included in the analysis: age at seroconversion, CD4 cell count at the first visit after seroconversion, viral load set point, the rate of change in CD4 cell count, and time between HIV seroconversion and ART initiation. Viral load set point was defined as the median $\log_{10}$ viral load, excluding viral load results from the first HIV-positive visit, the visit prior to ART initiation, and any visits after ART initiation. Pearson correlation coefficients and their respective p values and 95% confidence intervals were used to describe the relationships between the factors analyzed. We also compared the time to ART initiation among individuals who experienced a decline in antibody breath between 9 months and 2 years, and those who had stable or increasing antibody breadth in this period. Statistical significance between the breadth measures and time to ART initiation was assessed using a non-parametric log-rank test and the semi-parametric Cox proportional-hazards model with a dichotomized variable for change in breadth rate (decreasing versus stable/increasing). Individuals who did not initiate ART were treated as right-censored. Survival curves were plotted based on the resulting hazard functions for the two groups.

Identification of peptides for estimating duration of HIV infection—The observed duration of infection ($\log_{10}$ transformed) was regressed on each of the normalized read count for each peptide, and the peptide with the strongest association was selected. To select additional peptides with independent information about duration of infection, we correlated the “residuals” (i.e., the differences between the observed and fitted values) from the above linear model against each of the remaining peptides, selected the peptide with the strongest association, and repeated this step twice more to generate a list of four peptides. Two of the four peptides had increased antibody binding over time since infection (positively associated with duration of infection), and two had decreasing antibody binding over time (negatively associated with duration of infection). Since the absolute values of the four peptide parameter estimates were almost identical, a simpler prediction model for duration of infection was calculated, using the sum of the normalized read counts for the positively-associated peptides, minus the sum of the normalized read counts for the negatively-associated peptides, as independent variable (reducing the number of linear model parameters from five to two). For the analysis of predicted duration of infection, generalized estimating equations (GEE) were used to account for auto-regressive correlation structure of samples from the same individual. No substantial model violations were detected by inspecting residual errors by histograms and quantile-quantile plots when investigating the linear models used to predict observed duration of infection. The validity and usefulness of the final model was corroborated using a completely independent validation dataset (72 samples from 32 participants in the GS Study; Table S1).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Funding for this work was provided by grants from the US National Institute of Allergy and Infectious Diseases (NIAID) (R01-AI095068 and UM1-AI068613 to S.H.E. and U24-AI118633 to H.B.L.) and grant OPP1155863 from the Bill and Melinda Gates Foundation to S.J.E. Additional support was provided by the Division of Intramural Research, NIAID, NIH. The authors thank the study team and the Zimbabwean and Ugandan participants of the GS study for providing samples and data for this research. The authors thank Dr. Joseph Margolick for critically reviewing the manuscript. K.K.'s primary appointment is in the JHSOM and the SKCCC at JH, which is partially funded through National Cancer Institute P30-CA006973.

REFERENCES

Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS, Certain L, Mandaliya K, Jaoko W, and Overbaugh J (2007). HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. J. Infect. Dis 195, 1177–1180. [PubMed: 17357054]

Brookmeyer R, Laeyendecker O, Donnell D, and Eshleman SH (2013). Cross-sectional HIV incidence estimation in HIV prevention research.J. Acquir. Immune Defic. Syndr 63 (Suppl 2), S233–S239.

Busch MP, Pilcher CD, Mastro TD, Kaldor J, Vercauteren G, Rodriguez W, Rousseau C, Rehle TM, Welte A, Averill MD, and Garcia Calleja JM; WHO Working Group on HIV Incidence Assays (2010). Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation. AIDS 24, 2763–2771. [PubMed: 20975514]

Coates TJ, Kulich M, Celentano DD, Zelaya CE, Chariyalertsak S, Chingono A, Gray G, Mbwambo JK, Morin SF, Richter L, et al.; NIMH Project Accept (HPTN 043) study team (2014). Effect of community-based voluntary counselling and testing on HIV incidence and social and behavioural outcomes (NIMH Project Accept; HPTN 043): a cluster-randomised trial. Lancet Glob. Health 2, e267–e277. [PubMed: 25103167]

Costello C, Tang J, Rivers C, Karita E, Meizen-Derr J, Allen S, and Kaslow RA (1999). HLA-B*5703 independently associated with slower HIV-1 disease progression in Rwandan women. AIDS 13, 1990–1991. [PubMed: 10513667]

Curtis KA, Kennedy MS, Charurat M, Nasidi A, Delaney K, Spira TJ, and Owen SM (2012). Development and characterization of a bead-based, multiplex assay for estimation of recent HIV type 1 infection. AIDS Res. Hum. Retroviruses 28, 188–197. [PubMed: 21585287]

Delhalle S, Schmit JC, and Chevigné A (2012). Phages and HIV-1: from display to interplay. Int. J. Mol. Sci 13, 4727–4794. [PubMed: 22606007]

Dotsey EY, Gorlani A, Ingale S, Achenbach CJ, Forthal DN, Felgner PL, and Gach JS (2015). A high throughput protein microarray approach to classify HIV monoclonal antibodies and variant antigens. PLoS ONE 10, e0125581. [PubMed: 25938510]

Fahey JL, Taylor JM, Detels R, Hofmann B, Melmed R, Nishanian P, and Giorgi JV (1990). The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. N. Engl. J. Med 322, 166–172. [PubMed: 1967191]

Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH, and Busch MP (2003). Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS 17, 1871–1879. [PubMed: 12960819]

Geiß Y, and Dietrich U (2015). Catch me if you can—the race between HIV and neutralizing antibodies. AIDS Rev. 17, 107–113. [PubMed: 26035168]

Guy R, Gold J, Calleja JM, Kim AA, Parekh B, Busch M, Rehle T, Hargrove J, Remis RS, and Kaldor JM; WHO Working Group on HIV Incidence Assays (2009). Accuracy of serological assays for detection of recent infection with HIV and estimation of population incidence: a systematic review. Lancet Infect. Dis 9, 747–759. [PubMed: 19926035]

Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T, Kang S, Ceradini D, Jin Z, Yazdanbakhsh K, et al. (1996). The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat. Med 2, 1240–1243. [PubMed: 8898752]

Justman JE, Mugurungi O, and El-Sadr WM (2018). HIV population surveys—bringing precision to the global response. N. Engl. J. Med 378, 1859–1861. [PubMed: 29768142]

Kadelka C, Liechti T, Ebner H, Schanz M, Rusert P, Friedrich N, Stiegeler E, Braun DL, Huber M, Scherrer AU, et al.; Swiss HIV Cohort Study (2018). Distinct, IgG1-driven antibody response landscapes demarcate individuals with broadly HIV-1 neutralizing activity. J. Exp. Med 215, 1589–1608. [PubMed: 29794117]

Kassanjee R, Pilcher CD, Keating SM, Facente SN, McKinney E, Price MA, Martin JN, Little S, Hecht FM, Kallas EG, et al.; Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA) (2014). Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. AIDS 28, 2439–2449. [PubMed: 25144218]

Koenig O, Walker T, Perle N, Zech A, Neumann B, Schlensak C, Wendel HP, and Nolte A (2013). New aspects of gene-silencing for the treatment of cardiovascular diseases. Pharmaceuticals (Basel) 6, 881–914. [PubMed: 24276320]

Konikoff J, Brookmeyer R, Longosz AF, Cousins MM, Celum C, Buch-binder SP, Seage GR 3rd, Kirk GD, Moore RD, Mehta SH, et al. (2013). Performance of a limiting-antigen avidity enzyme immunoassay for cross-sectional estimation of HIV incidence in the United States. PLoS ONE 8, e82772. [PubMed: 24386116]

Laeyendecker O, Brookmeyer R, Mullis CE, Donnell D, Lingappa J, Celum C, Baeten JM, Campbell MS, Essex M, de Bruyn G, et al.; Partners in Prevention HSV/HIV Transmission Study Team (2012a). Specificity of four laboratory approaches for cross-sectional HIV incidence determination: analysis of samples from adults with known nonrecent HIV infection from five African countries. AIDS Res. Hum. Retroviruses 28, 1177–1183. [PubMed: 22283149]

Laeyendecker O, Brookmeyer R, Oliver AE, Mullis CE, Eaton KP, Mueller AC, Jacobson LP, Margolick JB, Brown J, Rinaldo CR, et al.; Multicenter Aids Cohort Study Macs (2012b). Factors associated with incorrect identification of recent HIV infection using the BED capture immunoassay. AIDS Res. Hum. Retroviruses 28, 816–822. [PubMed: 22014036]

Laeyendecker O, Brookmeyer R, Cousins MM, Mullis CE, Konikoff J, Donnell D, Celum C, Buchbinder SP, Seage GR 3rd, Kirk GD, et al. (2013). HIV incidence determination in the United States: a multiassay approach. J. Infect. Dis 207, 232–239. [PubMed: 23129760]

Laeyendecker O, Konikoff J, Morrison DE, Brookmeyer R, Wang J, Celum C, Morrison CS, Abdool Karim Q, Pettifor AE, and Eshleman SH (2018). Identification and validation of a multi-assay algorithm for cross-sectional HIV incidence estimation in populations with subtype C infection.J. Int. AIDS Soc. 21, e25082.

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. [PubMed: 19261174]

Larman HB, Zhao Z, Laserson U, Li MZ, Ciccia A, Gakidis MA, Church GM, Kesari S, Leproust EM, Solimini NL, and Elledge SJ (2011). Autoantigen discovery with a synthetic human peptidome. Nat. Biotechnol 29, 535–541. [PubMed: 21602805]

Longosz AF, Morrison CS, Chen PL, Arts E, Nankya I, Salata RA, Franco V, Quinn TC, Eshleman SH, and Laeyendecker O (2014). Immune responses in Ugandan women infected with subtypes A and D HIV using the BED capture immunoassay and an antibody avidity assay. J. Acquir. Immune Defic. Syndr 65, 390–396. [PubMed: 24583613]

Longosz AF, Morrison CS, Chen PL, Brand HH, Arts E, Nankya I, Salata RA, Quinn TC, Eshleman SH, and Laeyendecker O (2015). Comparison of antibody responses to HIV infection in Ugandan women infected with HIV subtypes A and D. AIDS Res. Hum. Retroviruses 31, 421–427. [PubMed: 25317854]

Margolick JB, Gange SJ, Detels R, O'Gorman MR, Rinaldo CR Jr., and Lai S (2006). Impact of inversion of the CD4/CD8 ratio on the natural history of HIV-1 infection. J. Acquir. Immune Defic. Syndr 42, 620–626. [PubMed: 16868499]

Mohan D, Wansley DL, Sie BM, Noon MS, Baer AN, Laserson U, and Larman HB (2018). PhIP-Seq characterization of serum antibodies using oligonucleotide encoded peptidomes. Nat. Protoc 13, 1958–1978. [PubMed: 30190553]

Monaco D, Kottapalli D, Yuan T, Breitwieser F, Anderson D, Wijaya L, Tan K, Chia WN, Kammers K, Caturegli M, et al. (2018). Deconvoluting virome-wide antiviral antibody profiling data. bioRxiv. 10.1101/333625.

Morrison CS, Demers K, Kwok C, Bulime S, Rinaldi A, Munjoma M, Dunbar M, Chipato T, Byamugisha J, Van Der Pol B, et al. (2010). Plasma and cervical viral loads among Ugandan and Zimbabwean women during acute and early HIV-1 infection. AIDS 24, 573–582. [PubMed: 20154581]

Morrison CS, Chen PL, Nankya I, Rinaldi A, Van Der Pol B, Ma YR, Chipato T, Mugerwa R, Dunbar M, Arts E, and Salata RA (2011). Hormonal contraceptive use and HIV disease progression among women in Uganda and Zimbabwe. J. Acquir. Immune Defic. Syndr 57, 157–164. [PubMed: 21358412]

Murphy G, and Parry JV (2008). Assays for the detection of recent infections with human immunodeficiency virus type 1. Euro Surveill. 13, 36.

Ng OT, Laeyendecker O, Redd AD, Munshaw S, Grabowski MK, Paquet AC, Evans MC, Haddad M, Huang W, Robb ML, et al. (2014). HIV type 1 polymerase gene polymorphisms are associated with phenotypic differences in replication capacity and disease progression. J. Infect. Dis 209, 66–73. [PubMed: 23922373]

Pantaleo G, Demarest JF, Schacker T, Vaccarezza M, Cohen OJ, Daucher M, Graziosi C, Schnittman SS, Quinn TC, Shaw GM, et al. (1997). The qualitative nature of the primary immune response to HIV infection is a prognosticator of disease progression independent of the initial level of plasma viremia. Proc. Natl. Acad. Sci. USA 94, 254–258. [PubMed: 8990195]

Sharma SK, and Soneja M (2011). HIV & immune reconstitution inflammatory syndrome (IRIS). Indian J. Med. Res 134, 866–877. [PubMed: 22310819]

Touloumi G, Pantazis N, Pillay D, Paraskevis D, Chaix ML, Bucher HC, Kücherer C, Zangerle R, Kran AM, and Porter K; CASCADE collaboration in EuroCoord (2013). Impact of HIV-1 subtype on CD4 count at HIV seroconversion, rate of decline, and viral load set point in European seroconverter cohorts. Clin. Infect. Dis 56, 888–897. [PubMed: 23223594]

Wei X, Liu X, Dobbs T, Kuehl D, Nkengasong JN, Hu DJ, and Parekh BS (2010). Development of two avidity-based assays to detect recent HIV type 1 seroconversion using a multisubtype gp41 recombinant protein. AIDS Res. Hum. Retroviruses 26, 61–71. [PubMed: 20063992]

Wendel SK, Longosz AF, Eshleman SH, Blankson JN, Moore RD, Keruly JC, Quinn TC, and Laeyendecker O (2017). Short communication: the impact of viral suppression and viral breakthrough on Limited-Antigen Avidity assay results in individuals with clade B HIV infection. AIDS Res. Hum. Retroviruses 33, 325–327. [PubMed: 27875908]

Xu GJ, Kula T, Xu Q, Li MZ, Vernon SD, Ndung'u T, Ruxrungtham K, Sanchez J, Brander C, Chung RT, et al. (2015). Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science 348, aaa0698. [PubMed: 26045439]

Yuan T, Mohan D, Laserson U, Ruczinski I, Baer A, and Larman HB (2018). Improved analysis of phage immunoprecipitation sequencing (PhIP-Seq) data using a z-score algorithm. bioRxiv. 10.1101/285916.

Highlights

- Reactivity to >3,300 HIV peptides is quantified using a phage display system
- Antibody diversity (breadth) is associated with CD4 decline and HIV treatment
- Defined sets of peptides have increasing or decreasing reactivity over time
- A prototype four-peptide “serosignature” predicts duration of HIV infection

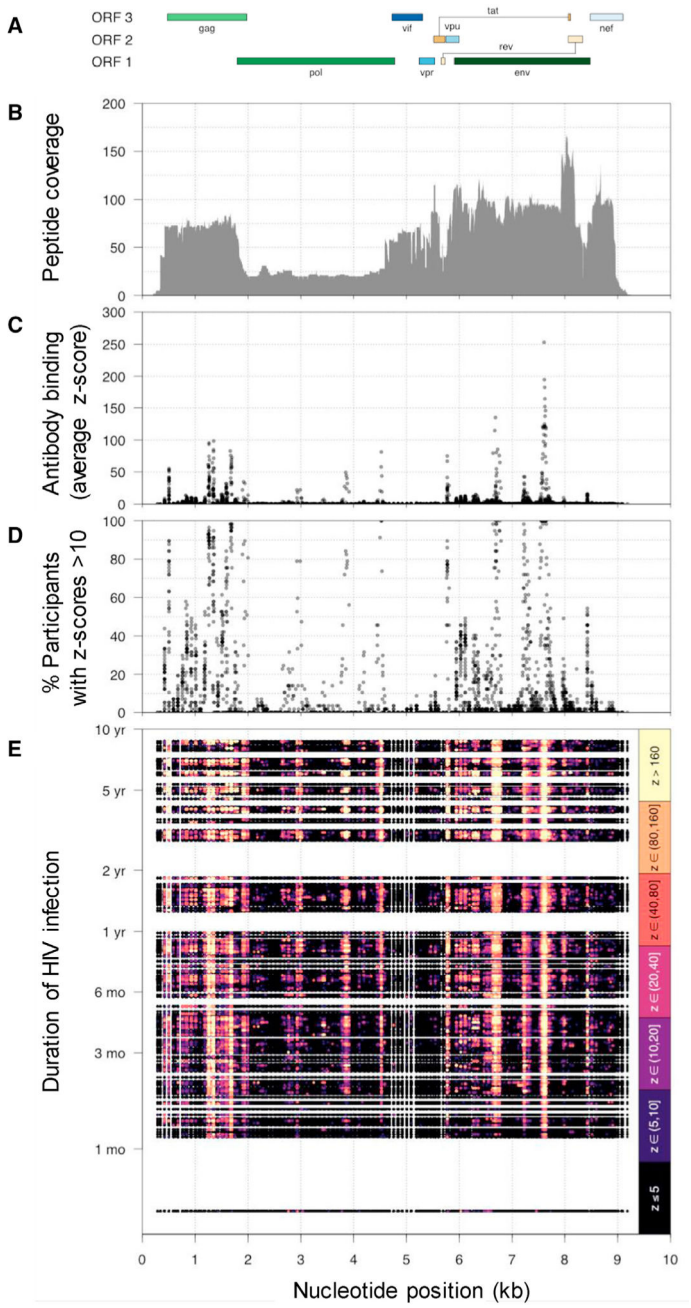


Figure 1. Antibody Reactivity to Peptides Spanning the HIV Proteome

(A) The size and position of open reading frames (ORFs) in the HIV genome.

(B–D) Plotted relative to genomic coordinates for HIV (HXB2, NCBI #NC_001802), shown at the bottom of the figure.

(B) The plot shows the number of peptide tiles encoded by the VirScan library at each position across the HIV genome.

(C) The plot shows the average level of antibody binding (average z-score) for each peptide for the 403 samples in the discovery sample set; each dot represents antibody binding for a single peptide in the VirScan library.

(D) The plot shows the percentage of study participants who had a high level of antibody binding for each peptide (z-score > 10).
(E) A heatmap of the level of antibody binding for peptides in the VirScan library as a function of duration of HIV infection. The position of peptides is shown on the x axis; the duration of infection is shown on the y axis. z-scores are noted according to the color bar on the right; lighter colors (higher z-scores) indicate a higher level of antibody binding. For each sample, data are plotted in order of increasing z-scores, since many points were overlapping. ORF, open reading frame; mo, months; yr, years; kb, kilobases.

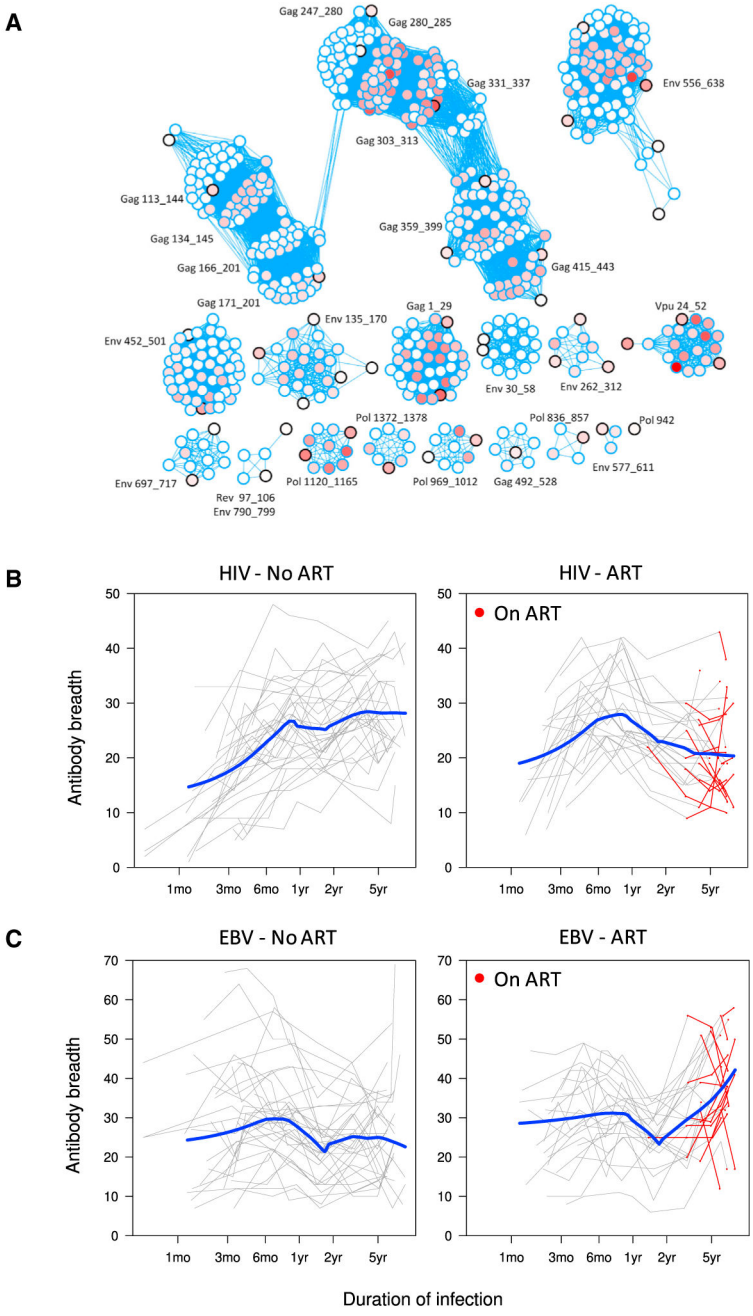


Figure 2. Breadth of Antibody Reactivity

(A) The relationships between peptides that were highly enriched (z-score > 10) are displayed as a network graph; data are from a single representative sample. Peptides (nodes) are indicated by circles. Darker red indicates peptides with higher z-scores. Overlapping peptides that share amino acid sequences form clusters in the graph; the position of the peptides in HIV proteins is noted for each cluster (the HIV protein is listed first, followed by numbers that represent the range of amino acid positions of the N termini of peptides in the cluster). Peptides are linked (connected by lines) if they share an identical sequence of at least seven consecutive amino acids. In this case, network graph analysis of 573 reactive

peptides identified 45 unique peptide specificities (circles outlined in black), corresponding to an antibody breadth value of 45.

(B and C) Antibody breadth is plotted as a function of duration of HIV infection. The top two graphs (B) show breadth data for HIV peptides; the bottom two graphs (C) show breadth data for EBV peptides. Each line represents results from a single study participant. The two graphs on the left show data for participants who did not start anti-retroviral treatment in the GS study (no ART; N = 33); the two graphs on the right show data for participants who reported starting antiretroviral treatment (ART; N = 24). Data from samples collected after treatment initiation are shown in red (on ART). Dark blue lines indicate the locally weighted regression (lowess) curves for all participants in each graph. Additional analyses of these data are shown in Figure S1.

Env, envelope; Pol, polymerase; Gag, group-specific antigen; Rev, HIV regulatory protein; Vpu, viral protein U; ART, antiretroviral treatment; mo, months; yr, years; EBV, Epstein-Barr virus.

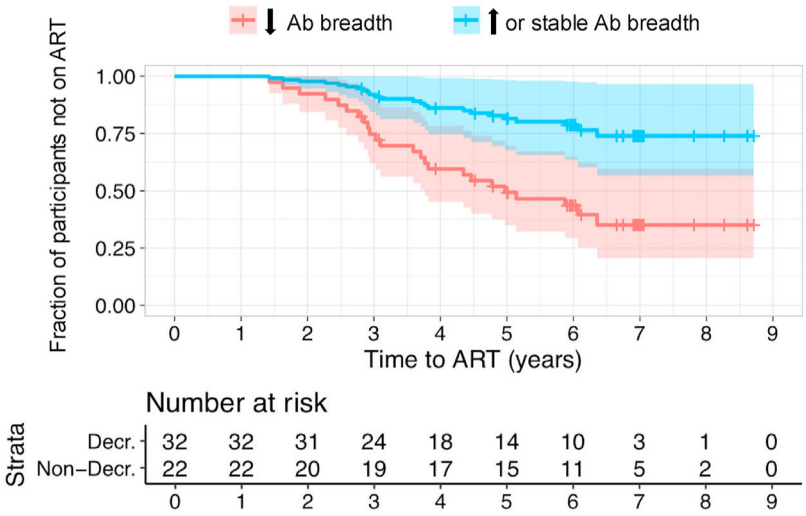


Figure 3. Relationship between Changes in Antibody Breadth and Time to ART

Time-to-event (survival) analysis is shown for the outcome of time from HIV infection to antiretroviral treatment initiation (time to ART), comparing participants with declining versus stable or increasing antibody breadth (shown in red and blue, respectively). The change in antibody breadth was calculated for the time period between 9 months and 2 years after HIV infection, using samples collected closest to these dates. The median sample collection times were 0.8 years for the visit 9 months after infection (range, 0.55–0.98 years) and 1.5 years for the visit 2 years after infection (range, 1.26–3.12 years); the median time to ART initiation was 3.34 years (range, 1.16–6.35 years). Data from two participants were removed for this analysis (one did not have viral load data and one started ART <2 years after HIV infection). The survival curves are based on estimated hazard ratios (lines) with 95% confidence intervals (shaded areas). The number of participants at risk (number at risk; not yet on ART) at each time point is shown below the graph for each participant group. Ab, antibody; ART, antiretroviral therapy; Decr, decreasing antibody breadth; Non-Decr, stable or increasing antibody breadth.

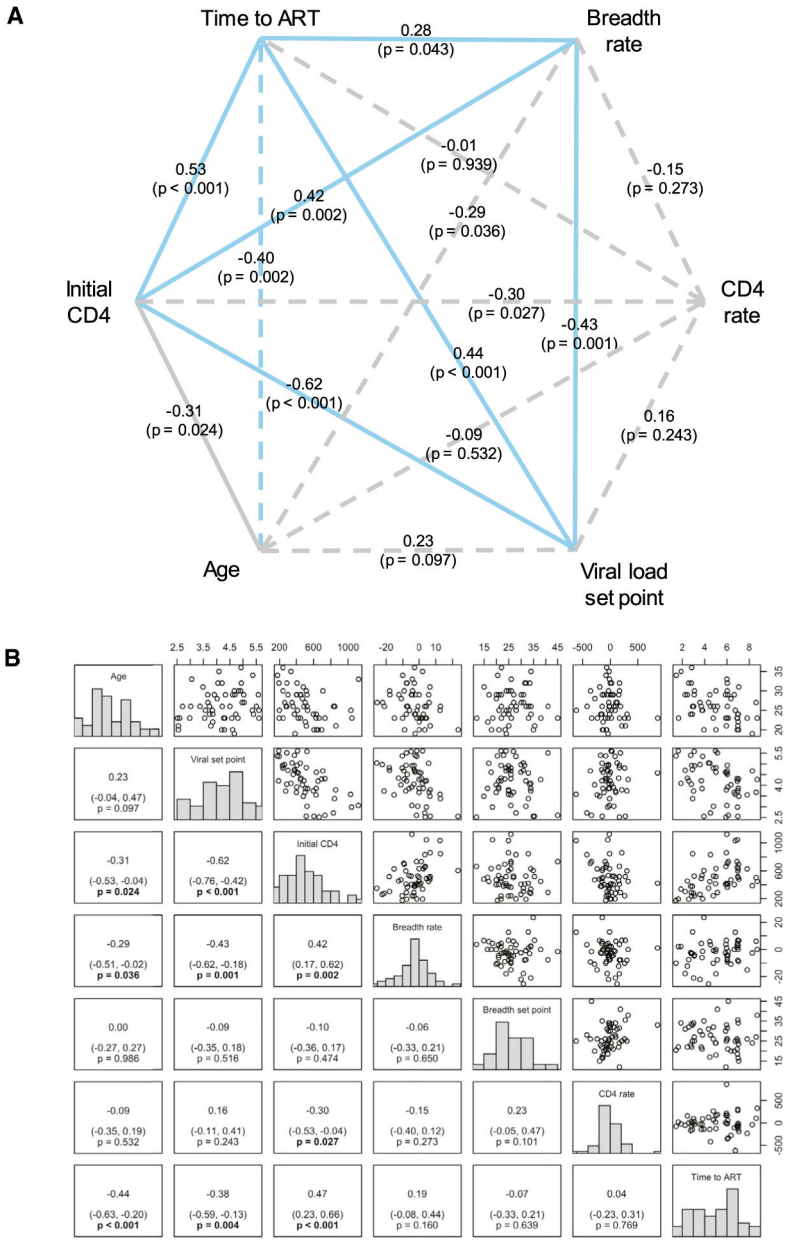


Figure 4. Association of Changes in Antibody Breadth and Other Factors

We evaluated the relationship between the changes in antibody breadth between 9 months and 2 years after infection, time to initiation of antiretroviral therapy (ART), and other factors.

(A) This plot shows univariate (pairwise) associations, reported as estimated Pearson correlation coefficients and respective p values, between pairs of factors. Solid lines indicate correlations that were statistically significant after correction for multiple comparisons ($p < 0.05/ 15 = 0.0033$).

(B) The array shows histograms of data for factors evaluated for their association with time to ART initiation (diagonal). The array also shows scatterplots of the data (upper right) and

summary statistics (lower left) for all pairwise comparisons. Summary statistics include the estimated Pearson correlations with 95% confidence intervals and the respective p values. Units for variables are as follows: age (years); viral load set point (log_{10} copies/mL); baseline CD4 cell count (baseline CD4; cells/mm^3); change in the antibody breadth between 9 months and 2 years after HIV infection; change in CD4 cell count between 9 months and 2 years after HIV infection (cells/mm^3); time to ART (years).

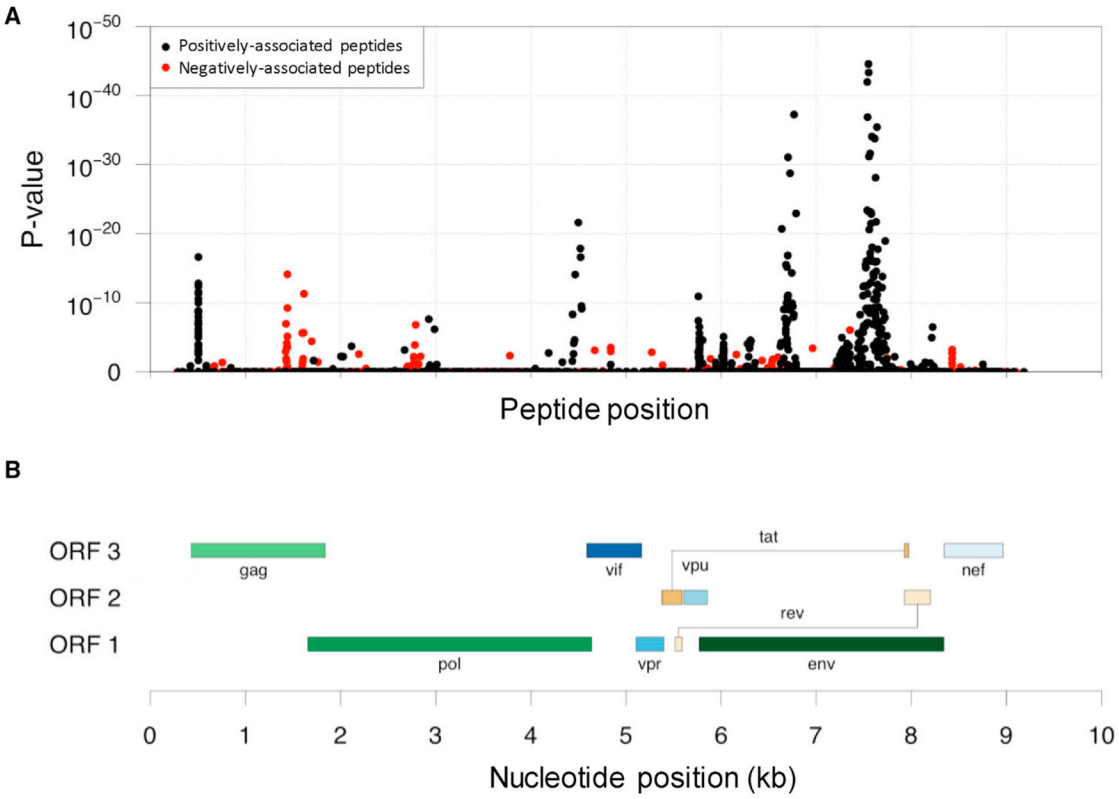


Figure 5. Association of Antibody Binding and the Duration of HIV Infection

(A) The plot shows data evaluating the association of antibody binding (normalized read counts) and the duration of HIV infection for 3,327 peptides in the VirScan library that had well-defined positions in the HIV genome. p values were calculated using generalized estimation equations to account for the dependency between measurements over time from the same individual, and were adjusted using the Bonferroni correction based on all 3,384 identified HIV peptides. The x axis shows the position of each peptide, and the y axis shows the corresponding Bonferroni adjusted p value. Black dots represent peptides where antibody binding was positively associated with the duration of infection (266 peptides with adjusted $p < 0.05$); red dots represent peptides where antibody binding was negatively associated with the duration of infection (43 peptides with adjusted $p < 0.05$).

(B) The figure shows the position of open reading frames (ORFs) in the HIV genome (reproduced from Figure 1A).

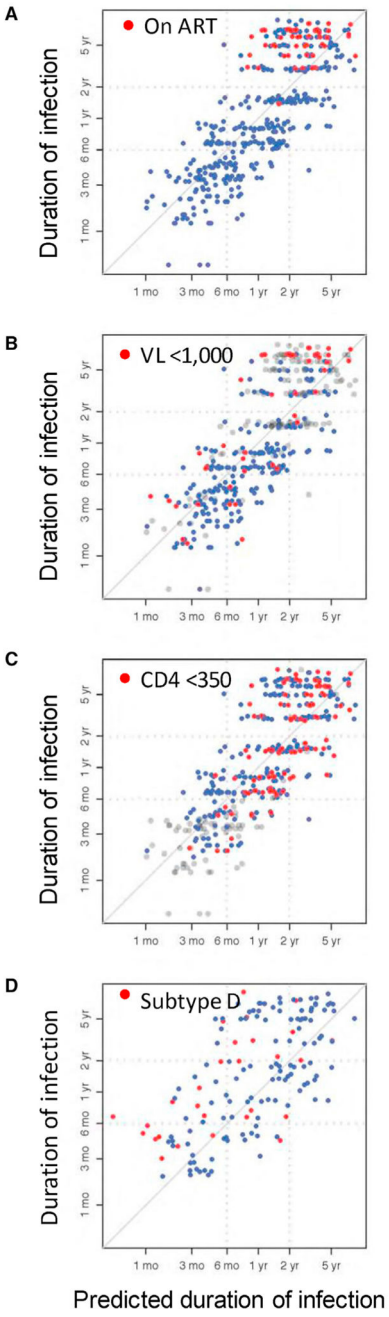


Figure 6. Use of a Four-Peptide Model to Predict Duration of HIV Infection

Four peptides were selected from the VirScan library that had the strongest independent association between antibody binding and the duration of HIV infection. This included two peptides that had increasing antibody binding over time, and two peptides that had decreasing antibody binding over time (Figure S3).

(A–C) Data from these four peptides (normalized read counts) were summed to generate a composite antibody binding score for each of the 403 samples in the discovery sample set that was used to identify the four peptides (Table S1). The plots show the observed duration of HIV infection (y-axes) and the duration of HIV infection that was predicted using a

simple linear regression model based on the composite antibody binding score for the four peptides (x-axes). In the graphs, each dot represents data from a single sample. The same data are plotted in (A)–(C). Red dots represent data obtained for samples collected after antiretroviral treatment (ART) initiation (A), for samples with viral load <1,000 copies/mL (B), and for samples with CD4 cell counts <350 cells/mm^3 (C).

(D) The four-peptide model described above was used to predict the duration of HIV infection in an independent sample set that included 72 samples from 32 participants in the GS study (validation sample set; Table S1).

Data were analyzed and plotted using the same methods used for (A)–(C). Red dots represent data obtained for samples with subtype D HIV. Correlation values are r = 0.79 and r = 0.64 for (A)–(C) and (D), respectively, under the assumption that data points are independent.

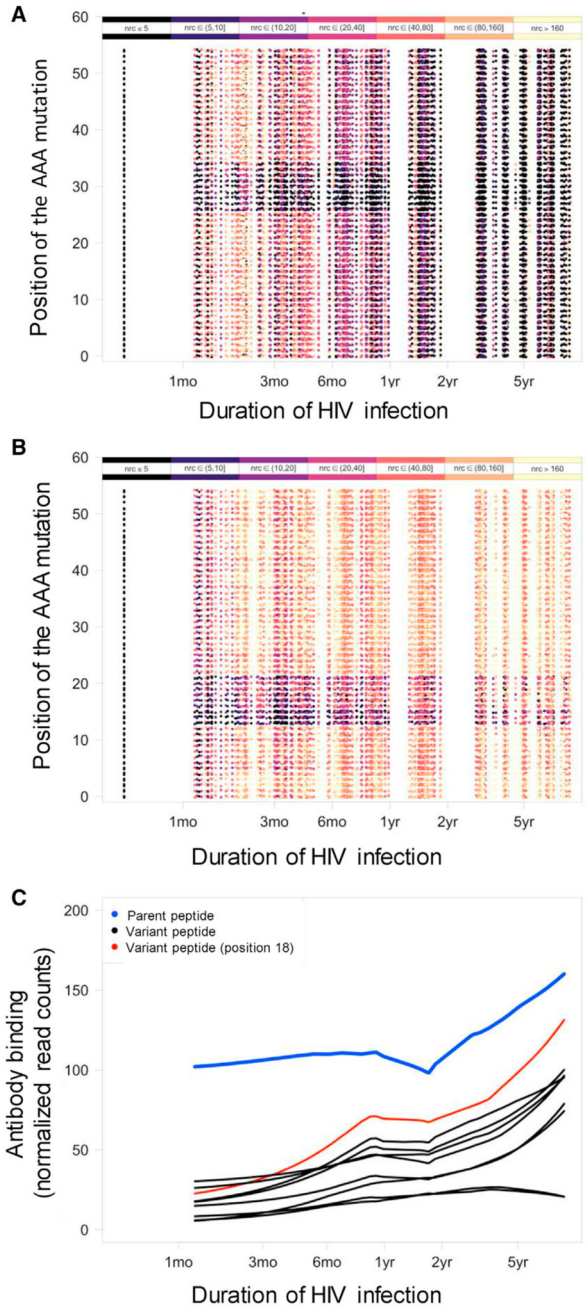


Figure 7. Peptide Engineering

Antibody data are shown for two representative parent peptides and their respective variant peptides generated by alanine scanning mutagenesis. High levels of antibody binding (z-scores > 10) were observed in samples for all but one of the 57 participants for parent peptide A (98.2%) and for all 57 participants for parent peptide B.

(A and B) These panels show heatmaps of antibody binding for each set of peptides (the parent peptide and 54 variant peptides with triple alanine substitutions at different positions within the peptide); the position of the alanine substitution in each variant peptide is shown on the y-axes. These parent peptides were selected due to their decreasing reactivity (A) or

increasing reactivity (B) over time. Antibody binding data are shown as a function of duration of HIV infection (x-axes).

(C) The blue line shows antibody binding data (normalized read counts) for the parent peptide included in the analysis in (B) (parent peptide B) and selected variant peptides. Black lines show data for variant peptides with triple alanine substitutions at amino acids 12–17 and 19–21; the red line shows data for the variant peptide with the triple alanine substitution at amino acid 18.

nrc, normalized read count; mo, months; yr, years.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat anti-human IgG-UNLB	Southern Biotech	Cat. # 2040–01; RRID:AB_2795640
Goat F(ab′)2 anti-human IgG-HRP	Southern Biotech	Cat. # 2042–05; RRID:AB_2795660
Human IgG Isotype Control	Life Technologies	Cat. # 02–7102
Bacterial and Virus Strains		
BLT5403	EMD Millipore Sigma	Cat. # 70014–3
T7Select-10–3b	EMD Millipore Sigma	Cat. # 70548–3
Biological Samples		
See Table S1		N/A
Chemicals, Peptides, and Recombinant Proteins		
Phosphate Buffered Saline	Thermo Fisher Scientific	Cat. # 10010023
Fetal Bovine Serum	Corning	Cat. # 35–011-CV
Dynabeads Protein A	Invitrogen	Cat. # 10002D
Dynabeads Protein G	Invitrogen	Cat. # 10004D
Tris-HCI	Sigma-Aldrich	Cat. # RES3098T-B7
150 mM Sodium Chloride	Sigma-Aldrich	Cat. # S7653
Tween-20	Sigma-Aldrich	Cat. # P1379
NP-40	Sigma-Aldrich	Cat. # 492016
Herculase II polymerase	Agilent	Cat. # 600679
dNTPs	Applied Biosystems	Cat. # N8080261
Critical Commercial Assays		
KAPA Library Quantification Kit	KAPA Biosystems	Cat # KK4828
NucleoSpin Gel and PCR Clean-up	Macherey–Nagel	Cat # 740609.50
Oligonucleotides		
PCR1 Forward Primer 5′-ATAAAGGTGAGGGTAATGTC–3′	Integrated DNA Technology	T7-Pep2_PCR1_F
PCR1 Reverse Primer 5′–CTGGAGTTCAGACGTGTGCTCTT CCGATCAGTTACTCGAGCTTATCGTC–3′	Integrated DNA Technology	T7-Pep2_PCR1_R+ad_min
PCR2 Forward Primer 5′–AATGATACGGCGACCACCGAGA TCTACAC-XX-GGAGCTGTCGTATTCCAGTC–3′	Integrated DNA Technology	T7-Pep2_PCR2_F_BCX_P5
PCR2 Reverse Primer 5′–CAAGCAGAAGACGGCATACGAGATXX-GTGACTGGAGTTCAGACGTGTGCTC–3′	Integrated DNA Technology	Reverse ad_min_BCX_P7
Sequencing Primer 5′-GGTGTGATGCTCGGGGATCCAGG AATTCCGCTGCGT–3′	Integrated DNA Technology	T7-VirScan_SP

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sequencing Primer 5′–GATCGGAAGAGCACACGTCTGAA CTCCAGTCAC–3′	Integrated DNA Technology	i7_SP