2016

Within-Host Evolution Of Hiv-1: Novel Pathways Of Virus Escape From Cellular And Humoral Immunity

Edward Kreider
University of Pennsylvania, fkreider@upenn.edu

Within-Host Evolution Of Hiv-1: Novel Pathways Of Virus Escape From Cellular And Humoral Immunity

Abstract
Longitudinal HIV-1 single genome sequencing (SGS), which permits unambiguous genetic characterization of circulating viral strains without introduction of PCR error, can be used to identify sites in the viral genome that are under selective pressure. Following transmission, the earliest sites under positive selection often fall in cytotoxic T lymphocyte (CTL) epitopes. During escape from CTL immune pressure, viral sequences typically exhibit nonsynonymous mutations within the span of the cognate T cell epitope. I applied SGS to study sequence evolution in the HIV-1 5' leader sequence, which is thought to be translationally silent. I observed mutational patterns consistent with CTL escape and demonstrated that the HIV-1 5' leader expresses T cell antigens from non-canonical one-off AUG codons (e.g. CUG). While these non-canonical start codons can be mutated during CTL escape, a reverse transcriptase overextension error periodically restores a one-off AUG within the 5' leader. As infection ensues, sites under selection within the gene encoding the viral envelope glycoprotein (Env) often fall within autologous neutralizing antibody epitopes. In a subset of individuals, the strain-specific neutralizing antibody response develops into a broadly cross-reactive neutralizing antibody (bnAb) response. To understand what factors influence bnAb ontogeny, I used SGS to study Env evolution both during natural infection and immunotherapy. I found viral diversification in bnAb contact residues and divergence of the virus population into multiple persistent lineages to precede bnAb development. Taken together, these data demonstrate that longitudinal HIV-1 SGS can be used to discover novel aspects of virus biology and host-pathogen interactions.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Beatrice H. Hahn

Subject Categories
Microbiology

WITHIN-HOST EVOLUTION OF HIV-1: NOVEL PATHWAYS OF VIRUS ESCAPE FROM CELLULAR AND HUMORAL IMMUNITY

Edward F. Kreider

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

_____

Dr. Beatrice H. Hahn, MD

Professor, Departments of Medicine and Microbiology

Graduate Group Chairperson

_____

Dr. Daniel S. Kessler, PhD

Associate Professor of Cell and Developmental Biology

Dissertation Committee

James A. Hoxie, MD (Chair), Professor, Department of Medicine

Robert W. Doms, MD, PhD, Professor, Department of Pathology and Laboratory Medicine

Scott E. Hensley, PhD, Associate Professor, The Wistar Institute

Rahul M. Kohli, MD, PhD, Assistant Professor, Department of Medicine

ACKNOWLEDGMENTS

I would first and foremost like to thank my mentor, Dr. Beatrice Hahn. She has helped me flourish as a scientist and provided me with unbelievable opportunities. Her mentorship over the past four years has molded me as a both a person and an investigator. I cannot thank her enough.

Dr. George Shaw has also played an integral role both in my training and mentorship. His insistence that we understand every stripe in every highlighter plot forced me to think critically about every datum and even led to unexpected discoveries.

My thesis committee members Dr. James Hoxie, Dr. Robert Doms, Dr. Rahul Kohli, and Dr. Scott Hensley have all provided me with guidance and mentorship on personal, academic, and scientific matters. Dr. Doms is a major reason why I am here today and I am so thankful for the guidance and opportunities he has provided me.

Dr. Yingying Li welcomed me to the lab, taught me everything that I know about sequencing, and cared for me every day of my PhD training. Dr. Hui Li is an incredible collaborator and friend whose door was always open. Dr. Gerald Learn provided me with a foundation in sequence analysis and is a treasured colleague. Dr. Katherine Bar was a mentor and, to my great pleasure, a collaborator. To the Hahn lab graduate students who have been a family throughout these past four years, in particular Hannah Barbian and Shilpa Iyer, thank you so much for being there. And Ranjit Warrier, whose calming presence got me through tough days. Additionally, many thanks are due to all of the collaborators who welcomed me into their labs: Dr. Persephone Borrow, Dr. Andrew McMichael, and Dr. Bette Korber. As well to the CHAVI Duke family, particularly Dr. Barton Haynes and Dr. Mattia Bonsignori

I also need to thank my social support network: Liz, Sarah, Andrew, Alice, Justin, and my entire MSTP class, and Skip, Maggie, and Maureen. Finally, to my family and Fraz, thank you for putting up with me.

ABSTRACT

WITHIN-HOST EVOLUTION OF HIV-1: NOVEL PATHWAYS OF VIRUS ESCAPE FROM CELLULAR AND HUMORAL IMMUNITY

Edward F. Kreider

Beatrice H. Hahn, MD

Longitudinal HIV-1 single genome sequencing (SGS), which permits unambiguous genetic characterization of circulating viral strains without introduction of PCR error, can be used to identify sites in the viral genome that are under selective pressure. Following transmission, the earliest sites under positive selection often fall in cytotoxic T lymphocyte (CTL) epitopes. During escape from CTL immune pressure, viral sequences typically exhibit nonsynonymous mutations within the span of the cognate T cell epitope. I applied SGS to study sequence evolution in the HIV-1 5' leader sequence, which is thought to be translationally silent. I observed mutational patterns consistent with CTL escape and demonstrated that the HIV-1 5' leader expresses T cell antigens from non-canonical one-off AUG codons (*e.g.* CUG). While these non-canonical start codons can be mutated during CTL escape, a reverse transcriptase overextension error periodically restores a one-off AUG within the 5' leader. As infection ensues, sites under selection within the gene encoding the viral envelope glycoprotein (Env) often fall within autologous neutralizing antibody epitopes. In a subset of individuals, the strain-specific

neutralizing antibody response develops into a broadly cross-reactive neutralizing antibody (bnAb) response. To understand what factors influence bnAb ontogeny, I used SGS to study Env evolution both during natural infection and immunotherapy. I found viral diversification in bnAb contact residues and divergence of the virus population into multiple persistent lineages to precede bnAb development. Taken together, these data demonstrate that longitudinal HIV-1 SGS can be used to discover novel aspects of virus biology and host-pathogen interactions.

TABLE OF CONTENTS

List of Tables

List of Illustrations

CHAPTER 1

INTRODUCTION

Edward F. Kreider

Departments of Medicine and Microbiology, Perelman School of Medicine at the University of Pennsylvania

All figures appear at the end of this chapter

Section 1.1 – AIDS and HIV-1

In 1981, the Centers for Disease Control published a Morbidity and Mortality Weekly Report describing five cases *P. jiroveci* pneumonia (formerly known as *P. carinii* pneumonia) in young, otherwise healthy gay men (www.cdc.gov/mmwr). This MMWR would mark the first official reporting of Acquired Immunodeficiency Syndrome (AIDS) in the United States (*1*). Since 1981, over 35 million people have died from AIDS-related illnesses worldwide (www.UNAIDS.org). The causative agent is a lentivirus called human immunodeficiency virus type 1 (HIV-1, Figures 1.1 and 1.2). HIV-1 primarily infects CD4+ T cells (*2*, *3*). Viral entry into a new target cell is mediated by the Env glycoprotein, which engages the host cell receptor (CD4) and a coreceptor (*e.g.* CCR5, CXCR4) (*3*). These events trigger a series of Env conformational changes that ultimately result in fusion of the viral and host membranes and delivery of the vial payload into the host cell cytoplasm (*3*). After entry, HIV-1 undergoes reverse transcription and integration (*4*). During reverse transcription, the single stranded RNA genome is converted into a double stranded DNA provirus (*5*). This DNA provirus is then trafficked to the nucleus where it is integrated into the host genome (*4*).

Section 1.2 – Biology of the 5' leader sequence

The HIV-1 5' leader is a conserved RNA element that lies at the 5' end of every HIV-1 RNA transcript and performs multiple essential replicative functions. It extends from nucleotide 1 of R through the *gag* AUG, approximately 336 nucleotides in most subtype B and C viruses (Figure 1.3). Given its placement upstream of all coding sequences, the 5' leader has the alternative designation of the 5' untranslated region. The various functions of this region depend on proper folding of mRNA secondary structure (*6*, *7*) and/or exact sequence identity (*5*).

R elements

The first two RNA features of the 5' leader – TAR and the polyA signal – fall in the retroviral Repeat element (R). As shown in Figure 1.1, the DNA provirus is flanked by two long terminal repeats (LTRs). Promoter and enhancer elements within the Unique 3 (U3) element of the 5' LTR are bound by a variety of host transcription factors, like NF-κB (*8*), that regulate transcription of viral RNA (*9*). Early studies of the HIV-1 LTR revealed that an additional regulatory element fell downstream of the transcription start site and required expression of a factor in *trans* for efficient transcription (the transactivator of transcription, Tat). This additional *cis* acting regulatory domain called TAR is encoded by the first 59 nucleotides (nts) of R (*9*). It was found that TAR RNA, not DNA, is bound by a viral protein called Tat via a bulge in the stem loop structure (*10*). If the TAR RNA element is properly folded and Tat is expressed, viral RNA is

efficiently transcribed (*11*, *12*). If Tat is not expressed, HIV-1 elongation is potently restricted by a host factor called negative elongation factor (*9*). As a result, Tat acts in a positive feedback loop in which Tat expression drives higher levels of transcription and more Tat expression. The next R element, the polyA signal, is silenced at the 5' end of the genome such that premature polyadenylation does not occur (*13-15*).

Primer binding site and reverse transcription

The next 5' leader element, the primer binding site (PBS), plays multiple essential roles in virus reverse transcription. Retrovirus reverse transcription is primed by a host tRNA(Lys,3), which is incorporated into budding virions during assembly (Figure 1.4 A) (*16-18*). The tRNA anneals to the viral PBS via the 18 nucleotides (nts) at its 3' end (Figure 1.4, bent cyan arrow). As minus strand DNA is polymerized, the RNase H domain of RT cleaves RNA in RNA:DNA hybrids (*5*). Once RT reaches the 5' RNA cap (Figure 1.4 B), the newly synthesized DNA undergoes the first strand transfer (*19*). In this reaction, the R element of the minus strand DNA is annealed to the complementary R element at the 3' end of the genome (Figure 1.4 C). Post-strand transfer, minus strand synthesis commences once again along with concomitant digestion of the RNA template by RNase H (*5*). Stretches of the plus strand RNA that are rich in purines (called polypurine tracts) are RNase-resistant (Figure 1.4 D) and can serve as the primers for plus strand DNA synthesis (*20*). RT extends plus strand synthesis past the 5' end of the negative strand DNA, transcribing 18 nts of the host tRNA and generating the PBS in the

plus strand intermediate (Figure 1.4 E) (*5*, *21*, *22*). It is hypothesized that RT reverse transcribes the host tRNA until it encounters a modified base, which is a methyl-adenosine at position 58 of the tRNA, 19 bases from the 3' end (*22-24*). Subsequently RT performs the second strand transfer reaction (Figure 1.4 F) and completes synthesis a full length, 2-LTR DNA provirus is generated (Figure 1.4 G).

Ψ Stem Loops

The final set of RNA elements in the 5' leader are collectively referred to as the Ψ packaging stem loops and are involved in mRNA splicing and virion assembly. 3 Ψ stem loops fall within the 5' leader and have been named stem loop 1, 2, and 3 (*25*). Within the loop of Stem loop 1, HIV-1 encodes a palindromic 6-8 nt sequence called the dimerization initiation site (DIS). During virion assembly, two HIV-1 genomes base pair via the DIS and initiate the genomic dimerization reaction (*25-27*). Co-packaging of HIV-1 genomes is a requirement for virus particle infectivity (*28*) and can result in recombination if two genetically distinct viruses infect the same cell (*29*). Stem loop 2 encodes the first, or major, splice donor (SD1), which is used as the splice donor in every HIV-1 splicing reaction (*30*). Although this secondary structure has only been recently been directly detected using nuclear magnetic resonance methods (*31*), the proper folding of this motif is crucial to viral fitness (*32*). Collectively, these three stem loops (plus one that falls downstream of the *gag* AUG) play an important role in accurate genomic

packaging. When the Ψ signal is removed, cellular RNAs are nonspecifically packaged into budding virions (*33*).

HIV-1 translation

In addition to the conserved functions discussed above, the 5' leader plays an important role in regulation of virus translation. Because all HIV-1 RNAs are 5' 7-methylguanosine capped, it has been proposed that translation initiation should be mediated by typical, cap-dependent ribosomal scanning (*34*). When ligated to a reporter, however, regions of the HIV-1 5' leader inhibit ribosomal scanning and downstream translation initiation (*35*, *36*). Thus, many alternative mechanisms of HIV-1 translation regulation have been proposed (reviewed in (*37*)). Multiple groups have identified Internal Ribosomal Entry Sites (IRESs) in various regions of the HIV-1 genome (*38-40*), however assays used to characterize these elements are subject to false positive results (*37*). Thus, new methods of assessing the mechanism HIV-1 translation initiation are needed.

Additionally, a growing body of literature concerning eukaryotic and viral translation has suggested that "untranslated regions," like the HIV-1 5' leader, may actually be translated. While AUG codons are heavily disfavored in the 5' leader, one-off AUGs (e.g. CUG) have been shown to mediate translation initiation (*41*, *42*). Further, ribosomal profiling in the presence of lactimidomycin or harringtonine – two drugs that that lead to the accumulation of ribosomes at translation initiation sites – has supported

the notion that upstream, one-off AUGs initiate translation. For example, 96 of 506 mapped initiation sites in the human cytomegalovirus genome were one-off AUGs (*43*). These findings, while further removed from HIV-1 itself, shed light on the complexities of eukaryotic translation and may inform our ultimate understanding of HIV-1 translation regulation.

Section 1.3 –HIV-1 evolution and the T cell response

Acute HIV-1 infection

For 7-21 days following transmission, newly infected individuals generally show no signs or symptoms of HIV-1 infection (*44*). During this eclipse phase, the virus undergoes an initial local expansion, infecting CD4+ T cells in the genital mucosa, submucosa and draining lymphoid tissue (*45*). When local infection disseminates into the bloodstream, the first clinical markers of HIV-1 infection are detected (*44*). These include viral RNA (Fiebig stage I), Gag p24 antigen (Fiebig stage II), and, in some cases, a "flu-like" acute retroviral syndrome (*46*, *47*). Once viral RNA is detected in the blood, the virus population increases in size exponentially with a median doubling time of ~16 hours and a R_o of 8 (*48*). Infection spreads from the local site of infection to the gut associated lymphoid tissue (GALT), leading to a massive depletion of mucosal CD4+ T cells (*45*, *47*, *49*). During this viral ramp up, viral loads frequently reach >1 million copies/mL (*44*, *47*). Analysis of viral sequences taken at this time reveals random mutation with no evidence of positive selection. If these sequences demonstrate both a star-like phylogeny with coalescence to a single ancestral sequence and Poisson distribution of Hamming Distances (*50*, *51*), it can be inferred that infection was established by a single transmitted founder virus. In 2008, Keele and colleagues studied this process using single genome sequencing. Single genome amplification of end-point diluted plasma followed by direct amplicon sequencing (single genome sequencing or SGS) eliminates *Taq*-induced errors, *Taq*-mediated recombination, and cloning bias (*51-53*). Using this model, paired with single genome sequencing, it has been determined that

70-80% of sexually transmitted HIV-1 infections are established by a single transmitted founder (TF) virus (*47*, *50*, *54*).

Around 22 days post-infection, viral load peaks and begins to decline. This drop in viral load is accompanied by the appearance of HIV-1 specific, non-neutralizing antibodies (*44*, *55*) and the first detectable T cell response (*56*, *57*). As an HIV-infected person transitions into chronic infection, viral load reaches a quasi-stable set point (usually between 100-100,000 copies/mL) that is maintained for years (*46*).

Factors affecting within patient HIV-1 evolution

Longitudinal HIV-1 sequence analysis has revealed that the genetically homogeneous population present during acute infection accumulates remarkable genetic diversity over time (*46*, *58*, *59*). Two parameters contribute to this feature of the virus: the mutation rate and the turnover rate. While estimates of the substitution rate vary slightly based on method, it is generally agreed that the HIV-1 substitution rate ranges between 1.4-3.4 x 10^{-5} mutations/base pair/cycle (*5*, *51*, *60*, *61*). With a genome size of approximately 10 kilobases, it can be conservatively estimated that 1 in 10 progeny genomes will harbor a base substitution relative to the parental virus per generation. The next parameter, the virus turnover rate, has been estimated using data from monotherapy trials. Mathematical modeling has demonstrated that infected cells and virions have a short half-life of 1-2 days and <6 hours, respectively (*46*, *62-65*). Based on this clearance rate, ~10^{10} new virions must be produced each day in an infected individual (*65*). The

combined effect of the HIV-1 error and turnover rates is staggering – every substitution at every position of the viral genome is sampled hundreds to thousands of times per day (*64*).

Another hallmark of HIV-1 evolution within a host is the iterative fixation of immune escape mutations over time (*59*, *66*). Longitudinally collected sequences taken during chronic infection exhibit a time-ordered phylogenetic structure, characteristic of a pathogen evolving under positive selection (*66*, *67*). When single genome sequencing is used to generate these data, the molecular pathway of virus evolution can be ascertained (*51*, *52*, *68*). An important consequence of SGS is the maintenance of genetic linkage across the entire amplicon. The application of this tool is best illustrated through an example. Figure 1.5 depicts longitudinally collected single genome sequences from subject CH40 in a synonymous/nonsynonymous highlighter plot (*68*). In this plot, horizontal lines represent sequences generated using plasma from days 45-412 post-trial enrollment, with the genomic map shown at the bottom. Tick marks represent nucleotide changes relative to inferred transmitted founder virus, with green ticks representing synonymous changes, red ticks nonsynonymous, and blue ticks changes outside of viral ORFs. Concentrated non-random mutations away from the TF (positive selection) are observed in nearly all day 45 sequences in five regions throughout the viral genome. Figure 1.6 zooms in on such region in Nef that demonstrates a pattern of mutation called toggling. As shown in Figure 1.6, sequences from days 45-412 exhibited toggling of one of three nonsynonymous changes in Nef: S188N, R192H, and R196Q. These three mutations are embedded in a peptide – SLAFRHVAR – that was encoded by the

autologous TF virus. Multiple studies have shown that this pattern of mutation is consistent with escape from an autologous T cell response (*68*-*71*), which can be measured using interferon-γ ELISPOT assays (Figure 1.7). As subjects are followed for longer durations of time, more shared mutations away from the TF are observed – in CH40 at least 14 sites within the viral genome exhibit concentrated, non-random mutations at day 412. In this way, virus evolutionary patterns found in SGS datasets can be used to precisely characterize host-pathogen interactions and directly measure positive selection. Similar analysis has revealed that this technique can also be applied to study escape from the neutralizing antibody response (*53*, *72*-*74*).

In summary, acute and chronic HIV-1 infection are characterized by different evolutionary processes. During acute infection, the virus exhibits random diversification and exponential population growth (*50*). During chronic infection, the population demonstrates continuous, strong positive selection (*66*). As demonstrated by the Nef-SR9 example, mutations can serve as footprints of selection that provide insight into immune pressure *in vivo* (*68*, *71*).

However, certain selected mutations in longitudinal SGS datasets represet synonymous changes or fall outside of canonical HIV-1 ORFs. As shown in Figure 1.3, CH40 sequences taken at day 412 have shared synonymous mutations in viral ORFs (green tick marks, in *pol* and *env*, Figure 1.5) and shared mutations outside of canonical coding regions (blue tick marks). These selected mutations, while less commonly observed than their nonsynonymous counterparts, could potentially represent a number of processes. They may represent escape from immune responses directed at alternative

reading frame peptides (*75*, *76*); alterations to functional RNA elements and secondary structures (*7*); or genetic "hitchhikers" that are linked to neighboring CTL escape mutations (*77*). A subset of these selected changes fall in the 5' leader sequence (blue tick marks, Figure 1.3), which has the alternative designation of "5' untranslated region." In Chapters 2 and 3, I will investigate two such patterns of non-random mutation within longitudinally collected HIV-1 5' leader sequences. The first pattern is a set of non-random mutations, similar to T cell escape mutations observed within canonical HIV-1 epitopes like Nef-SR9 described above. The second pattern of mutation falls in a short 5' leader motif called the primer overextension sequence.

Section 1.4 – HIV-1 and the neutralizing antibody response

Following the initial T cell response, neutralizing antibodies (nAbs) against the Env glycoprotein rise. The fully folded mature Env protein is a trimer of non-covalently bound heterodimers (*78*, *79*). The protein is expressed as a gp160 polyprotein that is cleaved by the host protease furin into gp120 and gp41 subunits (*3*). gp120 is partitioned into constant regions (C1-5) and variable loops (V1-5), which harbor hypervariable regions that are not readily aligned when comparing strains (*3*). gp41 consists of an extracellular region, a transmembrane domain, and a long cytoplasmic tail. Like the cytotoxic T lymphocyte response, the neutralizing antibody (nAb) response potently drives selection for virus escape mutations (*53*, *72*, *74*, *80*).

In a subset of individuals, the autologous strain-specific antibody response matures into a broadly cross-reactive one (*81*, *82*). A cross-sectional analysis of plasma neutralization breadth from 205 individuals using a heterologous panel of 219 viruses revealed that neutralization breadth exists on a spectrum from narrowly strain-specific to broadly cross-reactive during chronic infection (*83*). Advances in B cell cloning and Ig locus amplification have permitted the rapid isolation and cloning of broadly neutralizing monoclonal antibodies (bnAbs) from subjects who exhibit broad plasma neutralization (*84*, *85*). These monoclonal bnAbs recapitulate the neutralization patterns observed in plasma and exhibit atypical properties such as: unusually long heavy chain complementarity determining region 3s (HCDR3s), high levels of somatic hypermutation (upwards of 30% mutated relative to the germline), and autoreactivity (*84*, *86*).

Monoclonal bnAbs bind at multiple sites of vulnerability on the Env trimer (Figure 1.8). The three most commonly targeted sites are (1) the high mannose patch at the base of V3, (2) the CD4 binding site (CD4bs), and (3) V2 at the apex of the trimer (*79*). V3 glycan bnAbs that target the high mannose patch often engage Env at different angles relative to the viral membrane (*87*). These antibodies all contact a short motif in the descending arm of V3 (^{324}GDIR^{327}) and an N-linked glycan at position 332 (*87*, *88*). CD4bs bnAbs come in multiple flavors: VRC01-like bnAbs are restricted in which V_H gene is used (1-2 and 1-46), while others converge on a common mode of Env-binding via their HCDR3s (*73*, *89*, *90*). V2 apex bnAbs typically have long HCDR3s that interact with strand C of the V2 loop and a set of adjacent glycans at N156 and N160 (*91*). The remaining bnAb classes target additional sites on the Env trimer, but may be elicited less frequently during natural HIV-1 infection (*79*, *84*). These include gp120-gp41 interface, membrane proximal external region, and fusion peptide-directed bnAbs (*84*, *92*).

bnAb/virus co-evolution

It has been proposed that molecular characterization of virus/antibody co-evolution during bnAb development may inform HIV vaccine design (*92*, *93*). Through genetic and phenotypic characterization of both B cell and virus evolution, these types of studies identify crucial steps along the path of bnAb development and the viral antigens that drove this maturation. Here, I will introduce the first and most well characterized individuals studied with this technique, CH0505 (*72-74*). CH0505 is an African man who

was infected with a single Subtype C virus and sampled regularly over more than six years of infection, prior to initiation of antiretroviral therapy (ART, Haynes, B., unpublished). An CD4 binding site bnAb lineage designated CH103 was isolated from CH0505 and extensively characterized for its autologous and heterologous binding and neutralization properties (*72*). Interestingly, the unmutated common ancestor (UCA) of the CH103 lineage bound, but did not neutralize, the autologous transmitted founder (TF) virus (*72*). Evaluation of virus sequence evolution during the first 4 weeks of infection revealed that a small subset of Envs harbored changes in Loop D of the CD4 binding site. Interestingly, the sampled Loop D mutations sensitized autologous Envs to neutralization by the CH103 lineage (*72*, *74*). A second, strain-specific CD4bs antibody, called CH235, was then isolated from CH0505. Loop D variants were neutralization resistant when assayed with CH235 antibodies, suggesting that these mutations were driven by immune pressure from CH235 nAbs *in vivo*. In other words, the CH235 antibody lineage cooperated with the CH103 lineage, sensitizing Envs to CH103 binding and neutralization (*74*). Env sequences subsequently showed significantly increased diversification within CD4 binding site contact residues, prior to the development of heterologous neutralization breadth when compared to subject who did not develop bnAbs (*72*). A recent follow up study revealed that the CH235 lineage, which was initially a strain-specific nAb, underwent continued affinity maturation for five years of infection and developed into an even broader bnAb than CH103 (*73*). Thus, in CH0505, the development of bnAb lineages involved bnAb precursor engagement of the TF Env,

cooperation between multiple nAb lineages to maintain Env neutralization sensitivity, diversification in bnAb contacts, and extended affinity maturation over many years.

Antibody/virus coevolution in other subjects has supported the findings in CH0505. For example, human subjects CAP177 and CAP331 both developed broadly cross-reactive neutralizing serum that was dependent on the glycan at 332 for neutralization (V3 glycan bnAbs). The transmitted founder viruses from both of these individuals lacked this glycan, raising the question of what stimulated these N332-glycan dependent bnAbs (*94*). Likely as a result of autologous immune escape from another nAb lineage, the plasma virus acquired the glycan at 332 ~6 months into infection and heterologous neutralization breadth subsequently developed. While monoclonal antibodies were not isolated from these individuals, this pattern in plasma neutralization is consistent with the idea of cooperating nAb lineages. In fact, findings consistent with cooperating antibody lineages were seen in CAP256 (*95*, *96*), PC76 (*97*), and CAP257 (*98*). Together, these data suggest that, perhaps unsurprisingly, bnAb lineages do not develop in isolation during natural HIV-1 infection.

Unfortunately, comprehensive Env sequencing is not available for all of these subjects for direct comparison to virus evolution in CH0505. In one subject, CAP256, deep sequencing of V2 (the target of the bnAb response) revealed that, like CH0505, the plasma quasispecies exhibited diversification prior to neutralization breadth development (*95*). Future studies in more subjects will be needed to explore whether similar patterns of Env diversification precede breadth. In Chapter 4, I will present one such study, the

virus-antibody co-evolution in a human subject, CH0848, who developed V3-glycan specific bnAbs.

Section 1.5 – Monoclonal bnAbs in prevention and therapy

The recent explosion in the number of cloned monoclonal bnAbs has led to the development of exciting prevention and therapeutic modalities. Non-human primate studies have demonstrated that passive immunization of rhesus macaques with monoclonal bnAbs prevents simian/human immunodeficiency virus (SHIV) infection (*99*). These and other findings have led to the hypothesis that bnAb expression at the time of transmission may prevent infection in humans (*92*). To test this hypothesis, ~4,200 HIV-negative adults will be infused with a CD4bs bnAb called VRC01 and monitored for HIV-1 acquisition. This study's findings will be important for the HIV vaccine field, no matter the outcome.

bnAbs have also been tested as possible therapeutics. Monoclonal antibodies may provide benefit over small molecule inhibitors based on their ability to engage the host innate immune response (*100*). Initial studies in humanized mice demonstrated that treatment with 3-5 bnAbs were required to completely suppress viremia (*101*). This finding was followed up by two non-human primate trials in which suppression of simian/human immunodeficiency viruses (SHIV) replication in macaques was achieved with one or a combination of bnAbs (*102*, *103*). Given the success of these non-human studies, monoclonal bnAb infusions have now been tested in humans in two settings: viremic individuals (*104*, *105*) and antiretroviral therapy (ART)-suppressed individuals (*106*). In these trials, bnAb treatment was safe and, in some cases, may have offered therapeutic benefit in the form of a delay in rebound viremia and a restriction in the number of rebounding viruses (*106*). In all cases, however, bnAb treatment led to the

emergence of resistant viral strains. Future studies will be needed to test if combination bnAb treatment can fully suppress viremia and prevent the development of resistance. Further, comparisons must be made to small molecule inhibitors to determine whether bnAb infusion has added benefit over combination ART. In Chapter 5, I will present a study that monitored bnAb and virus evolution after bnAb immunotherapy.

Section 1.6 – References

1. B. S. Taylor, S. M. Hammer, The challenge of HIV-1 subtype diversity. *N Engl J Med*. **359**, 1965–1966 (2008).

2. D. Klatzmann *et al.*, Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes. *Science*. **225**, 59–63 (1984).

3. C. B. Wilen, J. C. Tilton, R. W. Doms, HIV: cell binding and entry. *Cold Spring Harb Perspect Med*. **2**, a006866–a006866 (2012).

4. R. Craigie, F. D. Bushman, HIV DNA Integration. *Cold Spring Harb Perspect Med*. **2**, –a006890 (2012).

5. W.-S. Hu, S. H. Hughes, HIV-1 Reverse Transcription. *Cold Spring Harb Perspect Med*. **2**, –a006882 (2012).

6. K. Lu *et al.*, NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*. **334**, 242–245 (2011).

7. E. Pollom *et al.*, Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. *PLoS Pathog.* **9**, e1003294 (2013).

8. L. A. Pereira, K. Bentley, A. Peeters, M. J. Churchill, N. J. Deacon, A compilation of cellular transcription factor interactions with the HIV-1 LTR

promoter. *Nucleic Acids Res.* **28**, 663–668 (2000).

9. J. KARN, C. M. Stoltzfus, Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med*. **2**, a006916 (2012).

10. C. DINGWALL *et al.*, Hiv-1 Tat Protein Stimulates Transcription by Binding to a U-Rich Bulge in the Stem of the Tar Rna Structure. *Embo J*. **9**, 4145–4153 (1990).

11. B. Berkhout, R. H. SILVERMAN, K. T. Jeang, Tat Trans-Activates the Human Immunodeficiency Virus Through a Nascent Rna Target. *Cell*. **59**, 273–282 (1989).

12. M. J. SELBY, E. S. BAIN, P. A. LUCIW, B. M. PETERLIN, Structure, Sequence, and Position of the Stem Loop in Tar Determine Transcriptional Elongation by Tat Through the Hiv-1 Long Terminal Repeat. *Genes & Development*. **3**, 547–558 (1989).

13. M. P. ASHE, P. GRIFFIN, W. JAMES, N. J. PROUDFOOT, Poly(a) Site Selection in the Hiv-1 Provirus - Inhibition of Promoter-Proximal Polyadenylation by the Downstream Major Splice Donor Site. *Genes & Development*. **9**, 3008–3025 (1995).

14. M. P. ASHE, L. H. Pearson, N. J. PROUDFOOT, The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *Embo J*. **16**, 5752–5763 (1997).

15. A. T. Das, B. Klaver, B. Berkhout, A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. *Journal of Virology*. **73**, 81–91 (1999).

16. A. T. Das, S. E. Koken, B. B. Essink, J. L. van Wamel, B. Berkhout, Human immunodeficiency virus uses tRNA(Lys,3) as primer for reverse transcription in HeLa-CD4+ cells. *FEBS Lett.* **341**, 49–53 (1994).

17. M. Jiang *et al.*, Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. *Journal of Virology*. **67**, 3246–3253 (1993).

18. L. Kleiman, tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life*. **53**, 107–114 (2002).

19. A. T. Panganiban, D. Fiore, Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science*. **241**, 1064–1069 (1988).

20. P. Charneau, M. Alizon, F. Clavel, A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *Journal of Virology*. **66**, 2814–2820 (1992).

21. R. Muthuswami *et al.*, The HIV plus-strand transfer reaction: determination of replication-competent intermediates and identification of a novel lentiviral element, the primer over-extension sequence. *Journal of Molecular Biology*. **315**, 311–323 (2002).

22. S. Auxilien, Role of Post-transcriptional Modifications of Primer tRNALys,3 in the Fidelity and Efficacy of Plus Strand DNA Transfer during HIV-1 Reverse Transcription. *Journal of Biological Chemistry*. **274**, 4412–4420 (1999).

23. H. Ben-Artzi *et al.*, Molecular analysis of the second template switch during reverse transcription of the HIV RNA template. *Biochemistry*. **35**, 10549–10557 (1996).

24. B. P. Burnett, C. S. McHenry, Posttranscriptional modification of retroviral primers is required for late stages of DNA replication. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7210–7215 (1997).

25. W. I. Sundquist, H.-G. Kraeusslich, HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb Perspect Med*. **2**, –a006924 (2012).

26. M. D. Moore *et al.*, Dimer initiation signal of human immunodeficiency virus type 1: its role in partner selection during RNA copackaging and its effects on recombination. *Journal of Virology*. **81**, 4002–4011 (2007).

27. S. F. Johnson, A. Telesnitsky, Retroviral RNA dimerization and packaging: the what, how, when, where, and why. *PLoS Pathog.* **6**, e1001007 (2010).

28. M. D. Moore, W.-S. Hu, HIV-1 RNA dimerization: It takes two to tango. *AIDS Rev*. **11**, 91–102 (2009).

29. W. S. Hu, H. M. TEMIN, Genetic consequences of packaging two RNA

genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1556–1560 (1990).

30. K. E. Ocwieja *et al.*, Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* **40**, 10345–10355 (2012).

31. S. C. Keane *et al.*, RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*. **348**, 917–921 (2015).

32. T. E. M. Abbink, B. Berkhout, RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *Journal of Virology*. **82**, 3090–3098 (2008).

33. S. J. Rulli *et al.*, Selective and nonselective packaging of cellular RNAs in retrovirus particles. *Journal of Virology*. **81**, 6623–6631 (2007).

34. A. G. Hinnebusch, eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem. Sci.* **31**, 553–562 (2006).

35. A. P. Geballe, M. K. Gray, Variable inhibition of cell-free translation by HIV-1 transcript leader sequences. *Nucleic Acids Res.* **20**, 4291–7. PMC334138. (1992).

36. N. T. Parkin *et al.*, Mutational analysis of the 5' non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation. *Embo J.* **7**, 2831–7. PMC457075. (1988).

37. C. Bolinger, K. Boris-Lawrie, Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. *Retrovirology*. **6**, 8 (2009).

38. C. B. Buck *et al.*, The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. *Journal of Virology*. **75**, 181–191 (2001).

39. T.-D. M. Plank, J. T. Whitehurst, J. S. Kieft, Cell type specificity and structural determinants of IRES activity from the 5' leaders of different HIV-1 transcripts. *Nucleic Acids Res.* **41**, 6698–6714 (2013).

40. A. Brasey *et al.*, The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *Journal of Virology*. **77**, 3939–3949 (2003).

41. S. R. Starck *et al.*, Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science*. **336**, 1719–1723 (2012).

42. A. Corbin, A. C. Prats, J. L. Darlix, M. Sitbon, A nonstructural gag-encoded glycoprotein precursor is necessary for efficient spreading and pathogenesis of murine leukemia viruses. *Journal of Virology*. **68**, 3857–3867 (1994).

43. N. Stern-Ginossar *et al.*, Decoding human cytomegalovirus. *Science*. **338**, 1088–1093 (2012).

44. E. W. Fiebig *et al.*, Dynamics of HIV viremia and antibody seroconversion in

plasma donors: implications for diagnosis and staging of primary HIV infection. *AIDS*. **17**, 1871–1879 (2003).

45. A. T. Haase, Targeting early infection to prevent HIV-1 mucosal transmission. *Nature*. **464**, 217–223 (2010).

46. J. Coffin, R. Swanstrom, HIV Pathogenesis: Dynamics and Genetics of Viral Populations and Infected Cells. *Cold Spring Harb Perspect Med*. **3**, –a012526 (2013).

47. G. M. Shaw, E. Hunter, HIV transmission. *Cold Spring Harb Perspect Med*. **2**, a006965–a006965 (2012).

48. R. M. Ribeiro *et al.*, Estimation of the initial viral growth rate and basic reproductive number during acute HIV-1 infection. *Journal of Virology*. **84**, 6096–6102 (2010).

49. R. S. Veazey *et al.*, Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection. *Science*. **280**, 427–431 (1998).

50. B. F. Keele *et al.*, Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7552–7557 (2008).

51. J. F. Salazar-Gonzalez *et al.*, Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification

and sequencing. *Journal of Virology*. **82**, 3952–3970 (2008).

52. S. Palmer *et al.*, Multiple, Linked Human Immunodeficiency Virus Type 1 Drug Resistance Mutations in Treatment-Experienced Patients Are Missed by Standard Genotype Analysis. *Journal of Clinical Microbiology*. **43**, 406–413 (2005).

53. K. J. Bar *et al.*, Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. *PLoS Pathog.* **8**, e1002721 (2012).

54. D. C. Tully *et al.*, Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLoS Pathog.* **12**, e1005619 (2016).

55. G. D. Tomaras *et al.*, Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *Journal of Virology*. **82**, 12449–12463 (2008).

56. P. Borrow, H. Lewicki, B. H. Hahn, G. M. Shaw, M. B. Oldstone, Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *Journal of Virology*. **68**, 6103–6110 (1994).

57. R. A. Koup *et al.*, Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *Journal of Virology*. **68**, 4650–4655 (1994).

58. B. Korber *et al.*, Evolutionary and immunological implications of contemporary HIV-1 variation. *British Medical Bulletin*. **58**, 19–42 (2001).

59. R. Shankarappa *et al.*, Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *Journal of Virology*. **73**, 10489–10502 (1999).

60. L. M. MANSKY, H. M. TEMIN, Lower in-Vivo Mutation-Rate of Human-Immunodeficiency-Virus Type-1 Than That Predicted From the Fidelity of Purified Reverse-Transcriptase. *Journal of Virology*. **69**, 5087–5094 (1995).

61. M. E. Abram, A. L. Ferris, W. Shao, W. G. Alvord, S. H. Hughes, Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. *Journal of Virology*. **84**, 9864–9878 (2010).

62. D. D. Ho *et al.*, Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature*. **373**, 123–126 (1995).

63. X. Wei *et al.*, Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*. **373**, 117–122 (1995).

64. J. M. Coffin, HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*. **267**, 483–489 (1995).

65. A. S. Perelson, A. U. Neumann, M. Markowitz, J. M. Leonard, D. D. Ho, HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral

generation time. *Science*. **271**, 1582–1586 (1996).

66. A. Rambaut, D. Posada, K. A. Crandall, E. C. Holmes, The causes and consequences of HIV evolution. *Nat Rev Genet*. **5**, 52–61 (2004).

67. B. T. Grenfell *et al.*, Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*. **303**, 327–332 (2004).

68. J. F. Salazar-Gonzalez *et al.*, Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* **206**, 1273–1289 (2009).

69. M. Kearney *et al.*, Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. *Journal of Virology*. **83**, 2715–27. PMC2648286. (2009).

70. B. Walker, A. McMichael, The T-cell response to HIV. *Cold Spring Harb Perspect Med*. **2** (2012), doi:10.1101/cshperspect.a007054.

71. N. Goonetilleke *et al.*, The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* **206**, 1253–1272 (2009).

72. H.-X. Liao *et al.*, Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*. **496**, 469–476 (2013).

73. M. Bonsignori *et al.*, Maturation Pathway from Germline to Broad HIV-1

Neutralizer of a CD4-Mimic Antibody. *Cell*. **165**, 449–463 (2016).

74. F. Gao *et al.*, Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell*. **158**, 481–491 (2014).

75. S. M. Mayrand, D. A. Schwarz, W. R. Green, An alternative translational reading frame encodes an immunodominant retroviral CTL determinant expressed by an immunodeficiency-causing retrovirus. *J. Immunol.* **160**, 39–50 (1998).

76. A. Bansal *et al.*, CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med.* **207**, 51–59 (2010).

77. P. S. Pennings, S. Kryazhimskiy, J. Wakeley, Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genet.* **10**, e1004000 (2014).

78. P. D. Kwong *et al.*, HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature*. **420**, 678–682 (2002).

79. M. Pancera *et al.*, Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*. **514**, 455–461 (2014).

80. X. Wei *et al.*, Antibody neutralization and escape by HIV-1. *Nature*. **422**, 307–312 (2003).

81. E. Landais *et al.*, Broadly Neutralizing Antibody Responses in a Large Longitudinal Sub-Saharan HIV Primary Infection Cohort. *PLoS Pathog.* **12**, e1005369 (2016).

82. P. Rusert *et al.*, Determinants of HIV-1 broadly neutralizing antibody induction. *Nature Medicine* (2016), doi:10.1038/nm.4187.

83. P. Hraber *et al.*, Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS*. **28**, 163–169 (2014).

84. P. D. Kwong, J. R. Mascola, Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity*. **37**, 412–425 (2012).

85. J. F. Scheid *et al.*, Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. **333**, 1633–1637 (2011).

86. B. F. Haynes *et al.*, Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science*. **308**, 1906–1908 (2005).

87. D. Sok *et al.*, Promiscuous glycan site recognition by antibodies to the high-mannose patch of gp120 broadens neutralization of HIV. *Sci Transl Med*. **6**, 236ra63–236ra63 (2014).

88. D. Sok *et al.*, A Prominent Site of Antibody Vulnerability on HIV Envelope Incorporates a Motif Associated with CCR5 Binding and Its Camouflaging Glycans. *Immunity*. **45**, 31–45 (2016).

89. X. Wu *et al.*, Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-1 Infection. *Cell*. **161**, 470–485 (2015).

90. T. Zhou *et al.*, Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting

the CD4 Supersite in 14 Donors. *Cell*. **161**, 1280–1292 (2015).

91. J. S. McLellan *et al.*, Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*. **480**, 336–343 (2011).

92. B. F. Haynes *et al.*, HIV-Host Interactions: Implications for Vaccine Design. *Cell Host and Microbe*. **19**, 292–303 (2016).

93. J. R. Mascola, B. F. Haynes, HIV-1 neutralizing antibodies: understanding nature's pathways. *Immunol. Rev.* **254**, 225–244 (2013).

94. P. L. Moore *et al.*, Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. *Nature Medicine*. **18**, 1688–1692 (2012).

95. N. A. Doria-Rose *et al.*, Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*. **509**, 55–62 (2014).

96. J. N. Bhiman *et al.*, Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nature Medicine*. **21**, 1332–1336 (2015).

97. D. T. MacLeod *et al.*, Early Antibody Lineage Diversification and Independent Limb Maturation Lead to Broad HIV-1 Neutralization Targeting the Env High-Mannose Patch. *Immunity*. **44**, 1215–1226 (2016).

98. C. K. Wibmer *et al.*, Viral escape from HIV-1 neutralizing antibodies drives

increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. *PLoS Pathog.* **9**, e1003738 (2013).

99. R. Gautam *et al.*, A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges. *Nature*. **533**, 105–109 (2016).

100. S. Bournazos, J. V. Ravetch, Fcγ receptor pathways during active and passive immunization. *Immunol. Rev.* **268**, 88–103 (2015).

101. F. Klein *et al.*, HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature*. **492**, 118–122 (2012).

102. M. Shingai *et al.*, Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. *Nature*. **503**, 277–280 (2013).

103. D. H. Barouch *et al.*, Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature*. **503**, 224–228 (2013).

104. M. Caskey *et al.*, Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature*. **522**, 487–491 (2015).

105. R. M. Lynch *et al.*, Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Sci Transl Med*. **7**, 319ra206–319ra206 (2015).

106. J. F. Scheid *et al.*, HIV-1 antibody 3BNC117 suppresses viral rebound in humans

during treatment interruption. *Nature*. **535**, 556–560 (2016).

107. Landmarks of the HIV Genome. *LANL.gov*, (available at http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html).

108. *LANL.gov*.

109. M. H. Malim, M. Emerman, HIV-1 accessory proteins--ensuring viral survival in a hostile environment. *Cell Host and Microbe*. **3**, 388–398 (2008).

Section 1.7 – Figures

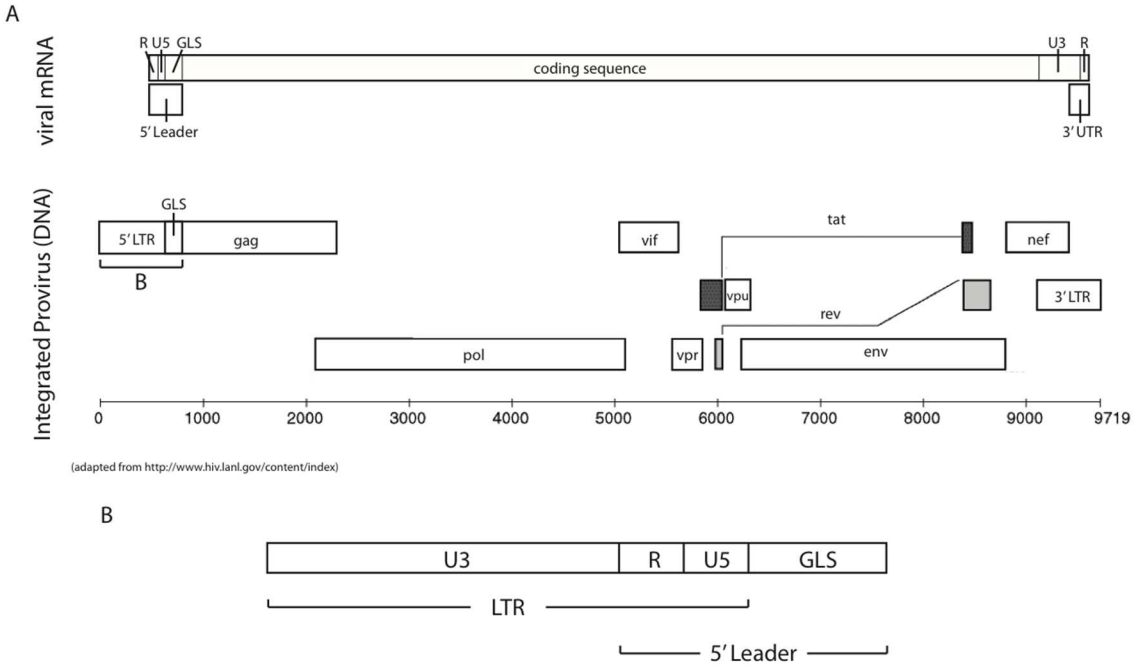


Figure 1.0-1 HIV-1 RNA and DNA Genomic Organization

(A) During the HIV-1 lifecycle, the viral genome is converted between a single stranded, positive-sense RNA molecule and a double stranded DNA molecule (top and bottom, respectively). All transcripts begin with the 5' leader sequence. It harbors the retroviral repeat (R) element, the unique 5 (U5) element, and the *gag* leader sequence (GLS). Following are the nine canonical HIV-1 open reading frames (ORFs): *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef*. The viral group specific antigen (*gag*) encodes structural proteins including matrix, capsid, and nucleocapsid (*3*, *25*, *107*). The *pol* gene encodes the enzymes protease, reverse transcriptase, and integrase (*4*, *5*, *25*). *rev* and *tat* encode the two essential viral regulatory factors that mediate viral RNA nuclear export and transactivation, respectively (*9*, *108*). *env* encodes the Envelope glycoprotein, which mediates viral entry (*3*). The remainder of the viral ORFs encode so-called accessory proteins: Nef, Vif, Vpu, and Vpr (*109*). These proteins perform functions that are sometimes dispensable for replication *in vitro*, but affect pathogenesis *in vivo* (*109*). Finally, the RNA genome ends with a retroviral unique 3 element (U3) and the 3' R element. (B) Upon reverse transcription into a DNA provirus, the HIV-1 genome is flanked by two complete long terminal repeats (LTRs), each consisting of U3, R, and U5 elements. The spans of these elements are depicted. Abbreviations: R – repeat element, U5 – unique 5 element, GLS – gag leader sequence, U3 – unique 3 element, 3' UTR – 3' untranslated region, LTR – long terminal repeat. Adapted from www.hiv.lanl.gov.

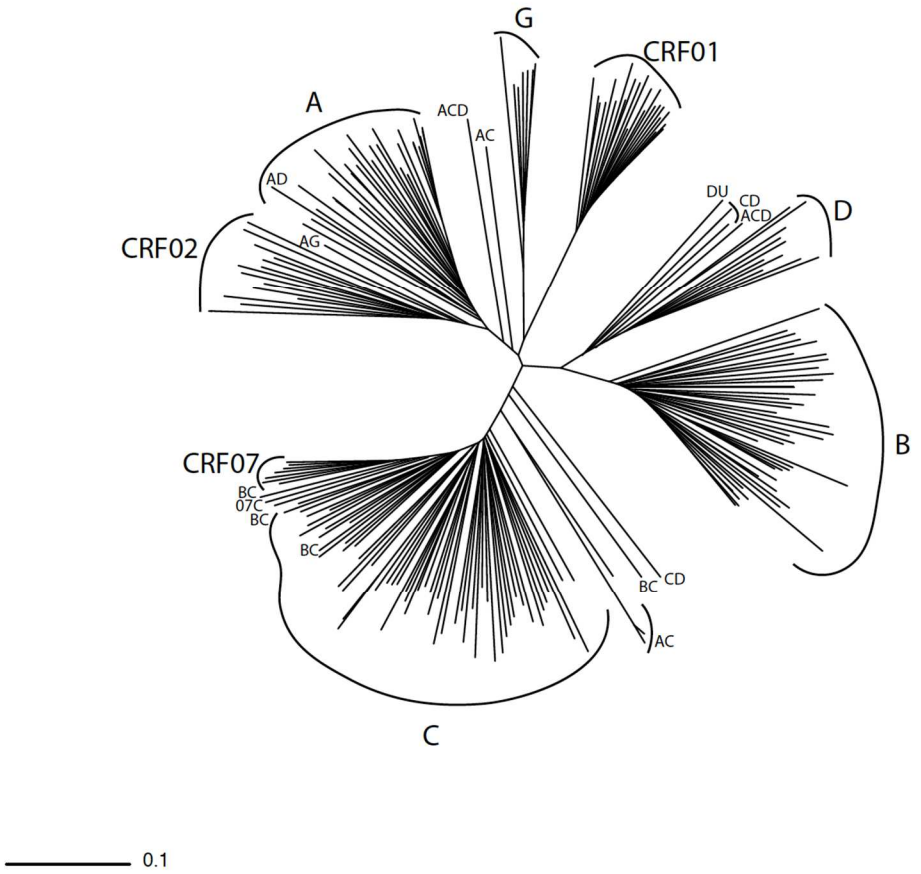


Figure 1.0-2 HIV-1 Group M *env* diversity.

HIV-1 strains can be categorized into four lineages, which are referred to as Groups M, N, O, and P. Group M strains, which cause the vast majority of HIV-1 infections, have been further classified as one of nine subtypes (A-D, F-H, J, K) or recombinants thereof. Depicted is a maximum likelihood *env* phylogenetic tree with 208 representative sequences from major Group M subtypes and circulating recombinant forms (CRFs). Adapted from Bonsignori, et al., submitted.

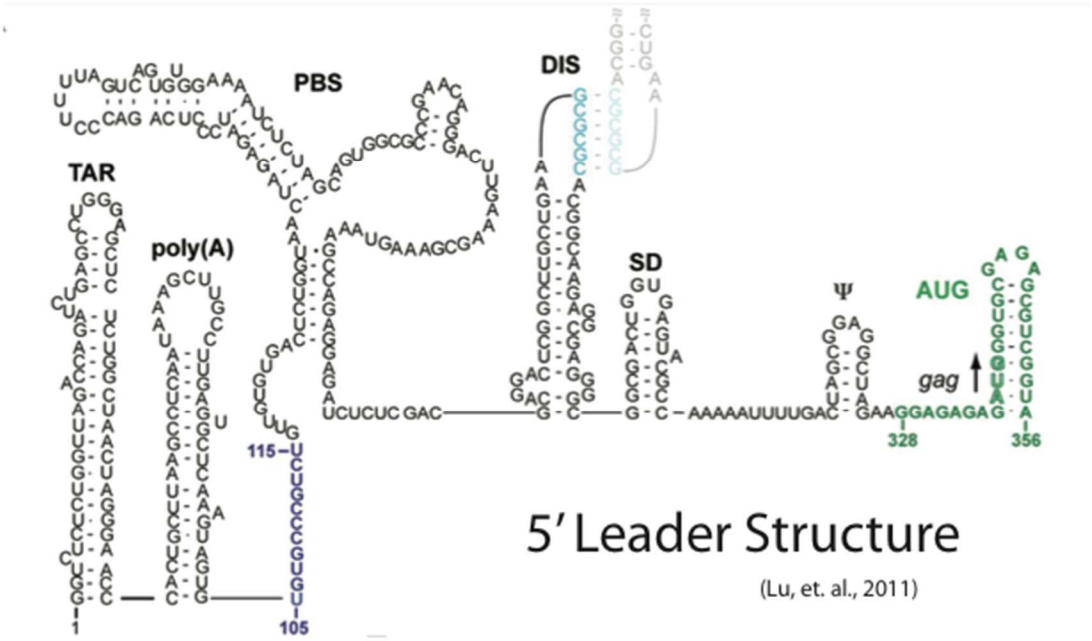


Figure 1.0-3 Secondary structure of the 5' leader sequence.

Depicted is a graphical representation of the RNA secondary structure of the 5' leader. Adapted from Lu, et al., 2011. Abbreviations – transactivation response element (TAR), poly-adenylation signal (polyA), primer binding site (PBS), dimerization initiation site (DIS), major splice donor (SD), psi packaking signal stem loop 3 (Ψ).

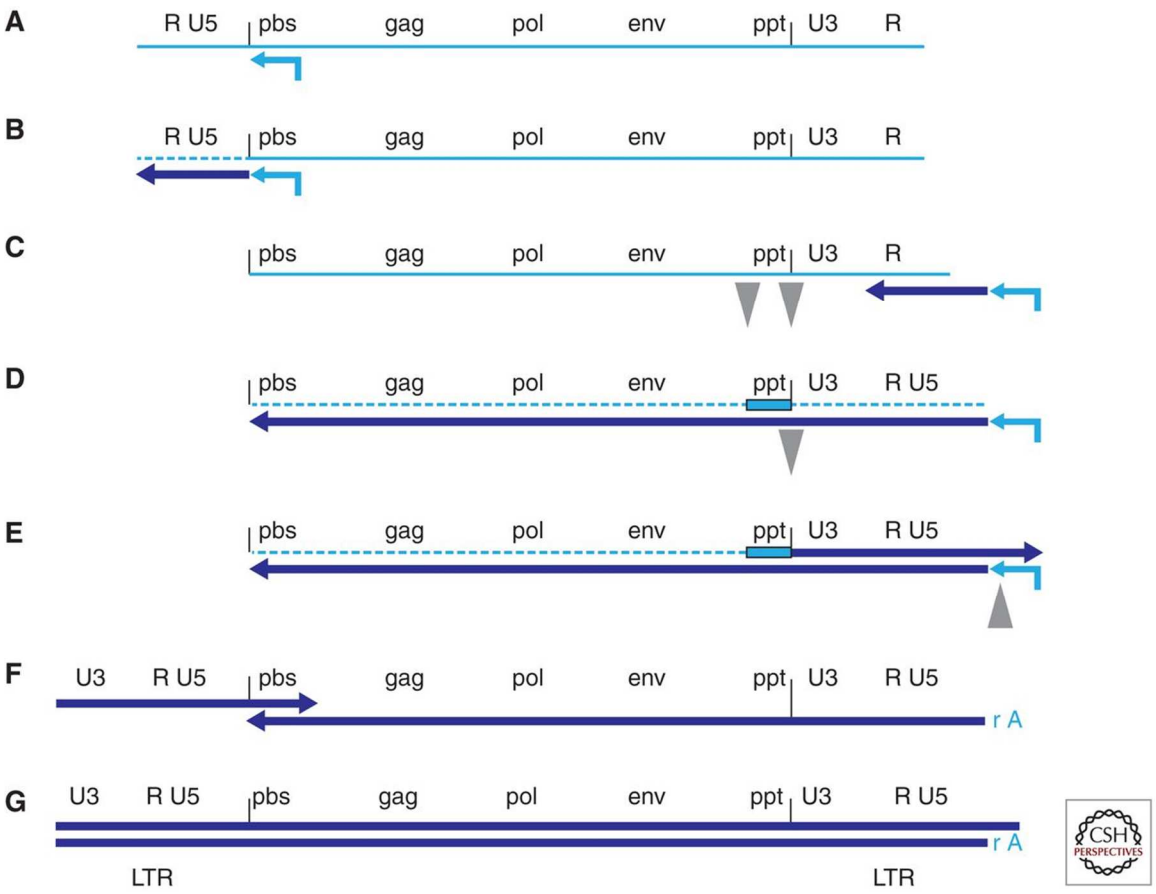


Figure 1.0-4 Reverse trascription

As described in the text, depicted are the steps of reverse transcription. tRNA(Lys,3) is represented as a bent cyan arrow, plus strand RNA in cyan (digested RNA dotted), and DNA in dark blue. (A) Priming from tRNA(Lys,3). (B) reverse transcription of U5 and R and digestion of viral RNA in RNA:DNA duplex until the minus strand strong stop. (B) to (C) First strand transfer. (D) Continued minus strand synthesis and RNA digestion (except for the polypurine tracts, ppt). (E) plus strand synthesis using the ppt as a primer until plus strand strong stop, typically after 18 nts of the tRNA are copied. (E) to (F) Second strand transfer. (G) full extension of both plus and minus strands. Adapted from Hu and Hughes, 2011.

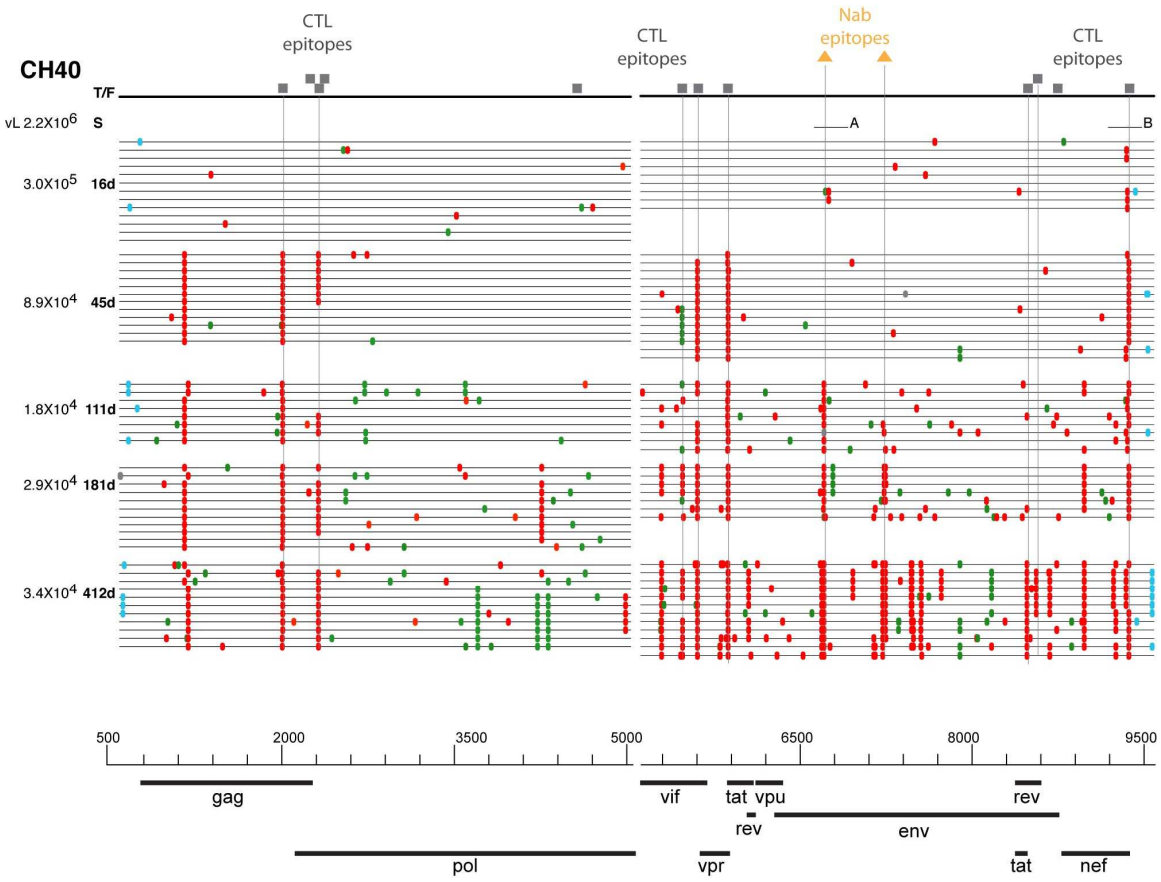


Figure 1.0-5 HIV01 sequence evolution in CH40

CH40 is a human subject who was identified during acute infection and sampled longitudinally. Depicted is a nucleotide highlighter plot in which horizontal lines represent single genome sequences derived at screening (S), day 6, day 45, day 11, day 181, and day 412 post-screening. Tick marks denote sequence changes relative to the transmitted founder (TF) virus. Red ticks are nonsynonymous, green are synonymous, and blue fall outside of known ORFs. Selected mutations are embedded in cytotoxic T lymphocyte (CTL) epitopes (grey squares) and neutralizing antibody (Nab) epitopes (orange triangles) and confer escape from these immune responses in *ex vivo* assays. Viral load (vL) is indicated to the left. Adapted from Salazar-Gonzalez, et al., 2009, Goonetilleke, et al., 2009, and Bar, et al., 2012.

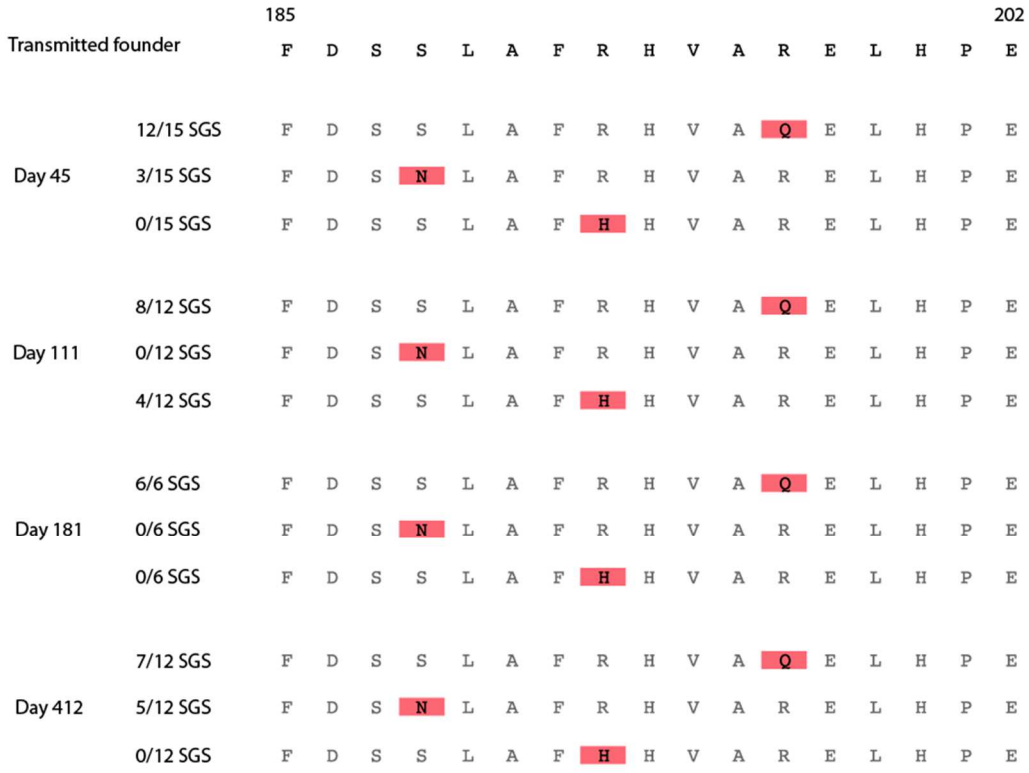


Figure 1.0-6 Amino acid toggling in Nef.

Depicted is an amino acid alignment of Nef positions 185-202 from CH40 (far right, figure 1.3). The inferred transmitted founder sequence is depicted at the top. Mutations relative to this sequence are highlighted in red. 15, 12, 6, and 12 single genome sequences were generated at days 45, 111, 181, and 412, respectively. The number of SGS matching each amino acid variant is indicated to the left. 100% of sequences from day 45 onward harbored a nonsynonymous mutation relative to the TF within a 9 amino acid span, SLAFRHVAR. Sequences that did not encode a full length, gross non-defective *nef* reading frame were excluded.

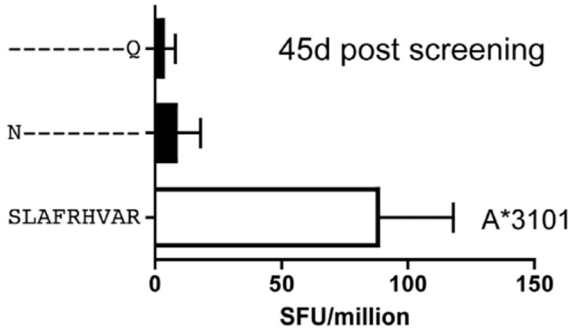


Figure 1.0-7 CH40 immune recognuition of Nef-SR9.

CH40 PBMCs taken at day 45 were stimulated with the autologous TF Nef peptide SLAFRHVAR or one of two Nef peptide variants (**N**LAFRHVAR and SLAFRHVA**Q**, see Figure 1.4) and assayed with an interferon-γ ELISPOT. Stimulation with the TF peptide resulted in more than 75 spot forming units/million PBMCs (SFU/million). Stimulation with Nef peptide variants that were selected observed day 45 and after yielded fewer than 20 SFU/million PBMCs. Adapted from Goonetilleke, et al., 2009.

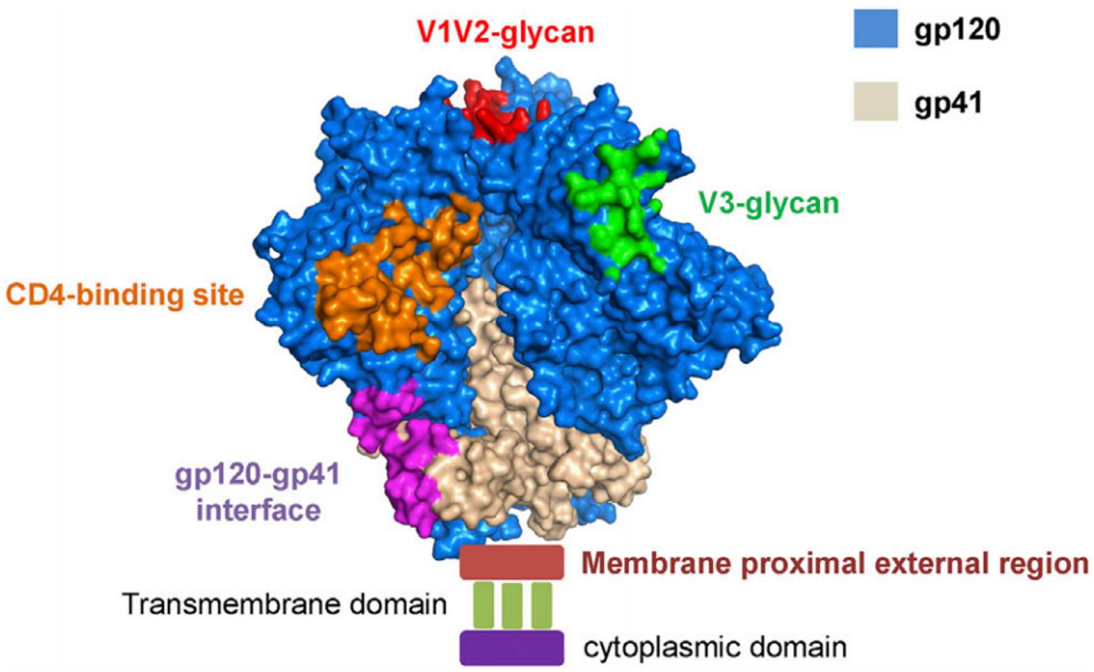


Figure 1.0-8 Broadly neutralizing antibody sites of vulnerability on the Env trimer.

Adapted from Haynes, et al., 2016

CHAPTER 2

NOVEL T CELL ANTIGENS ENCODED BY THE HIV-1 5' LEADER SEQUENCE

Edward F. Kreider[1,2], Hui Li[1,2], Ranjit Warrier[1,2], Katja Pfafferott[3], Thomas Partridge[3], Nicola Ternette[3], Michael Lopker[1,2], Yingying Li[1,2], Gerald H. Learn[1,2], Katherine Roberts[3], Benedikt Kessler[3], Andrew J. McMichael[3], Beatrice H. Hahn[1,2], Persephone Borrow[3], George M. Shaw[1,2]

Affiliations:

[1]Departments of Medicine and [2]Microbiology, Perelman School of Medicine at the University of Pennsylvania

[3]Nuffield Department of Medicine, Oxford University

Contributions: This study was initially conceived following work on SIVmac239/766 performed by ML, HL, and GMS. I designed, executed, and analyzed sequencing and immunologic experiments with technical assistance and insight from HL, KP, KR, RW, GHL, and YL and under the supervision of AJM, BHH, PB, and GMS. Mass spectrometry experiments were performed by TP, NT under the supervision of BK, AJM, and PB. I wrote this chapter with assistance from GMS. These data will be submitted as part of a larger manuscript in Fall 2016.

All figures appear at the end of this chapter

Section 2.1 – Abstract

Introduction: The lentiviral 5' leader is a conserved RNA structure that lies upstream of canonical coding regions and is conventionally considered "untranslated." Nonetheless, we recently observed concentrated, non-random mutations within the 5' leader of SIVmac strains sequenced longitudinally from Mamu-B*29+ macaques. Follow up studies revealed that this sequence evolution represented virus escape from a Mamu-B*29-restricted cytotoxic T lymphocyte (CTL) response targeting a 5' leader-encoded peptide called KA9 (KGAGRYQTA). We hypothesized that the HIV-1 5' leader also expressed peptides that were recognized by host T cells during infection.

Methods: Longitudinal plasma viral single genome sequences were generated from 27 human subjects who were acutely infected with HIV-1. *Ex vivo* interferon-γ ELISPOT assays were performed on human PBMCs stimulated with autologous transmitted founder virus peptides or putative escape variants encoded by the viral 5' leader and proteome. Mass spectrometry analysis of major histocompatibility complex (MHC)-presented epitopes was conducted on peptides purified using MHC immunoprecipitation, acid elution, and reverse phase liquid chromatography of lysates from HIV-infected CD4+ T cells.

Results: Longitudinal 5' leader sequencing from 13 out of 27 HIV-1-infected human subjects revealed mutational patterns consistent with escape from CTLs. Interferon-γ ELISPOT assays using PBMCs from three of these subjects (infected with subtype A or C HIV-1) demonstrated autologous T cell recognition of five HIV-1 5' leader peptides (5'L-ER10, 5'L-LL9, 5'L-IA9, 5'L-RS9, and 5'L-AA10). These peptides were encoded

by reading frames +1 and +3 within R or around Stem Loop 1. Mutations selected *in vivo* either conferred nonsynonymous changes that abrogated host CTL recognition or altered the 5' leader reading frame by through start codon deletion, nonsense mutation, or frameshifts. Morevoer, these putative escape mutations spared known functional RNA stems and loops and were predicted to not alter stem base-pairing. A sixth 5' leader peptide encoded by reading frame +2 within the R element of a subtype B virus was detected using mass spectrometry.

Conclusions: These data collectively demonstrate that 5' leader sequences from subtype A, B, and C HIV-1s express peptides from all three forward reading frames. During natural infection, these peptides elicit robust T cell responses that potently drive virus escape. These findings provide new insight into mechanisms of viral translation and will inform future vaccine efforts that target this region of the genome.

Section 2.2 – Introduction

The lentiviral 5' leader sequence is polyfunctional RNA structure that lies upstream of the canonical coding regions in every viral transcript. It spans from the first nucleotide of the long terminal repeat (LTR) R element to either the *gag* AUG or the first splice donor. Within its primary and secondary RNA structure, the HIV-1 5' leader encodes a series of functional domains that perform essential replicative functions throughout the viral lifecycle (1). These include the transactivation response element (TAR) (2,3), the poly-adenylation signal (polyA) (4,5), the tRNA primer binding site (PBS) (6,7), the genomic dimerization initiation site (DIS) (8), the major splice donor (SD1) (9), and the Ψ packaging signal (10). Proper RNA hairpin folding is required for many 5' leader functions (9,11-13) and is evolutionarily conserved among primate lentiviruses (14). Because the leader falls upstream of the canonical viral open reading frames (ORFs), it has the alternative designation of 5' "untranslated" region.

Based on this conserved structure and function, we became interested in evaluating whether the 5' leader sequence evolved over time during infection. We have previously developed an experimental strategy based on single viral template amplification followed by direct amplicon sequencing (single genome sequencing, SGS) that has permitted the genetic inference of transmitted HIV-1 strains (15,16) and molecular characterization of immune escape pathways (17-20). Application of this method to canonical viral ORFs has demonstrated that concentrated nonsynonymous mutations within a 30-45-nucleotide window (10-15 amino acids, toggling) serves as a strong indicator of an embedded cytotoxic T lymphocyte (CTL) epitope (18,19,21). Thus,

we applied this method to study virus evolution of the simian immunodeficiency virus (SIV) challenge strains SIVmac239/SIVmac766 during infection of Indian rhesus macaques. Strikingly, viruses from 14 of 14 Mamu-B*29:01+ macaques demonstrated rapid and complete selection confined to a 27-nucleotide (nt) span within Ψ Stem Loop 1 of the 5' leader (Figure 2.1A). 100% of mutations conferred nonsynonymous or nonsense changes within a predicted 9-amino acid span (9-mer), KGAGRYQTA (5'L-KA9). Thus, we hypothesized that 5'L-KA9 was expressed from the SIVmac 5' leader and targeted by a host CTL response.

To test this hypothesis, we performed interferon-γ (IFN-γ) ELISPOT assays, stimulating macaque PBMCs with predicted peptides from all 6 5' leader reading frames (3 forward and 3 reverse). Autologous CTLs demonstrated reactivity to the 5'L-KA9 peptide (Figure 2.1B) that was later confirmed with tetramer staining (Li, H., Shaw, G., unpublished). Ribosomal profiling of SIVmac766 demonstrated that the translation of 5'L-KA9 was initiated at two tandem one-off AUG codons (ACGACG) at positions 388-390 and 391-393, immediately upstream of 5'L-KA9 (Figure 2.1C, Warrier, R., Shaw, G., unpublished). Thus, our studies in SIV-infected macaques demonstrated that 5'L-KA9 is a novel Defective Ribosomal Product (DRiP) that is expressed from an ACG codon within the 5' leader and elicits a robust CTL response during natural infection.

Given this surprising discovery, we hypothesized that the 5' leader of HIV-1 also expressed DRiPs that would elicit host CTL responses. Here, we tested this hypothesis using longitudinal SGS, *ex vivo* immunophenotypic assays, and a mass spectrometric analysis of MHC-loaded peptides.

Section 2.3 – Results

Prevalence of 5' leader mutation during infection

To determine whether mutational selection within the 5' leader is common during HIV-1 infection, we evaluated longitudinally collected HIV-1 single genome sequences from 27 subjects (17-19,21). Sequences from 19 of these subjects were available from previous studies of viral evolution (17-19,21) (Li, H., Hahn, B., Shaw, G., unpublished) and we performed 5' leader-specific sequencing on the remaining eight subjects (CH0236, CH0694, R463F, R880F, CH0470, CH0256, CH0042, CH0850). Table 2.1 summarizes the sites and frequency of selection for longitudinal virus sequences from all 27 subjects, who were infected with either a subtype A, B, or C virus. Viruses from 13/27 subjects exhibited selection within the 5' leader over time, suggesting that 5' leader sequence evolution is common during HIV-1 infection.

HIV-1 Whole Genome Evolution in 3 Subjects

We selected three subjects, CH0236, CH0694, and R463F, for more in-depth study. These individuals were prospectively followed from acute HIV-1 infection for 1-4 years and exhibited diverse human leukocyte antigen (HLA) genotypes (Table 2.2). CH0236 and CH0694 were identified during acute HIV-1 infection, enrolled in a Center for HIV/AIDS Vaccine Immunology (CHAVI) acute infection cohort (CHAVI 001), and followed for 1.5 and 4 years of infection, respectively. R463F was identified as the acute recipient in a heterosexual transmission pair, enrolled into an International AIDS Vaccine

Initiative (IAVI) Protocol C acute infection cohort, and followed for 1 year of infection (22). Viral setpoint for all three subjects fell between 10^5-10^6 copies/mL and CD4+ T cell counts steadily declined over the course of follow up in all subjects (Figure 2.2). To compare patterns of 5' leader sequence evolution to that seen throughout other regions of the viral genome, we collected either half genome or near-full genome single genome sequences (SGSs) on human subject plasma. For CH0236 and CH0694, 3' and 5' half genome SGSs were generated at 6 and 4 time points, respectively (Figures 2.3 and 2.4). Near-full length SGSs from six R463F plasma samples had been previously published (Figure 2.5, GenBank accession numbers KJ190253-62; KP223729-75) (22).

Studies of virus sequence evolution are facilitated by genetic inference of the transmitted HIV-1 strain because it provides a reference point of comparison for all sequences obtained at later time points. 70-80% of mucosally transmitted HIV-1 infections are established by a single transmitted founder (TF) virus (23-25). To test if this was the case in CH0236, CH0694, and R463F, we applied the Poisson Fitter v2 tool. Poisson Fitter measures A) if sequences exhibit a star-like phylogeny with coalescence back to a single transmitted founder virus, B) if Hamming Distances between sequences exhibit a Poisson Distribution (nonsignificant p-value indicates a Poisson Distribution consistent with random diversification), and C) estimates time since infection. Using this tool, sequences from all three subjects coalesced to a single transmitted founder virus (TF virus, Table 2.2) (15,26). A note on the CH0236 TF inference is discussed in the Methods section.

We have previously shown that sites within the viral proteome that exhibit positive selection often overlap with cytotoxic T lymphocyte (CTL) epitopes (17-19). Therefore, we evaluated longitudinal sequences from these three subjects for regions under selection (mutually exclusive mutations relative to the TF in >67% of sequences in a 45-nt sliding window) (18,21). As expected, plasma virus SGSs from all three subjects accumulated an increasing number of sequence changes relative to the TF virus throughout the viral genome over time. Six months post trial enrollment, concentrated, nonrandom mutations relative to the TF were observed in 7-19 regions of the HIV-1 genome (Table 2.3), including the 5' leader. 17/44 (38%) of these selected sites fell within Env, which is consistent with previous studies that have linked to Env evolution to escape from either CTL responses (18,21,27) or neutralizing antibodies (19,20,27). Thus, these subjects exhibited typical HIV-1 sequence evolution throughout the proteome along side potent, early selection within the 5' leader.

5' leader evolution and immune recognition in CH0236

To further evaluate sequence evolution within the CH0236 viral 5' leader, we performed additional 5' leader SGS on eight longitudinally collected plasma samples. 5' leader sequences from CH0236 exhibited selection for non-TF variants at multiple sites during follow up (Figure 2.6A). At RNA position 169 in U5 (HXB2 position 624), a T-to-A substitution was observed in 16/36 (44%) week 24 sequences, 24/24 (100%) week 36 sequences, and 47/47 (100%) week 60 sequences. Additionally, toggling of mutually

exclusive mutations was observed in a 61-nt span (RNA 352-412, HXB2 680-740) in 11/15 (73%) week 1 sequences, 16/18 (89%) week 2 sequences, and >90% of sequences thereafter. 44/47 (94%) week 60 sequences shared an insertion mutation immediately following RNA position 255 (HXB2 710). Thus, CH0236 5' leader sequences exhibited mutational selection at two sites within the 5' leader.

To test if either of these patterns of 5' leader mutation correlated with immune escape from a CTL response, we performed autologous interferon-γ ELISPOT assays (18,21,22). Overlapping 18-mer peptides (OLPs) encoded by all six reading frames of the CH0236 TF 5' leader and the canonical viral proteome were synthesized (Sigma-Genosys or MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine). 5' leader peptides were pooled by reading frame and proteome peptides were pooled using a previously described matrix approach (18,21,28). CH0236 enrollment and week 12 cells reacted with pooled peptides encoded by the +1 reading frame (RF) of the 5' leader (Figure 2.6B). Deconvolution of this peptide pool revealed that this reactivity could be mapped to the 18-mer RRRTRLAEVHSARGEGRR (RR18). Stimulation of CH0236 PBMCs with 8-11-mers within this 18-mer span demonstrated host recognition of the autologous 11-mer AEVHSARGEGR (Figure 2.6C). To conserve cells, we then generated an RR18-specific, short-term CD8+ T cell line (29). Stimulation of this T cell line with the 11-mer AR11 and a shorter 10-mer (EVHSARGEGR, ER10) at a range of peptide concentrations revealed that ER10 was recognized by CH0236 PBMCs and was the minimal epitope (Figure 2.6C).

We next evaluated whether mutations observed in longitudinal 5' leader sequencing data overlapped with ER10. ER10 is encoded by a potential 5' leader open reading frame (ORF) that spans RNA positions 178-312, surrounding Stem Loop 1 of the Ψ packaging signal (Figure2.6D, HXB2 positions 633-765). While the ER10 ORF does not encode AUG codons, we have previously shown that SIVmac766 5'L-KA9 was expressed from a one-off AUG (*e.g.* CUG, ACG). Therefore, all one-off AUGs in the ER10 ORF are highlighted in green. Two observed mutations upstream of ER10 eliminated a one-off AUG codon (AUC) at RNA positions 226-228 (HXB2 681-683). Substitutions at 232 and 263 each incorporated a stop codon in the ER10 reading frame. Six additional substitutions were embedded in ER10. An insertion immediately downstream of RNA position 255 resulted in a frameshift within ER10. The five remaining mutations within in ER10 conferred a nonsynonymous changes at the first, fourth, or tenth residue. Autologous CH0236 T cells did not recognize any of these five ER10 variants (Figure 2.6E, **G**VHSARGER, **K**VHSARGER, EVH**L**ARGER, EVHSARGE**Q**, or EVHSARGE**W**), demonstrating that these mutations conferred CTL escape. And finally, the substitution at RNA position 287 (HXB2 742) confers a nonsynonymous change downstream of ER10. NetChop 3 analysis revealed that this mutation removes a predicted proteasomal cleavage site, and thus may affect peptide processing (30-32). In sum, toggling mutations either A) deleted a one-off AUG, B) incorporated a stop codon, C) shifted ER10 into another reading frame, or D) conferred a nonsynonymous change within or downstream of ER10.

Finally, we compared the kinetics of the ER10 response to T cell responses against the canonical viral proteome. CH0236 T cells recognized 19 autologous TF proteome peptides: Gag-QQ28, Gag-DI8, Pol-FL9, Pol-NNY9, Pol-NPY-9, Pol-KY9, Vif-KY9, Vif-SY9, Vif-DR9, Tat-KY9, Rev-AY9, Env-DL13, Env-DF9, Env-RY9, Env-GR9, Nef-WM8, Nef-QY9, Nef-LF9, and Nef-DW9 (Fig. 2.7). Concentrated, nonsynonymous mutations were embedded within 11/19 of these peptides (Figure 2.3). The 5'L-ER10 response in CH0236 was detected at trial enrollment but was smaller in magnitude when compared to the early Vif-KY9 (green) and Nef-QY9/NefLF9 (blue) responses (Fig. 2.7). Thus, the 5'L-ER10 CTL response was among the earliest CTL responses and potently drove virus escape.

5' leader evolution and immune recognition in CH0694

Similar to CH0236, 5' leader sequencing over three years of CH0694 infection revealed multiple, complex patterns of selection (Figure 2.8A). At week 24, 26/30 (87%) SGSs demonstrated toggling mutations between RNA positions 44-59 (HXB2 497-513). At subsequent time points, a base substitution at position 50 was observed in all variants. All week 24 variants also encoded a substitution at RNA position 179 (HXB2 633), which was observed for the remainder of follow up. Complex toggling was observed between positions 200-264 (HXB2 654-718). This toggling, which resembled that seen in CH0236 both in span and location, involved single and double mutations at RNA positions 200 and 201 (HXB2 654-655), a substitution at position 228 (HXB2 682), an

insertion downstream of position 238 (HXB2 692), an insertion downstream of position 254 (HXB2 708), and a substitution at position 263 (HXB2 717). One of these mutations was present in 30/34 (88%) week 96 sequences and 23/23 (100%) week 140 sequences.

Screening ELISPOT assays revealed that CH0694 PBMCs taken from week 16 and 48 reacted with pooled 5' leader peptides from +3 RF and +1 RF, respectively (Figure 2.8B). Deconvolution of the reading frame +3 pool response demonstrated that the reactive peptide was a 9-mer LAIWGTHCL, encoded by RNA positions 40-58 of the TAR stem loop (5'-LL9, Figure 2.8C). 5/6 toggling mutations fell within 5'L-LL9 and conferred nonsynonymous changes to the epitope, whereas the selected variant encoded a stop codon at the fourth amino acid (Figure 2.8D). Due to restrictions on cell availability, we were unable to test 5'L-LL9 variant peptides in IFN-γ ELISPOT assays.

Deconvolution of the RF+1 pool response revealed immune recognition of two partially overlapping peptides, ISRRRTRLA (5'L-IA9) and RTRLAEVHS (5'L-RS9, Figure 2.8E). 5'L-IA9, 5'L-RS9, and 5'L-ER10 were all encoded by the same open reading frame; that is, they shared upstream one-off AUG codons. Consistent with this finding, mutations in the CH0694 5'L-IA9/5'L-RS9 reading frame closely resembled those observed in CH0236 5'L-ER10 (Figure 2.8F). In the CH0236 TF virus, one-off AUGs upstream of 5'L-ER10 were observed at positions 193, 223, 226, and 238 (HXB2 positions 648, 678, 681, and 693). In the CH0694 TF virus, upstream one-off AUGs were found at 179, 194, 200, 227, and 239 (HXB2 positions 633, 648, 654, 681, and 693). Variants that deleted the one-off AUG at HXB2 position 681 were observed in both subjects (A**U**C, bolded residue mutated). Additional variants in CH0694 harbored

mutations in one-off AUGs at HXB2 positions 633 (RNA 179, **A**A**G**) and 654 (RNA 199, **UU**G), which were not present in the CH0236 TF virus. Thus, even though these two subjects mounted immune responses to different peptides in this 5' leader reading frame, virus variants from both subjects exhibited similar patterns of mutation, including deletion of the same one-off AUG.

Finally, we compared the kinetics of the anti-5' leader T cell responses against those targeting peptides derived from the viral proteome. CH0694 PBMCs from the first year of infection recognized 13 peptides from the proteome: Gag-YL9, Gag-TE18, Gag-TA18, Gag-GL9, Pol-Tl9, Vif-WI9, Vpr-CL9, Rev-IL9, Rev-DN18, Env-VL18, Env-RR11, Nef-KM9, and Nef-TL10 (Figure 2.9). Concentrated non-random, nonsynonymous changes were embedded in 10/13 (77%) of these epitopes during follow up. Interestingly, 43/52 (83%) week 24 sequences harbored a single nucleotide insertion in Env-RR11, throwing the cytoplasmic tail of this protein out of frame (Figure 2.4, orange tick marks). The resulting Env protein had a 19-amino acid cytoplasmic tail truncation that was later replaced by one of two other variants that restored the reading frame. Similar to CH0236, 5' leader responses were detected throughout CH0694 infection, albeit at lower magnitude than responses against canonical proteome epitopes: T cells specific for 5'L-LL9 were detectable at enrollment and quickly waned, whereas T cells specific for other 5' leader peptides arose later.

5' leader evolution and immune recognition in R463F

A previous study of subject R463F mapped autologous T cell responses against the entire viral proteome, but did not investigate 5' leader sequence evolution or T cell responses (22). We, therefore, performed 5' leader sequencing on three time points of infection (Figure 2.10A, plasma availability restricted analysis). Of note, the R463F TF virus harbors a 20-nt insertion downstream of the PBS that is seen ~20% of subtype A viruses and most CRF_01 subtype AE viruses (M group alignment, www.hiv.lanl.gov). Sequences derived from week 24 plasma demonstrated toggling of mutations between RNA positions 274-287 (HXB2 717-723). These included an insertion downstream of nucleotide 274 (HXB2 709), a single nucleotide deletion at RNA position 281 (HXB2 717), a substitution at 283 (HXB2 719), and a two-nt deletion at 286-287 (HXB2 722-723). Previously published sequences harbored two more sequences changes – an insertion at 257 (HXB2 692) and a substitution at 259 (HXB2 694) (22).

Screening ELISPOT assays demonstrated that R463F PBMCs reacted with pooled peptides from the +3 RF and deconvolution of this pooled mapped the minimal epitope to AECTQQEARA, which was encoded by Stem Loop 1 of the Ψ packaging signal (RNA positions 270-299, HXB2 726-757, Figure 2.10B and C, 5'L-AA10). Similar to CH0236 and CH0694, R463F viral variants observed during infection harbored a mutation in an upstream one-off AUG, a substitution within 5'L-AA10, and frameshift mutations. Week 46 PBMCs reacted with the TF 5'L-AA10 variant, but not the week 24 variant (AECTRQERA, Figure 2.10D). Thus, all three subjects mounted CTL responses against 5' leader encoded peptides that potently drove viral turnover and escape.

Escape mutation location on stem loop structure

Previous studies have shown that the RNA sequence and secondary structure of the HIV-1 5' leader play crucial roles in the leader's function (9,12,33). We, therefore, mapped the mutations observed during 5' leader sequence evolution in CH0236, CH0694, and R463F onto a previously published 5' leader secondary structure (14) to determine if any of these mutations fell in functional stem loops. This structure, which was determined using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), was used as a scaffold that we threaded each TF virus sequence into (Figure 2.11 and Figure 2.12) (14). Because the R463F encoded a 20-nt insertion in this region downstream of the primer binding sequence (PBS), it could not be aligned to the sequence used to build the SHAPE structure (NL4.3), and we omitted this part of the structure in Figure 2.11. 5'L-ER10, 5'L-IA9, 5'L-RS9 and 5'L-AA10 (bracketed) all overlapped with Ψ stem loop 1, which contains the palindromic dimerization initiation site (DIS). While this palindromic sequence can differ between subtypes, all three TF viruses encoded GUGCAC (Figure 2.11). Mutations observed in longitudinal sequence datasets all spared the DIS and, in CH0694 and CH0236, spared the residues in the stem. 5'L-LL9 was encoded by the TAR stem loop of R (Figure 2.12). The sites of mutations in CH0694 viruses are shown relative to the functional TAR loops/bulges, which have known functions in directly engaging both Tat and Cyclin T1 during transactivation (34-36). The G-to-A substitution at RNA position 50 (starred) was selected for during CH0694 follow up and alters a G-U base pair to a Watson-Crick A-U base pair. Thus,

escape mutations spared functional sequences like the DIS or Tat binding bulge and were predicted to maintain stem base-pairing.

Mass Spectrometry Detection of a 5' Leader Peptide

We next attempted to directly detect 5' leader DRiPs during *in vitro* infection. Ternette and colleagues recently developed a mass spectrometry-based approach for the identification of peptides loaded onto MHC complexes from cells infected with HIV-1 (37). We applied this technique HIV-1 IIIB-infected CD4+ T cells from a donor who expressed HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, HLA-C*07:01, and HLA-C*07:02 to screen for 5' leader peptides. Mass spectrometry analysis identified four HIV-1 peptides: two overlapping 5' leader peptides (LSLGALWLAREPTA and LSLGALWL, Figure 2.13) and two overlapping Gag peptides (FLGKIWPSY and FLGKIWPS, data not shown). To verify peak assignments, we synthesized LSLGALWLAREPTA (5'L-LA14) and collected spectra on this synthetic standard (Figure 2.13 bottom). Spectral matching between the infected-cell derived and synthetic peptide suggested that peak assignment was correct. Consistent with the HLA of this donor, both 5'L-LA14 and Gag-FS8 were predicted HLA-A*02:01 epitopes (NetMHC predicted affinity <500nM) (38). 5'L-LA14 was encoded by RNA positions 23-64 (HXB2 477-518) in a distinct reading frame from 5'L-LL9, but also overlapping with the TAR stem loop. Thus, mass spectrometry of *in vitro* infection can be readily used to

detect 5’ leader DRiPs and facilitated the identification of a protein encoded by the first ~60 nts of the HIV-1 RNA.

Section 2.4 – Conclusions

We recently demonstrated that the 5' leader of SIVmac766 expresses an immunogenic peptide called 5'L-KA9. In this study, we investigated whether diverse HIV-1 strains expressed peptides from their 5' leader sequences by correlating virus and CTL evolution and developing a mass spectrometry-based approach for the identification of MHC-loaded peptides. We provided evidence that at least six 5' leader peptides are expressed from four different subtype A, B, and C HIV-1 strains. Further, we reported indirect evidence of 5' leader peptide immune pressure and viral escape in 13/27 subjects.

Our findings, while rooted in the observation of immune escape variants that arise during infection, have implications for models of lentivirus translation regulation. During typical cap-dependent eukaryotic translation, the 43S preinitiation complex (comprised of the 40S ribosome and ternary complex) binds to the 7-methylguanylate 5' RNA cap and scans 5' to 3' through the RNA leader sequence until it reaches a translation initiation codon (39). Unlike most other eukaryotic 5' "untranslated" regions, the HIV-1 5' leader is long (~336 nts in unspliced transcripts) and exhibits extensive secondary structure, two features that impede cap-dependent ribosomal scanning *in vitro* (40-42). Whether HIV-1 translation initiation occurs via 5' RNA cap-dependent (43) or cap-independent (44-47) mechanisms is a matter of controversy within the field. The identification of 5' leader DRiPs in all three forward reading frames, however, suggests that cap-dependent ribosomal scanning is a major contributor to HIV-1 translation initiation. Moreover, these data provide direct evidence that ribosomes frequently initiate translation at one-off AUGs while scanning through this region (48-51). Future studies will be needed to

investigate the factors that impact translation initiation efficiency at these one-off AUGs, including Kozak context and RNA secondary structure (52-54).

This study also expands upon our previous observation of 5' leader DRiP expression from SIVmac766. Here, we identified 5' leader peptides encoded by all three forward reading frames of Subtype A, B, and C viruses using multiple different methods. Collectively, these findings suggest that 5' leader DRiP expression is a general feature of lentivirus infection. Further, mutations during escape from peptides derived from the same 5' leader ORF (5'L-ER10, 5'LIA9/5'L-RS9) often overlapped with one another and have help identify candidate ORF start codons. We have recently proposed that lentiviral 5' leader DRiPs offer an exciting new vaccine target and are in the process of testing 5' leader peptide vaccine efficacy using the RhCMV68.1 platform (55-57). Translation of these rhesus/SIV vaccines into an HIV-1 vaccine would require an understanding of which 5' leader ORFs are expressed and how often these ORFs are present in circulating HIV-1s. This study sheds light on both of these considerations and, along with future studies, will help define the rules underlying 5' leader DRiP expression.

Section 2.5 – Methods

Viral sequencing

Single genome amplification followed by direct amplicon sequencing (SGS) was performed for 3' half and 5' half genomes as previously described (17-19,21).

CH0236 TF Inference

CH0236 sequences exhibited shared differences from the first time point consensus within the *rev1-tat1* overlap, changes that could confound TF inference. Despite these shared changes, Poisson fitter analysis of both the 5' and 3' half genomes yielded a Poisson distribution of Hamming Distances (p=0.515 for the 3' half genome) and a star-like phylogeny. This pattern is consistent with one of two scenarios: recent infection by multiple viruses that differed from one another at one nucleotide or infection by a single TF virus that developed early escape mutations from an early autologous T cell response. 26 of 36 screening time point sequences harbored a single nucleotide difference when compared to the most recent common ancestor (MRCA) of these sequences. Of these 26 changes, 26 were nonsynonymous in a Rev peptide (MRCA variant AVRTKILLY, AY9). CH0236 PBMCs reacted with Rev-AY9 (data not shown), therefore, we concluded that CH0236 was infected with a single TF virus, whose genetic identity was the coalescent of the screening time point sequences. CH0694 and CH0236 were each infected with a subtype C virus and R463F was infected with a subtype A virus.

Interferon-γ ELISPOT assays

ELISPOT assays were performed as previously described (18,21). Short term cell lines were generated as previously described (29)

HIV IIIB infection

Primary $CD4^+$ T cells were isolated from Donor 97 (HLA type: A*01:01, A*02:01, B*07:02, B*08:01, C*07:01, C*07:02) with the $CD4^+$ T cell isolation kit (Miltenyi Biotec) according to the manufacturer's instructions. Cells were infected the lab-adapted strain HIV-1 IIIB at an MOI of 0.1 then cultured in RPMI 1640 medium containing 20% FBS and 20 U/ml IL-2 at 2 x 10^6/ml. Purification of MHC class I-bound peptides was performed as described previously (Ternette et al 2015 EJI). Briefly, peptide-MHC class I complexes were captured from cleared cell lysates by protein A resin coupled to W6/32 antibody. Complexes were washed with 50 mM Tris buffer pH 8.0 first with 150 mM NaCl, then 450 mM NaCl then without salt. Complexes were then eluted in 10% acetic acid, dried then loaded onto a 4.6 x 50 mm ProSwift RP-1S column (ThermoFisher Scientific) and eluted using a 500 ul/min flow rate over 10 minutes from 2-35% Buffer B (0.1% formic acid in acetonitrile) in Buffer A (0.1% formic acid in water) using an Ultimate 3000 HPLC system (ThermoFisher Scientific).

LC-MS/MS

Peptides were separated on an ekspert nanoLC 400 cHiPLC system (Eksigent) supplemented with a 15 cm x 75 µm ChromXP C18-CL, 3 µm particle size by application of a linear gradient from 8% buffer A to 35% buffer B at a flow rate of 300 nl/min for 60 min. Peptides were introduced by an electrospray source to a TripleTOF 5600 mass spectrometer (Sciex). Collision-induced dissociation was induced on the 30 most abundant ions per full MS scan using ramped collision energy and a unit quadrupole isolation width of 0.7 amu. All fragmented precursor ions were actively excluded from repeated selection for 15 s.

Raw data were analyzed using Peaks 7.0 software (Bioinformatics Solutions). Sequence interpretation of MS/MS spectra were performed using a database containing all annotated human SwissProt entries including translations of all six reading frames of the sequenced HIV-1 IIIB genome in addition to translations of all known assigned HIV-1 protein coding regions (GenBank KJ925006). The false discovery rate was determined by decoy database searches and a general cutoff below 5% was applied.

Section 2.6 – References

1. Berkhout B. Structure and function of the human immunodeficiency virus leader RNA. Prog Nucleic Acid Res Mol Biol. 1996 ed. 1996;54:1–34.

2. Berkhout B, SILVERMAN RH, Jeang KT. Tat Trans-Activates the Human Immunodeficiency Virus Through a Nascent Rna Target. Cell. 1989;59(2):273–82.

3. SELBY MJ, BAIN ES, LUCIW PA, PETERLIN BM. Structure, Sequence, and Position of the Stem Loop in Tar Determine Transcriptional Elongation by Tat Through the Hiv-1 Long Terminal Repeat. Genes & Development. 1989 Apr;3(4):547–58.

4. ASHE MP, GRIFFIN P, JAMES W, PROUDFOOT NJ. Poly(a) Site Selection in the Hiv-1 Provirus - Inhibition of Promoter-Proximal Polyadenylation by the Downstream Major Splice Donor Site. Genes & Development. 1995;9(23):3008–25.

5. ASHE MP, Pearson LH, PROUDFOOT NJ. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. Embo J. 1997;16(18):5752–63.

6. Beerens N, Klaver B, Berkhout B. A structured RNA motif is involved in correct placement of the tRNA(3)(Lys) primer onto the human immunodeficiency virus genome. Journal of Virology. American Society for Microbiology (ASM); 2000 Mar;74(5):2227–38.

7. Beerens N, Berkhout B. The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription. Journal of Virology. American Society for Microbiology (ASM); 2002 Mar;76(5):2329–39.

8. Ennifar E, Yusupov M, Walter P, Marquet R, Ehresmann B, Ehresmann C, et al. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. Structure. 1999 Nov 15;7(11):1439–49.

9. Abbink TEM, Berkhout B. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. Journal of Virology. American Society for Microbiology; 2008 Mar;82(6):3090–8.

10. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. Science. 2011 ed. 2011 Oct 14;334(6053):242–5.

11. van Bel N, Das AT, Berkhout B. In Vivo SELEX of Single-Stranded Domains in the HIV-1 Leader RNA. Journal of Virology. 2014 Jan 31;88(4):1870–80.

12. van Bel N, Ghabri A, Das AT, Berkhout B. The HIV-1 leader RNA is exquisitely sensitive to structural changes. Virology. 2015 Sep;483:236–52.

13. Das AT, Klaver B, Berkhout B. A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation

site selection. Journal of Virology. American Society for Microbiology (ASM); 1999 Jan;73(1):81–91.

14. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, et al. Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. PLoS Pathog. 2013 ed. 2013;9(4):e1003294.

15. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci USA. 2008 May 27;105(21):7552–7.

16. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. Journal of Virology. 2008 Apr;82(8):3952–70.

17. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med. 2009 Jun 8;206(6):1273–89.

18. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov VV, et al. The first T cell response to transmitted/founder virus contributes to the

control of acute viremia in HIV-1 infection. J Exp Med. 2009 ed. 2009 Jun 8;206(6):1253–72.

19. Bar KJ, Tsao C-Y, Iyer SS, Decker JM, Yang Y, Bonsignori M, et al. Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. PLoS Pathog. 2012;8(5):e1002721.

20. Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature. 2013 Apr 25;496(7446):469–76.

21. Liu MK, Hawkins N, Ritchie AJ, Ganusov VV, Whale V, Brackenridge S, et al. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. J Clin Invest. 2012 ed. 2013 Jan;123(1):380–93.

22. Yue L, Pfafferott KJ, Baalwa J, Conrod K, Dong CC, Chui C, et al. Transmitted virus fitness and host T cell responses collectively define divergent infection outcomes in two HIV-1 recipients. PLoS Pathog. 2015 ed. 2015 Jan;11(1):e1004565.

23. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci USA. 2008 ed. 2008 May 27;105(21):7552–7.

24. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et

al. Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. Swanstrom R, editor. PLoS Pathog. Public Library of Science; 2016 May;12(5):e1005619.

25. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. Journal of Virology. 2008 ed. 2008 Apr;82(8):3952–70.

26. Giorgi EE, Funkhouser B, Athreya G, Perelson AS, Korber BT, Bhattacharya T. Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. BMC Bioinformatics. BioMed Central; 2010;11(1):532.

27. Gao F, Bonsignori M, Liao H-X, Kumar A, Xia S-M, Lu X, et al. Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. Cell. 2014 Jul 31;158(3):481–91.

28. Roederer M, Koup RA. Optimized determination of T cell epitope responses. J Immunol Methods. 2003 Mar 1;274(1-2):221–8.

29. Goonetilleke N, Moore S, Dally L, Winstone N, Cebere I, Mahmoud A, et al. Induction of multifunctional human immunodeficiency virus type 1 (HIV-1)-specific T cells capable of proliferation in healthy subjects by using a prime-boost regimen of DNA- and modified vaccinia virus Ankara-vectored vaccines expressing HIV-1 Gag coupled to CD8+ T-cell epitopes. Journal of Virology.

American Society for Microbiology; 2006 May;80(10):4717–28.

30. Nielsen M, Lundegaard C, Lund O, Keşmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics. 2005 Apr;57(1-2):33–41.

31. Draenert R, Le Gall S, Pfafferott KJ, Leslie AJ, Chetty P, Brander C, et al. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. J Exp Med. 2004 ed. 2004 Apr 5;199(7):905–15.

32. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, Le Gall S, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. Journal of Virology. American Society for Microbiology; 2004 Jul;78(13):7069–78.

33. Berkhout B. Structure and function of the human immunodeficiency virus leader RNA. Prog Nucleic Acid Res Mol Biol. 1996;54:1–34.

34. Richter S, Cao H, Rana TM. Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. Biochemistry. 2002 May 21;41(20):6391–7.

35. Richter S, Ping Y-H, Rana TM. TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. Proc Natl Acad Sci USA. National Acad

Sciences; 2002 Jun 11;99(12):7928–33.

36. KARN J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med. 2012 ed. 2012 Feb;2(2):a006916.

37. Ternette N, Block PD, Sanchez-Bernabeu A, Borthwick N, Pappalardo E, Abdul-Jawad S, et al. Early Kinetics of the HLA Class I-Associated Peptidome of MVA.HIVconsv-Infected Cells. Journal of Virology. 2015 ed. 2015 Jun;89(11):5760–71.

38. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. Nucleic Acids Res. Oxford University Press; 2008 Jul 1;36(Web Server issue):W509–12.

39. Hinnebusch AG. eIF3: a versatile scaffold for translation initiation complexes. Trends Biochem Sci. 2006 Oct;31(10):553–62.

40. Miele G, Mouland A, Harrison GP, Cohen E, Lever AM. The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. Journal of Virology. 1996 ed. 1996 Feb;70(2):944–51.

41. Parkin NT, Cohen EA, Darveau A, Rosen C, Haseltine W, Sonenberg N. Mutational analysis of the 5' non-coding region of human immunodeficiency virus

type 1: effects of secondary structure on translation. Embo J. 1988 ed. 1988 Sep;7(9):2831–7.PMC457075.

42. Bolinger C, Boris-Lawrie K. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. Retrovirology. 2009 ed. 2009;6:8.

43. Berkhout B, Arts K, Abbink TE. Ribosomal scanning on the 5'-untranslated region of the human immunodeficiency virus RNA genome. Nucleic Acids Res. 2011 ed. 2011 Jul;39(12):5232–44.

44. Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. Journal of Virology. 2000 ed. 2001 Jan;75(1):181–91.

45. Brasey A, Lopez-Lastra M, Ohlmann T, Beerens N, Berkhout B, Darlix JL, et al. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. Journal of Virology. 2003rd ed. 2003 Apr;77(7):3939–49.

46. Vallejos M, Deforges J, Plank TD, Letelier A, Ramdohr P, Abraham CG, et al. Activity of the human immunodeficiency virus type 1 cell cycle-dependent internal ribosomal entry site is modulated by IRES trans-acting factors. Nucleic Acids Res. 2011 ed. 2011 Aug;39(14):6186–200.

47. Plank T-DM, Whitehurst JT, Kieft JS. Cell type specificity and structural

determinants of IRES activity from the 5' leaders of different HIV-1 transcripts. Nucleic Acids Res. 2013 May 9;41(13):6698–714.

48. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, et al. Decoding human cytomegalovirus. Science. 2012 ed. 2012 Nov 23;338(6110):1088–93.

49. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009 ed. 2009 Apr 10;324(5924):218–23.

50. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T, et al. Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. Science. 2012 Jun 29;336(6089):1719–23.

51. Starck SR, Shastri N. Non-conventional sources of peptides presented by MHC class I. Cell Mol Life Sci. 2011 Mar 10;68(9):1471–9.

52. Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 1987 Oct 26;15(20):8125–48.

53. Kozak M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol Cell Biol. American Society for Microbiology (ASM); 1989 Nov;9(11):5073–80.

54. Kozak M. Downstream secondary structure facilitates recognition of initiator

codons by eukaryotic ribosomes. Proc Natl Acad Sci USA. 1990 Nov;87(21):8301–5.

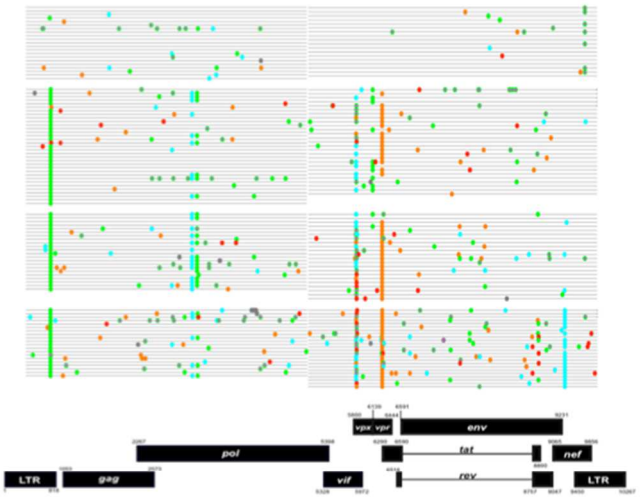
55. Hansen SG, Sacha JB, Hughes CM, Ford JC, Burwitz BJ, Scholz I, et al. Cytomegalovirus vectors violate CD8+ T cell epitope recognition paradigms. Science. 2013 May 24;340(6135):1237874.

56. Hansen SG, Vieville C, Whizin N, Coyne-Johnson L, Siess DC, Drummond DD, et al. Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. Nature Medicine. 2009 Mar;15(3):293–9.

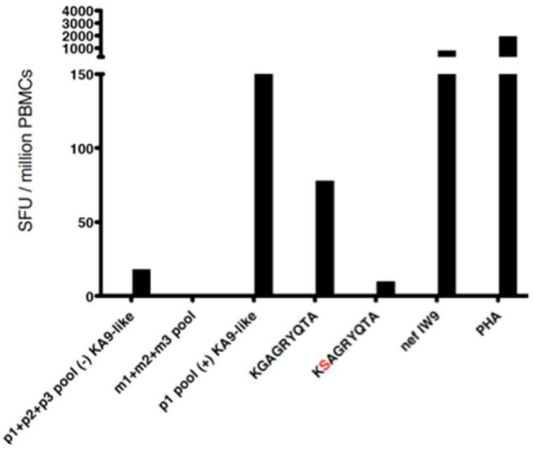
57. Hansen SG, Piatak MJ, Ventura AB, Hughes CM, Gilbride RM, Ford JC, et al. Immune clearance of highly pathogenic SIV infection. Nature. 2013 ed. 2013 Oct 3;502(7469):100–4.

Section 2.7 Figures and Tables

A)



B)



C)

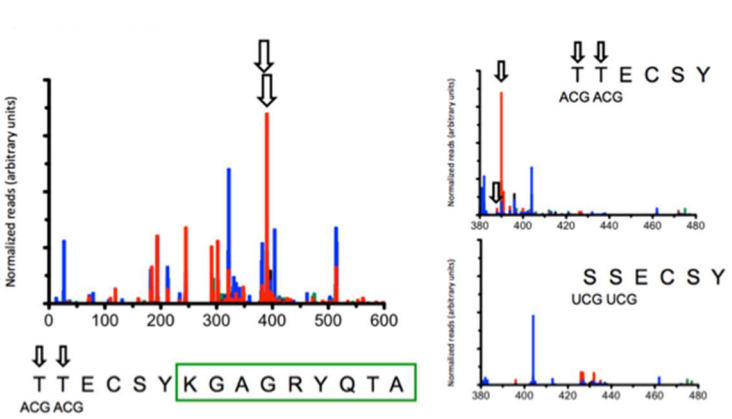


Figure 2-1 Expression and immune recognition of the SIVmac766 5' leader DRiP KA9

A) Highlighter plot of HIV-1 half genome single genome sequences (SGSs). Horizontal lines represent SGSs and tick marks denote mutations relative to the infecting stock (SIVmac766). Green = mutation to A, Blue = C, Orange = G, Red = T, and Grey = gap. B) Interferon-γ ELISPOT assays. Peptides and overlapping peptide pools used to stimulate autologous PBMCs from macaque HD35 are indicated at the bottom. Plotted on the y-axis are spot forming units/million PBMCs. All non-KA9 peptides encoded by the 5' leader (p1+p2+p3 pool (-) KA9-like and m1+m2+m3 pool representing plus frame and minus frame peptides, respectively) were not recognized by autologous PBMCs. KA9 and any KA9-containing peptide (*e.g.* CSYKGAGRYQTA) were recognized. The selected variant at week 4 (see A) – KSAGRYQTA – was not recognized, demonstrating complete viral escape. Positive controls include a Nef-(W9 peptide and phytohemagglutinin (PHA). C) Ribosomal profiling of the SIVmac 766 5' laeder. Shown are the sites of ribosome protected fragments (x-axis) and the normalized number of reads (y-axis) for experiments run under four conditions: red = lactimidomycin (LTM), freezes initiating ribosomes; green = cyclohexamide, freezes initiating or elongating ribosomes; blue = Pateamine A, freezes stalled ribosomes; and black = DMSO, drug solvent control. Enriched LTM signal (red) indicates initiating ribosomes. Arrows indicate two one-off AUG codons (ACGs) upstream of KA9 (shown below). Right: Comparison of the wildtype and a start site mutant SIVmac766 in which the putative start codons (ACG ACG, arrows) were replaced with two-off codons (UCG UCG). Panel depicts zoom in on the ribosomal profile of the KA9 ORF for the wildtype (top) and start site mutant (bottom) SIVmac766.

Table 2-1 Prevalence of selection in longitudinal HIV-1 5' leader sequence datasets

Subject	Subtype	R	U5/PBS	DIS/SD/Ψ
CH0236	C		x	x
CH0694	C	x	x	x
R463F	A			x
R880F	A	x		
CH0325	C			x
CH1432	C			x
CH0256	C	x		
CH0042	C			x
CH0164	C			x
CH0058	B		x	
CH0198	C			x
CH0040	B	x		
CAP210	C	x		
CH0185	C			
CH0107	C			
CH0067	C			
CH0077	B			
CH0131	C			
CH0159	C			
CH0162	C			
CH0264	C			
CH0443	C			
CH0569	C			
CH0752	C			
CH0470	B			
CH0850	C			

First column indicates subject ID, second column subtype of HIV-1 infection, and remainder represent regions of the 5' leader. X indicates selection detected

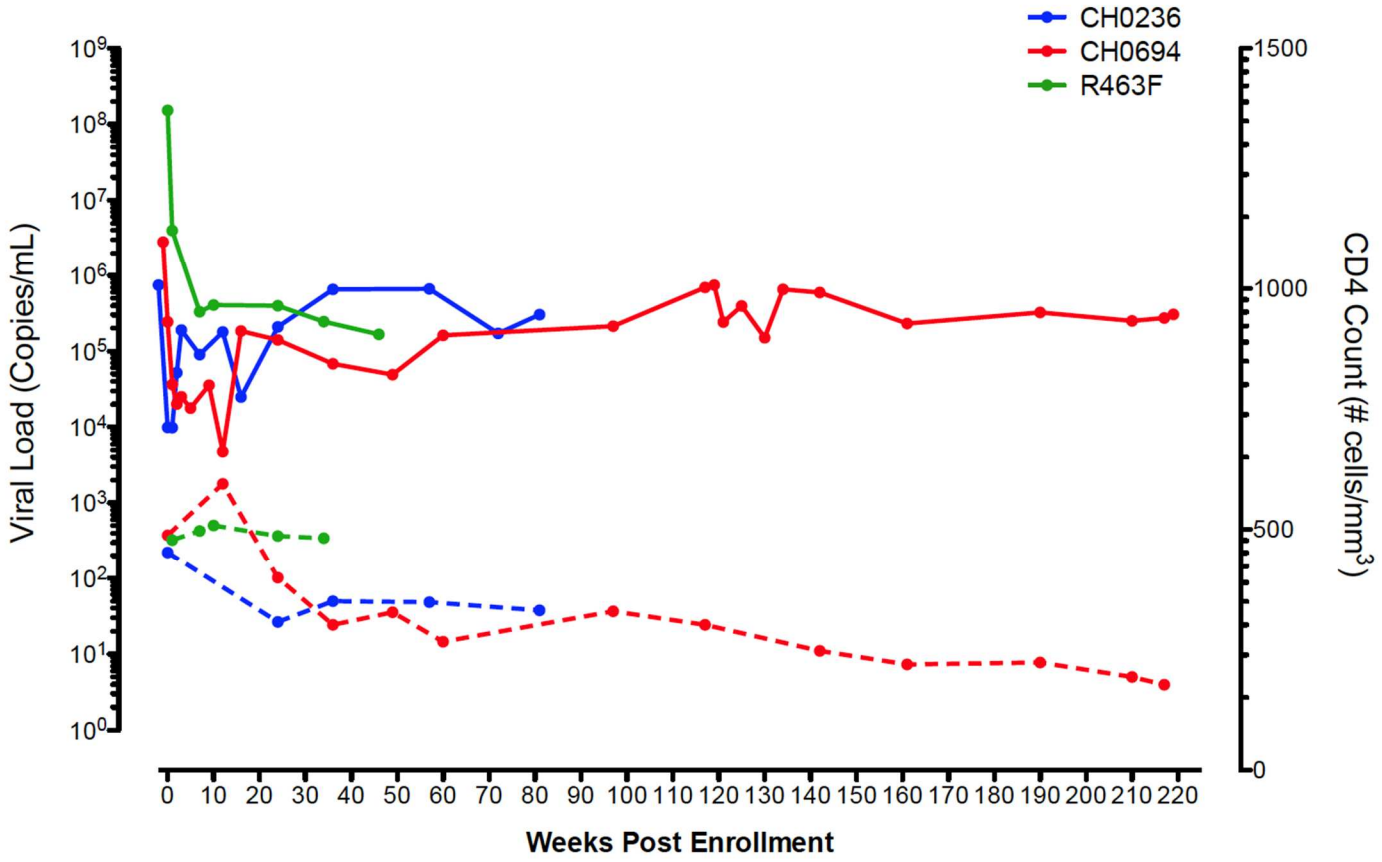


Figure 2-2 Viral load and CD4 counts for study subjects

Table 2-2 HLA typing for study subjects

Participant Id	CH0236	CH0694	R463F
HLA-A allele 1	A*30:02/33	A*30:01	A*01:01
HLA-A allele 2	A*66:01/08	A*74:01/02/03/09	A*30:02
HLA-B allele 1	B*08:01/19N	B*42:02	B*15:03
HLA-B allele 2	B*18:01/17N	B*15:10	B*45:01
HLA-C allele 1	C*07:01/06/18/52/ 57/116	C*17:01/02/03/04	C*02:10
HLA-C allele 2	C*17:01/02/03	C*03:04/06/09/19/23/24/26/32/37/44/46/48/54/ 57/63/64/72/73/74/77/78	C*06:02

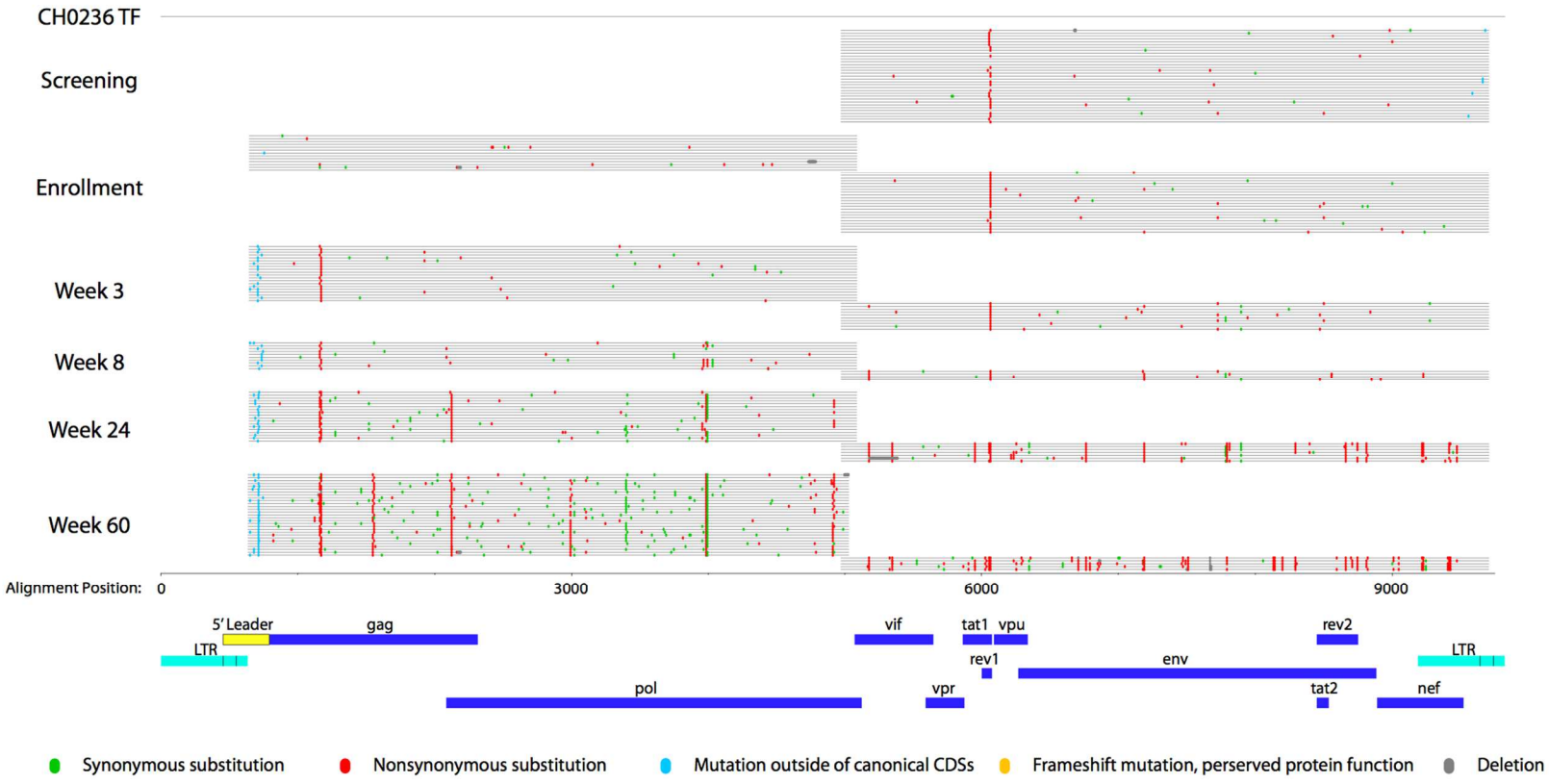


Figure 2-3 Longitudinal whole genome SGS for subject CH0236

Depicted is a synonymous/nonsynonymous mutation highlighter plot. Horizontal lines represent sequences and tick marks denote changes relative to the inferred transmitted founder virus (TF). Color code indicated at the bottom. Sequences are arranged

longitudinally from top to bottom with the trial time point indicated to the left. Alignment position (not HXB2) indicated at the bottom. Genomic organization is shown at the bottom with relevant viral landmarks noted.

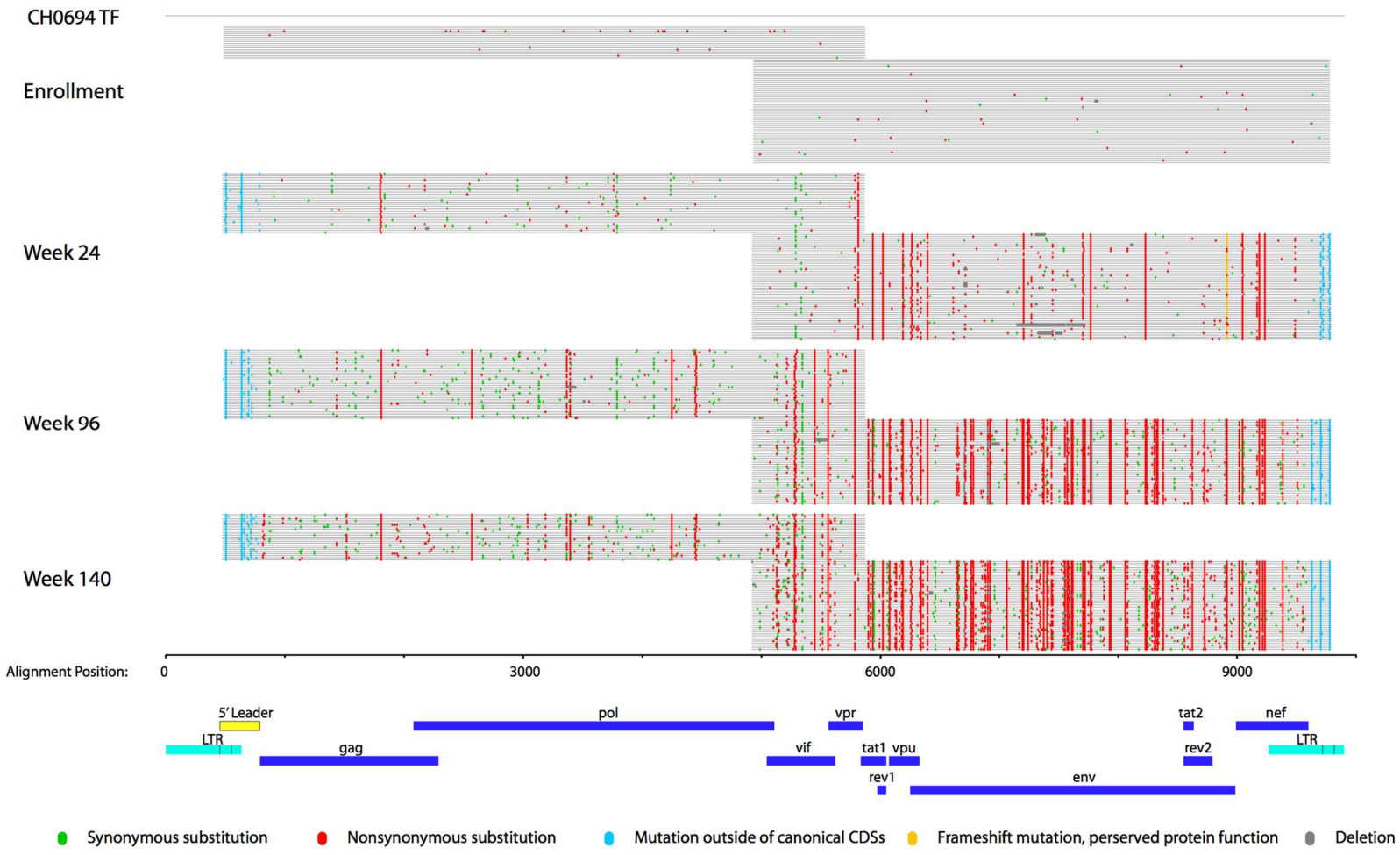

CH0694 TF
Enrollment
Week 24
Week 96
Week 140
Alignment Position:
0
3000
6000
9000
5' Leader
LTR
gag
pol
vif
vpr
tat1
vpu
rev1
env
tat2
rev2
nef
LTR
Synonymous substitution
Nonsynonymous substitution
Mutation outside of canonical CDSs
Frameshift mutation, perserved protein function
Deletion

Figure 2-4 Longitudinal whole genome SGS for subject CH0694

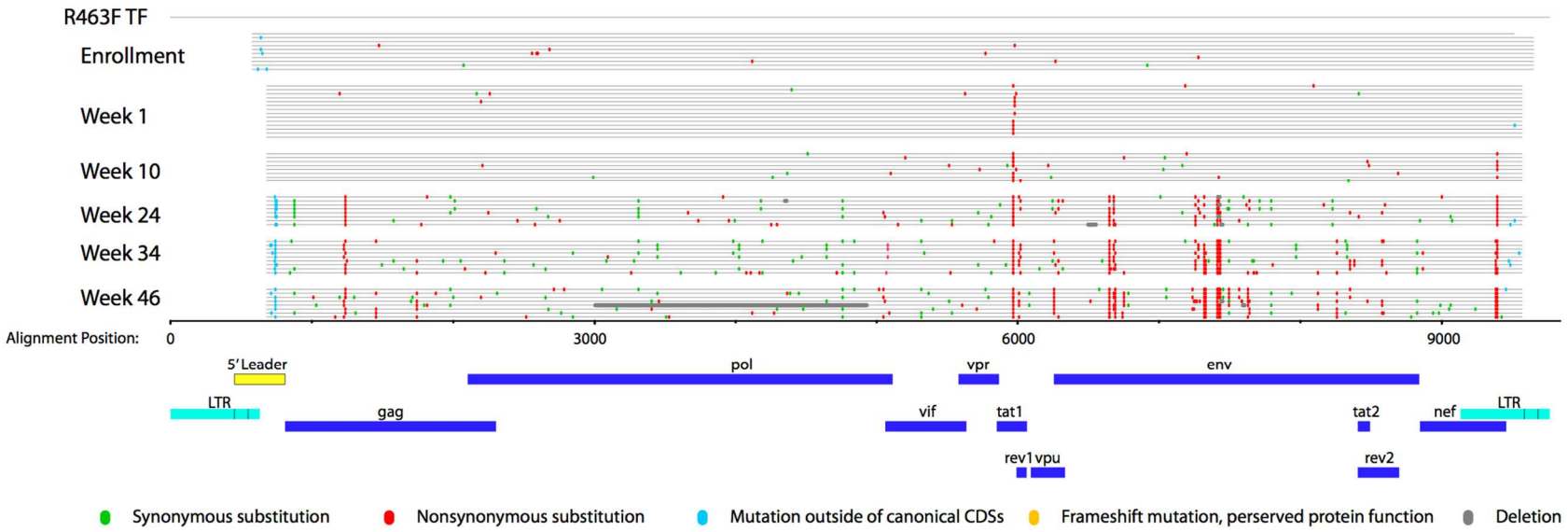


Figure 2-5 Longitudinal whole genome SGS for subject R463F

Table 2-3Poisson fitter analysis

Subject	Star-like phylogeny	p-value	Estimated Days Post Infection	Harmonic mean (Days)	95% Confidence Interval
CH0236.5	Yes	0.895	17	21	(18,24)
CH0236.3	Yes	0.515	27		
CH0694.5	Yes	0.390	9	9	(6,12)
CH0694.3	Yes	0.704	11		
R463F	Yes	0.668	12	N/A	(6,18)

Sequences from the earliest time point were analyzed using Poisson Fitter v2. Sequences used are indicated in the left column (.3 and .5 represent 3' and 5' half genomes, repsecitvely). Nonsignificant p-value indicates a Poisson distribution of Hamming Distances, consistent with random diversification from a single virus. Dates from infection estimates are indicated. To obtain a single estimate for subjects with two date estimates, the harmonic mean was taken. 95% confidence interval for the date of infection estimates is indicated to the far right.

Table 2-4 Sites under selection at week 24

CH0236										
Region	5'L	Gag	Pol	Pol	Vif	Vif	Tat	Rev	Env	
HXB2	680-719	1160-1177	2118-2130	3931-3965	5150	5318	5917	6022-6030	6705	
#SGS	18/18	18/18	18/18	18/18	7/7	7/7	7/7	5/7	7/7	
Region	Env	Env	Env	Env	Env	Env	Nef	Nef	Nef	
HXB2	7113	7718-7723	7826	8587g	8657	8718	9110-9120	9308	9522	
#SGS	7/7	7/7	5/7	7/7	5/7	6/7	6/7	5/7	5/7	
R463F										
Region	5'L	Gag	Tat	Env	Env	Env	Nef			
HXB2	717-723	1242-1247	5947-5949	6618-6621	6658-6662	7397-7408	9347			
#SGS	14/14	14/14	14/14	14/14	10/14	14/14	14/14			
CH0694										
Region	5'L	5'L	Gag	Vpr	Tat	Tat/Rev	Vpu	Env	Env	
HXB2	497-513	633	1804-1808	5789-5810	5933-5934	6019	6164-6172	6232-6246	6372	
#SGS	29/30	30/30	30/30	77/82	52/52	52/52	48/52	52/52	46/52	
Region	Env	Env	Env	Env	Env	Nef	Nef	Nef	U3	R
HXB2	7062-7063	7548-7555	7600-7601	8070	8728-8732	8867	9008	9049	9525-9538	9582-9591
#SGS	50/52	47/52	51/52	52/52	52/52	49/52	52/52	52/52	52/52	52/52

A sliding window of 45-nts (15 amino acids) was used to detect positive selection due to CTL escape (Goonetilleke, et al., 2009) in sequences taken ~6 months post infection. Indicated is the region of the genome that the selection fell in, the HXB2 coordinates of the span of nucleotide mutations, and the number of sequences of the total number with a mutation in that span.

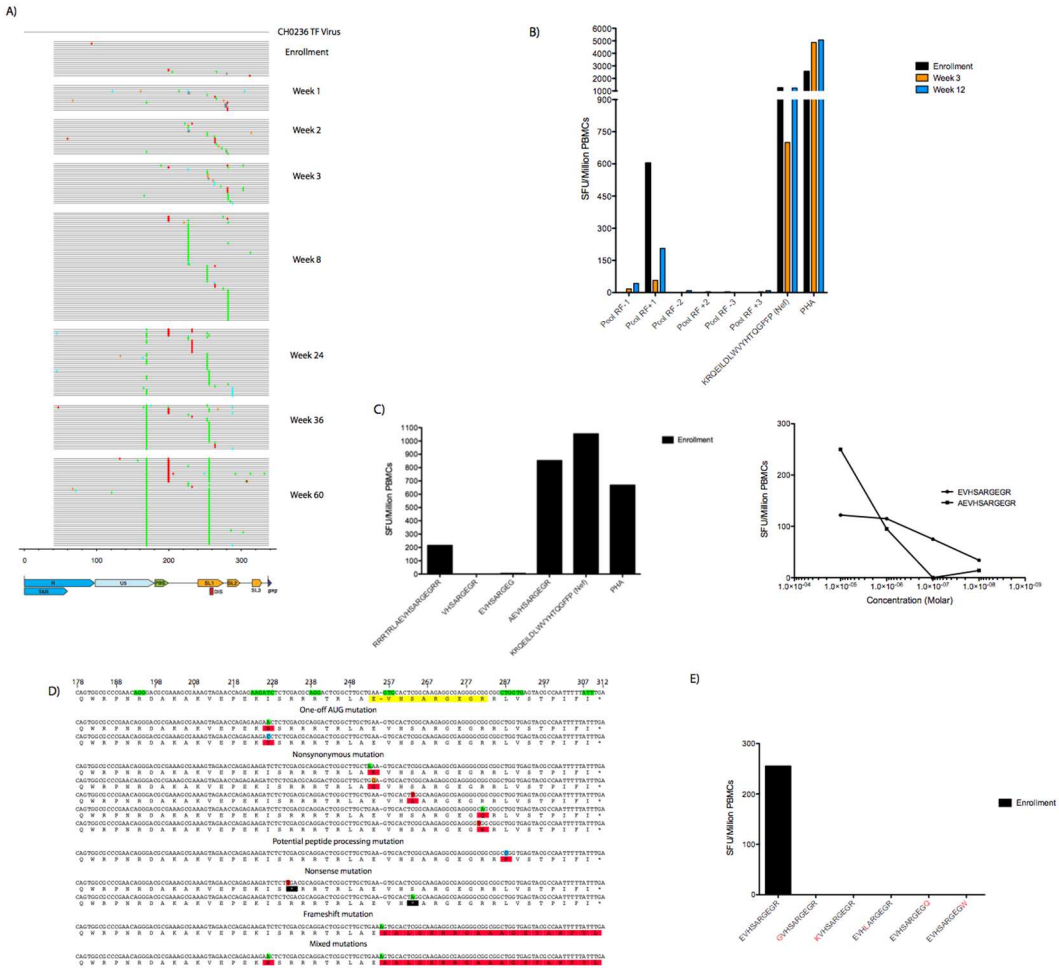


Figure 2-6 Immune recognition of a 5' leader DRiP during CH0236 infection

A) Nucleotide highlighter plot of 5' leader sequences taken from enrollment through week 60. Color code same as Figure 2.1A. Landmarks of the 5' leader are indicated at the bottom. B) Interferon-γ ELISPOT data using pooled peptides from the six reading frames of the 5' leader (+1, +2, +3, -1, -2, -3), an autologous Nef peptide and PHA. Reactivity above 50 SFUs was seen for the RF +1 pooled peptides. C) mapping of the 5' leader epitope within RF+1 peptides to RR18 and the AR11 (left). Comparison of AR11 and ER10 at 4 peptide concentrations demonstrates that ER10 is the minimal epitope. D) Nucleotide alignment translated into the ER10 RF demonstrating changes to the ER10 reading frame due to mutations observed in (A). Variants that were present in 3 or more sequences were included. ER10 ORF spans from RNA position 178-312. One-off AUGs highlighted in green, ER10 highlighted in yellow. Nucleotide mutations are highlighted with the same color code as (A). Nonsynonymous changes are highlighted in red at the amino acid level. E) IFN-γ ELISPOT assays of ER10 and variants observed over time demonstrates complete abrogation of immune recognition.

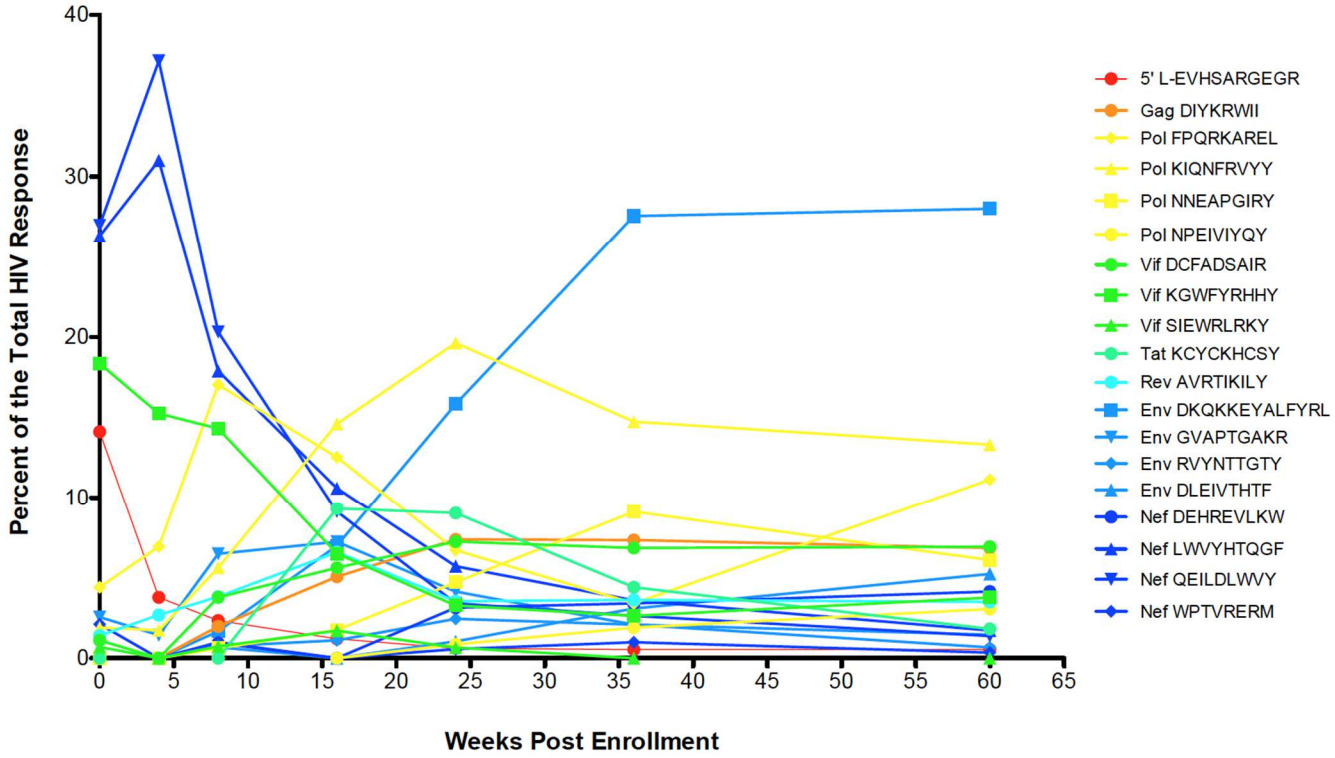


Figure 2-7 Kinetics of the autologous HIV-specific responses in CH0236

For the 19 autologous TF peptides recognized during the first year of infection (right), the percent of the total HIV-specific response is plotted over time (calculated as #SFU/total SFUs for that timepoint).

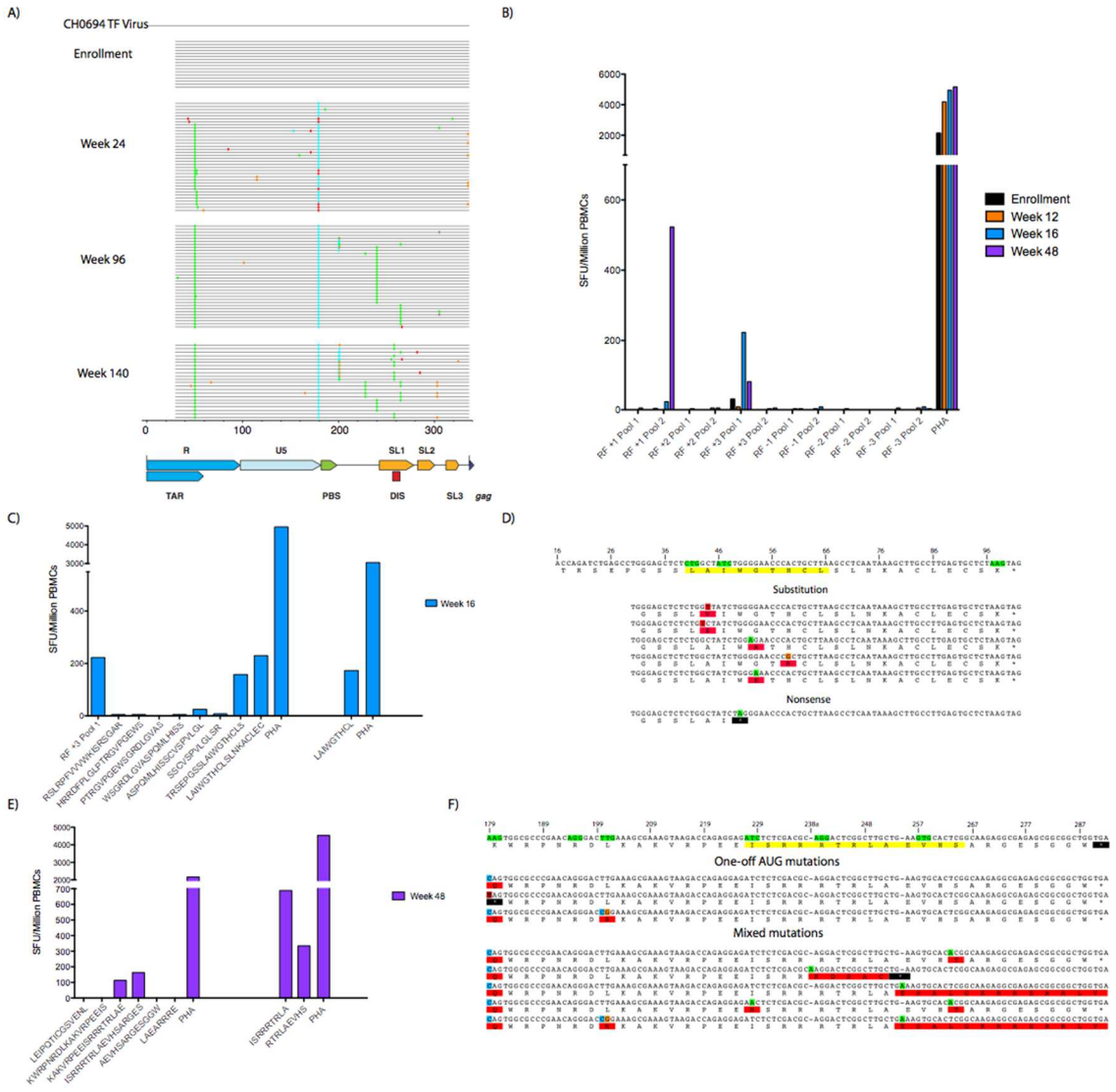


Figure 2-8 Immune recognition of the CH0694 5' leader DRiP during infection

A) Nucleotide highlighter plot of longitudinal 5' leader sequences, same as Figure 2.7A. B) IFN-γ ELISPOT assays using pooled peptides from all 6 5' leader reading frames demonstrate reactivity toward RF+1 and RF+3 pools. C) Deconvolution of the RF+3 pool using week 16 cells reveals the responsible peptide, LL9. D) Mutations mapped to the LL9 reading frame confer nonsynonymous changes or nonsense mutations. E) Deconvolution of the RF+1 pool from (B) reveals immune recognition of two partially overlapping peptides IA9 and RS9. F) Mutations mapped to the IA9/RS9 RF. Note – included variants were present in 3 or more sequences. Multiple variants with mutation at RNA positions 200+201 were present in the sequence dataset but did not appear 3 or more times.

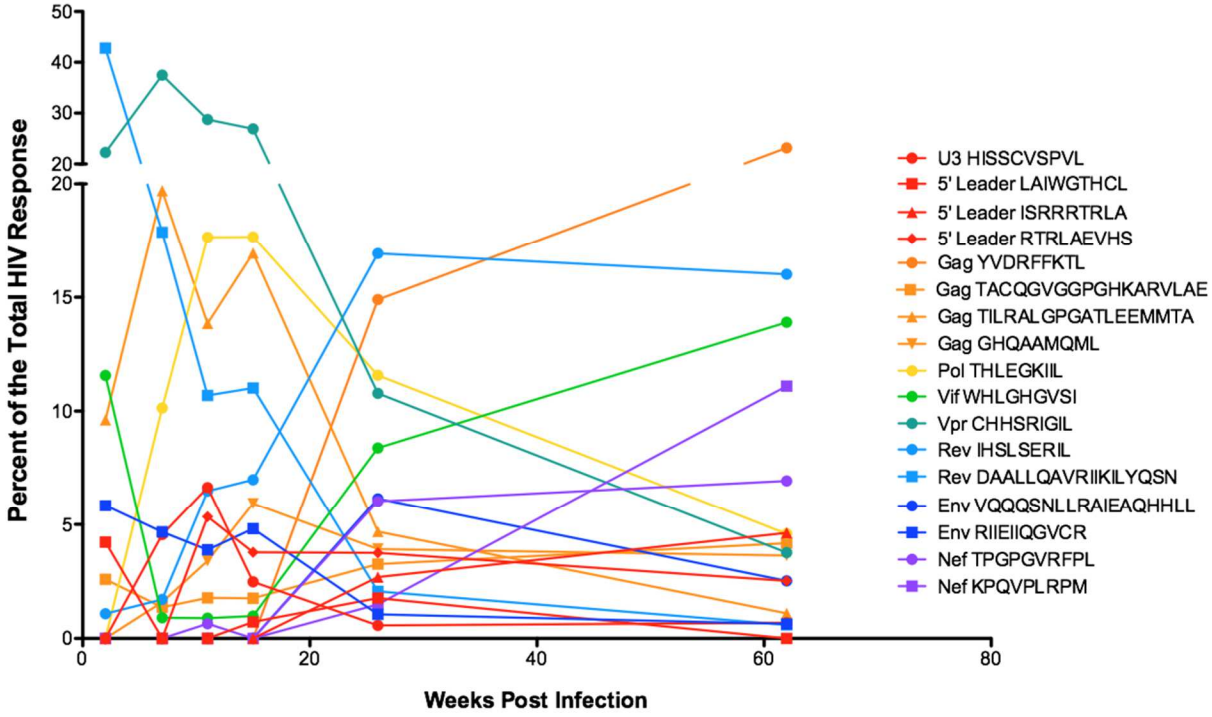


Figure 2-9 Kinetics of the autologous HIV-specific responses in CH0694

For the 17 autologous TF peptides recognized during the first year of infection (right), the percent of the total HIV-specific response is plotted over time (calculated as #SFU/total SFUs for that timepoint).

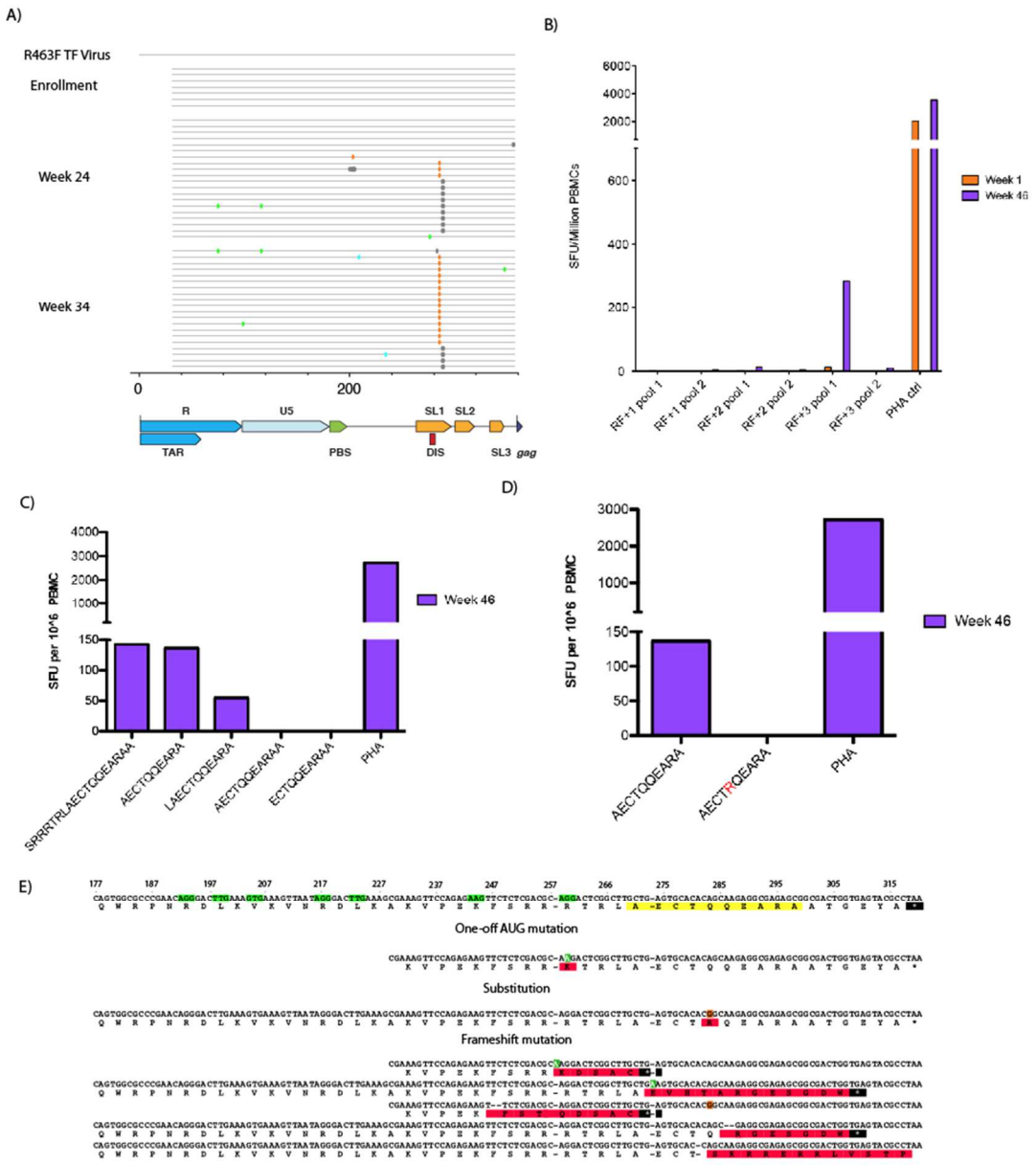


Figure 2-10 Immune recognition of a 5' leader DRiP during R463F infection

A) Nucleotide highlighter plot of longitudinal 5' leader sequences, same as Figure 2.7A. B) IFN-γ ELISPOT assays using pooled peptides from all 6 5' leader reading frames demonstrate reactivity toward RF+3 pool. C) Deconvolution of the RF+3 pool using week 46 cells reveals the responsible peptide, AA10. D) Immune recognition of AA10, but not variant AA10 observed at week 24 and beyond. E) Mutations mapped to the AA10 RF. Note – because of limited sample availability, sequences from the original near-full length sequencing were included here. The 5' end of these sequences is 228 due to primer placement.

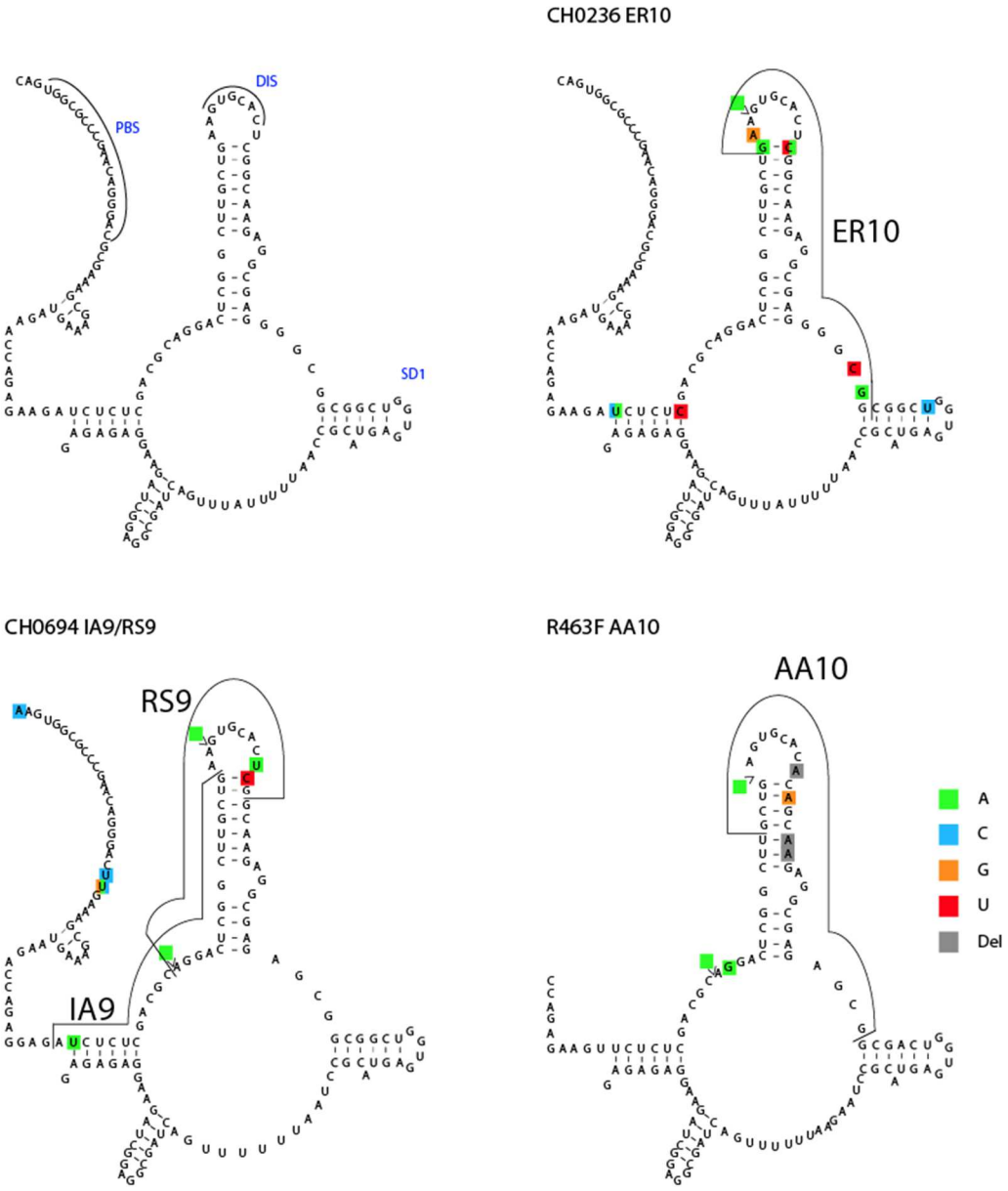


Figure 2-11 Location of 5' leader mutations on the RNA secondary structure, DIS ORF

3 subjects targeted peptides encoded by stem loop 1, also known as the DIS stem loop. Shown is the RNA structure of this region as determined by SHAPE (Pollom, et al., 2013) with the primer binding site (PBS), dimerization initiation site (DIS), and major splice donor (SD1) indicated. The TF sequence from each of these three subjects was substituted for the sequence used to determine the RNA folding. Watson-Crick base pairing indicated with black bars, G-U base pairing in grey. Highlighted residues demonstrated mutation during follow up (key bottom right). Carrots indicate insertions.

Due to a 20 nt insertion, the R463F does not have homology to other viruses immediately following the PBS. This region was therefore omitted.

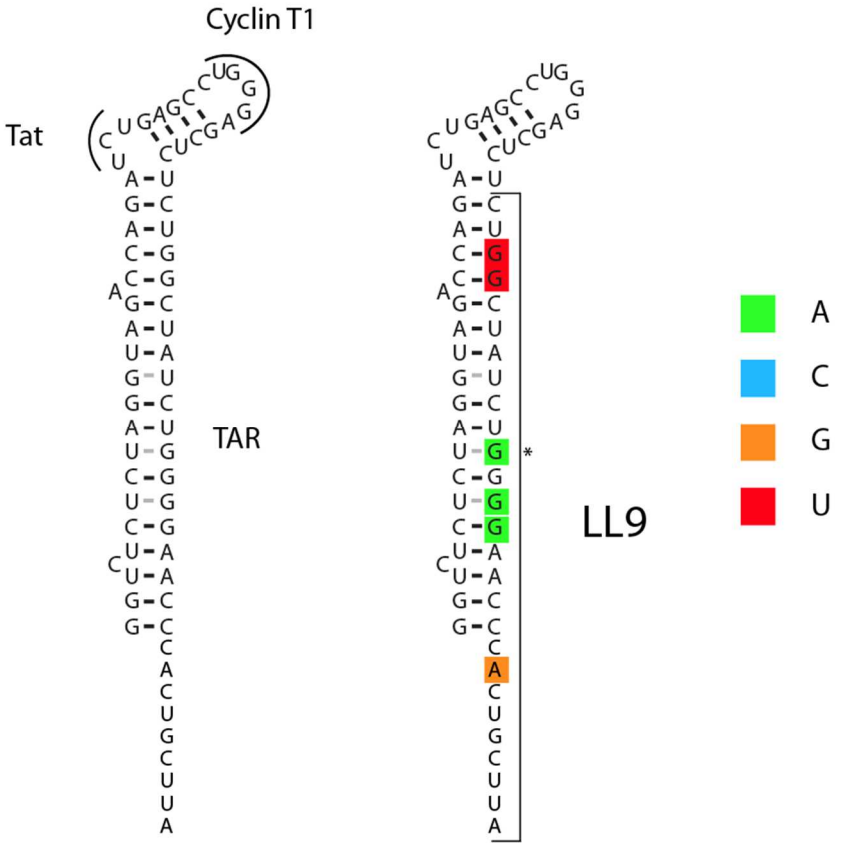


Figure 2-12 Location of the 5' leader mutations on the RNA secondary structure, R ORF

CH0694 targeted LL9, encoded by the TAR stem loop. Mutations are indicated as in Figure 2.11. The asterisk indicates the selected variant, which converts a G-U base pair to a Watson-Crick base pair.

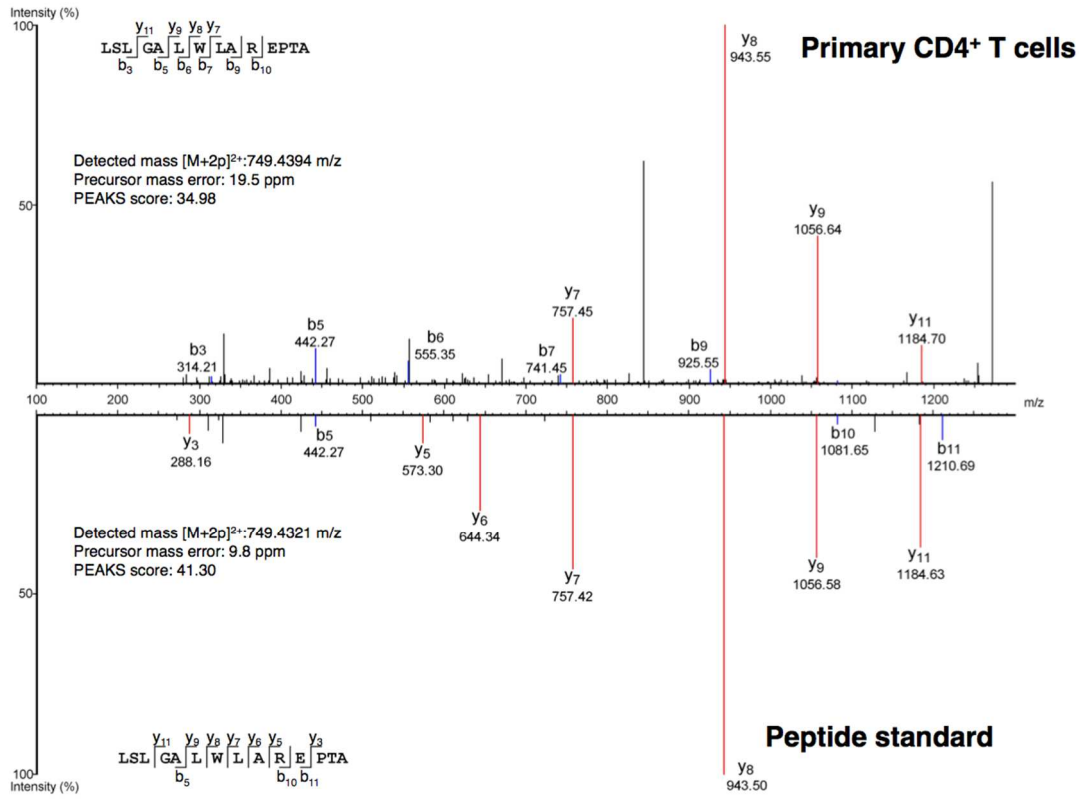


Figure 2-13 Comparison of experimenta LSLGALWLAREPTA fragmentation spectra with synthetic peptide standard

For each spectrum, the detected mass over charge ratio $[M+np]^n$ (where n is the number of charges), the PEAKS software peptide score and the measured precursor mass error (in parts per million) are shown. The detected fragment ions in each spectrum are displayed in the peptide sequence where b and y represent N- and C-terminal fragment ions, respectively.

CHAPTER 3

REVERSE TRANSCRIPTASE OVEREXTENSION GENERATES A ONE-OFF AUG INITIATION CODON IN THE HIV-1 5' LEADER

Edward F. Kreider[1], Emily K. Schutsky[1], Hui Li[1], Ranjit Warrier[1], Gerald H. Learn[1], Yingying Li[1], Peter T. Hraber[2], Rahul M. Kohli[1], George M. Shaw[1], Beatrice H. Hahn[1]

[1]Departments of Medicine and Microbiology, Perelman School of Medicine at the University of Pennsylvania

[2]Theoretical Biology and Biophysics Group, Los Alamos National Laboratory

Contributions: The experiments in this study were designed, executed, and analyzed by myself under the supervision of Beatrice H. Hahn, George M. Shaw, and Rahul M. Kohli. Emily K. Schutsky helped perform extension assay experiments and edited the paper. Hui Li, Ranjit Warrier, Gerald H. Learn, Peter T Hraber, and Yingying Li provided technical assistance. I wrote the paper with assistance from Beatrice H. Hahn

All figures appear at the end of this chapter

Section 3.1 – Abstract

Introduction: HIV-1 5' leader sequences exhibit multiple patterns of sequence evolution in infected subjects over time. One pattern was recently shown to reflect escape from CTL responses targeting 5' leader peptides expressed from one-off AUG codons (*e.g.* CUG). Here, we investigated another pattern within a motif called the primer overextension sequence (POS). *In vitro*, HIV-1 reverse transcriptase (RT) periodically overextends while copying the tRNA template by 4 nucleotides, incorporating TTGA into the POS.

Results: We identified 11 subjects who were infected with single transmitted founder (TF) viruses that did not match the canonical TTGA motif at 1 or 2 sites (non-overextension variants, GCGA, CCGA, CAGA, or CTGA). Variants that matched the TTGA motif (overextension variants) were subsequently observed in 10 of these subjects and shown to increase in frequency over time. Interestingly, only 52% of M group sequences matched the tRNA complement within the POS, leaving nearly half of HIV-1 strains with a one or two base pair mismatch relative to the tRNA. In an *in vitro* extension assay, HIV-1 RT initiated polymerization despite a two-base pair mismatch within the POS with 82% efficiency. We then identified 8 additional subjects whose TF viruses matched the canonical TTGA motif. In one subject, the POS mutated away from TTGA in concert with downstream cellular immune escape mutations in a newly identified 5' leader open reading frame. In this case, different mutations in the POS deleted a one-off AUG codon (bold, **TTG**A-> CGGA, CCGA, CAGA, TGGA, TCGA) and thus likely abrogated the expression of the 5' leader peptide.

Conclusions: Sequence analysis of HIV-1 quasispecies in infected subjects over time indicates that RT overextension incorporates a one-off AUG initiation codon within the HIV-1 5' leader. Cellular immune responses provide an antagonizing force that selects for POS mutations that delete this one-off AUG. This finding provides new insight into peptide expression from the HIV-1 5' leader and may inform future interventions that target this region of the genome.

Section 3.2 – Introduction

The HIV-1 5' leader is often cited as the most conserved region of the viral genome (*1-4*). Nonetheless, we recently observed multiple patterns of non-random mutation within the 5' leaders of HIV-1 and SIVmac strains that were sequenced over time (Chapter 2). One of these patterns – the toggling of mutations in a 27-62 nucleotide span – resembled cytotoxic T lymphocyte (CTL) escape that is typically observed within viral coding sequences (*5*, *6*). This finding led to the hypothesis that the 5' leader, a designated "untranslated" region, expressed peptides that could serve as CTL antigens. Using interferon-γ ELISPOT assays, we confirmed this hypothesis and demonstrated host CTL recognition of multiple HIV-1 5' leader encoded peptides (Chapter 2). Ribosomal profiling studies revealed that translation of these peptides is initiated at one-off AUG-like start codons (one-off AUGs, *e.g.* CUG, Chapter 2, Warrier, R., Shaw, G., unpublished) (*7-9*). Thus, based on virus sequence evolution, we were able to identify a novel class of HIV-1 Defective Ribosomal Products (DRiPs) that are expressed and presented in the context of major histocompatibility complexes, and may serve as early indicators of pathogen invasion (*10-13*).

In this study, we sought to identify other causes of 5' leader mutation and to correlate these to viral variants observed *in vivo*. Reverse transcriptase (RT) overextension offered one such candidate (*14*). HIV-1 reverse transcription is primed by an uncharged host tRNA(Lys,3), which anneals to the viral genome at the primer binding site (PBS) (*15*, *16*). During plus strand DNA synthesis, RT copies 18 nucleotides of this

primer, replicating the PBS (*14*, *17-20*). *In vitro* studies have suggested that RT pauses elongation upon encountering the plus strand strong stop signal, a modified adenosine at tRNA position 58 (m^1A-58) (*14*, *17*). This DNA intermediate must then undergo strand transfer, anneal to the minus strand via the PBS, and prime RT extension. *In vitro*, RT periodically overextends tRNA replication beyond the strong stop, yielding DNA intermediates that harbor 4 or 21 additional nucleotides of the host tRNA sequence (*21*, *22*). Muthuswami and colleagues demonstrated that the 4-nt overextended intermediate, but not the 21-nt overextended one, could undergo strand transfer and prime RT extension. Further, they found that >60% of HIV-1 strains matched the host tRNA complement in the four nts adjacent to the PBS and named this region the primer overextension sequence (POS, Figure 3.1). Based on their analysis, however, certain HIV-1 strains would not match the tRNA complement at position 1 and/or 2 of the POS (non-overextension variants). Strand transfer during reverse transcription of one of these non-overextension variants would result in donor-acceptor strand mismatches (Figure 3.1), which may either mutate the virus (*19*, *20*) or decrease RT extension efficiency (*23*, *24*). To test for evidence of RT overextension *in vivo*, we applied a previously developed experimental strategy based on single genome amplification of HIV-1 RNA and direct amplicon sequencing (single genome sequencing, SGS) that can be used to infer the genetic identity of the transmitted founder virus (*25*) and characterize molecular pathways of virus evolution (*26*, *27*). We applied this tool to evaluate POS sequence evolution in subjects who had been longitudinally followed from acute infection.

Section 3.3 – Results

We studied 5' leader sequence evolution in a cohort of 19 subjects who were followed starting from acute HIV-1 infection over 0.5-3 years. Single genome sequences from 12 of these subjects have been previously published (*26*, *28*). Poisson fitter analysis of 5' half genome sequences from the earliest time point demonstrated that each subjects was infected with a single transmitted founder (TF) virus (data not shown) (*25*, *26*, *28*, *29*). TF strains from 8 subjects matched the tRNA complement within the POS (TTGA, overextension variants), whereas 11 did not (non-overextension variants).

To test for mutations consistent with overextension, we evaluated longitudinal HIV-1 5' leader sequences from the 11 subjects who were infected by non-overextension variants. As shown in Table 1, these TF viruses demonstrated polymorphism at POS positions 1 and 2: four viruses encoded CTGA, one CAGA, one CCGA, and five GCGA. Sequences that encoded TTGA (overextension variants) were detected at subsequent time points during infection in 10 out of 11 of these subjects. The prevalence of these overextension variants increased over time: from 5.6% of sequences at enrollment to 25% at week 60 in CH0236; from 11% at enrollment to 25% at week 85 in CH077; and from 10% at enrollment to 40% at week 24 in CH1432 (Figure 3.2). Thus, examination of longitudinal HIV-1 5' leader sequences revealed mutations consistent with RT overextension *in vivo*.

To assess the diversity in POS motifs from globally circulating HIV-1 strains, we obtained 979 HIV-1 group M sequences from the Los Alamos National Laboratory HIV Database that included this region (HXB2 positions 635-659). Consistent with previous

findings (*22*), group M sequences showed conservation (>97% identity) throughout the PBS and at positions 3 and 4 of the POS (Figure 3.3). Positions 1 and 2 of the POS, however, were polymorphic. Only 512/979 sequences (52%) matched the host tRNA throughout the POS (TTGA, Figure 3.4). 197/979 (20%) harbored a single base pair mismatch and 234/979 (24%) harbored two base pair mismatches. These findings demonstrate that polymorphism within the POS is primarily at positions 1 and 2.

To test if these mismatches between the tRNA and the POS affected RT extension efficiency, we developed an *in vitro* extension assay. As outlined in Figure 3.5, fluorescein-labeled primers 18-22 nucleotides in length were annealed to a 31-nt template and incubated with purified HIV-1 reverse transcriptase. RT extension generated a 31-nt product (upper band) that could be visualized on a gel and quantified using densitometry. Three fluorescein-labeled primers were tested: an 18-nt primer representing the canonical plus strand DNA intermediate (PBS); a 20-nt primer that contained two additional Ts at its 3' end (PBS+2); and a 22-nt primer that represented the overextended intermediate (PBS+4, Figure 3.3). These primers were duplexed with one of two 31-nt templates: one that was complementary to TTGA in the POS (match, overextension variants) and one that was complementary to GCGA (mismatch, non-overextension variant). The PBS+2-mismatch template duplex was included as a control, since a two-base pair mismatch at the 3' end should confer a decreased extension efficiency as long as conditions are not saturating (*23*, *24*). The PBS+4/mismatch duplex (lane 9) primed RT extension with 82% extension efficiency as compared to duplexes with complete complementarity (lanes 2, 3, 5, and 8), which showed efficiencies ranging from 74% to 94% (Figure 3.5). Conversely,

a two base pair mismatch at the 3' hydroxyl group decreased extension efficiency to 40% (lane 6), suggesting that these conditions were not saturating. Along with the *in vivo* evidence of RT overextension, these data suggest that RT can efficiently initiate polymerization despite a two-base pair mismatch within the POS.

Based on the frequency of POS mutations observed in TF and group M viruses, we hypothesized that mutations away from TTGA would also occur during infection. To test this hypothesis, we examined sequence evolution in the 8 subjects infected with overextension variants (TTGA). Evaluation of 5' leader sequences in these subjects revealed three patterns of evolution (Table 2). Three of the eight subjects retained TTGA within the POS throughout the duration of follow up. Sequences from four of the five remaining subjects demonstrated the same T-to-C transitions at position 1 of the POS (HXB2 position 654, Figure 3.6). This CTGA variant is found in 13.5% of group M sequences (Figure 3.6) and more commonly in non-primate lentiviruses (*22*). Finally, viral sequences from subject CH0694 demonstrated substitutions at both positions 1 and 2 of the POS.

As described in Chapter 2, longitudinally collected 5' leader sequences from subject CH0694 accumulated non-random mutations during the course of his infection. We confirmed that CH0694 cytotoxic T lymphocytes recognized two autologous 5' leader peptides (IA9 and RS9, HXB2 positions 681-719) using interferon-γ ELISPOT assays (Chapter 2). Translation of these peptides occurs at one-off AUG-like codons (one-off AUGs), which can be mutated during virus escape from 5' leader peptide directed cellular immune responses (Chapter 2). Six one-off AUGs were observed

throughout the IA9/RS9 reading frame (HXB2 positions 633, 648, 654, 681, 693, and 711). Four of these codons were upstream IA9/RS9, including a UUG that fell within the POS (**UUG**A, green highlighting, Figure 3.7). The POS was mutated in 5 of 34 (14.7%) day 695 variants and 11 of 27 (40.7%) day 1009 variants. All but one of these mutations deleted the embedded one-off AUG. The only one-off AUG upstream of IA9 and RS9 that remained unaltered during immune escape fell within the invariant PBS (HXB2 position 648). We hypothesize that any mutations within this codon would rapidly revert to match the PBS, similar to what has been observed with other PBS substitutions (*19*). Thus, in the context of immune escape, one-off AUG codons within the POS can be deleted.

Section 3.4 – Conclusions

In this study, we evaluated intra-host sequence evolution of a 5' leader motif called the primer overextension sequence (POS) and correlated these observations with *in vitro* functional studies of the reverse transcriptase. We observed mutations consistent with RT overextension in subjects infected with non-overextension variants, providing evidence that this process is not simply an *in vitro* phenomenon (Table 3.1). Moreover, overextension variants gradually increased in population frequency (Figure 3.2), suggesting that canonical plus strand intermediates are generated more commonly than their overextended counterparts (ratio of ~3.3:1 *in vitro*) (*14*). Finally, we tested the effect of a two base-pair mismatch within the POS on RT priming and observed little to no loss in extension efficiency (Figure 3.5). These data indicate that RT overextension occurs *in vivo*, even when the viral POS does not match the tRNA.

Next, we evaluated sequence evolution in subjects infected with overextension variants and observed two different types of mutations. Viral sequences from four subjects exhibited a T-to-C transition at the first position of the POS (Table 2 and Figure 3.6). This finding is consistent with the observation that RT favors transition over transversion mutations (*30*) and that non-human lentiviruses often harbor CTGA within the POS (*14*), but seemingly contradicts *in vitro* studies that have shown that m^1A-58 is most commonly decoded as an adenosine (*17*). Future biochemical studies of RT base misincorporations while decoding modified bases may reconcile these differences. Finally, we observed mutations that deleted a one-off AUG codon embedded in the POS

(**U**UG**G**A), suggesting that this change confers immune escape from a host T cell response (Figure 3.7, Chapter 2).

Collectively, the data presented here suggest that RT overextension occurs *in vivo* and incorporates a non-canonical translation initiation codon in the 5' leader sequence. This one-off AUG can initiate expression of a downstream peptide, which in the case of subject CH0694 became the target of a CTL response. This link between RT overextension and 5' leader peptide expression is relevant to the development of HIV vaccines targeting 5' leader peptides. One concern for developing such a vaccine is that HIV-1 strains can delete one-off AUG codons upstream of targeted peptides. The data presented here indicates that certain mutated one-off AUGs can be restored by processes like RT overextension. Interestingly, another one-off AUG is encoded by the PBS (Figure 3.7 AGG at position 648), which we hypothesize will not tolerate mutation (*19*, *20*). Thus, the findings presented add to 5' leader biology and inform interventions that target peptides encoded by this region of the genome.

Section 3.5 – Methods

Single Genome Sequencing and Sequence Analysis. RNA extraction, complementary DNA synthesis, and single genome amplification were performed as previously described (*29*). Sequences were then aligned and analyzed using Geneious 7 software. Accession numbers for previously published sequences are as follows (fill in): CH0040: , CH0058: , CH0077: , CH0131: , CH0159: , CH0198: , CH0256: , CH0185: , CH0470: , CH0162: , CH0164: , CH0042: . M group sequences were obtained from the Los Alamos Compendium and all sequences that did not include the 18 nucleotides of the PBS and 6 nucleotides downstream were excluded.

RT extension assay. In each condition, FAM-labeled primer was annealed to excess, unlabeled template (1:1.5 ratio) by heating to 95C for 5 minutes and then slowly cooling to room temperature. A master mix of 10 mM dNTPs, 1x RT reaction buffer, and 0.7 units of RT (ThermoFisher) was assembled and aliquoted into reaction tubes. Extension reactions were initiated with 100nM of the primer:template duplexes, and were incubated at 37C for 2 minutes. The reactions were quenched with formamide buffer with 10 mM EDTA and were heated at 95C for 20 minutes. Samples were run on a preheated 20% acrylamide-TBE/Urea gel at 50C for 30 minutes. Gels were imaged using a Typhoon 9410 (Amersham Biosciences). Intensities of unextended and extended primer bands were calculated using Bio-Rad QuantityOne software after subtracting background signal, and percent extensions were calculated by dividing the intensity of the extended band by the total intensity of both unextended and extended bands.

Section 3.6 – Competing interests, Contributions, and Acknowledgments

The authors have no conflicts of interest to report.

Authors' contributions

EFK conceived of the study, designed and carried out experiments, and drafted the manuscript. EKS helped devise and carry out *in vitro* biochemical assays. HL, RW, GHL, and PTH participated in sequence alignment and analysis. RMK devised the *in vitro* biochemical assays. GMS and BHH helped conceive of the study and draft the manuscript.

Acknowledgements

EFK is supported by a NIH NIAID F30 fellowship (F30 AI112426). We want to thank Hannah Barbian and Elizabeth Traxler for reviewing the manuscript.

Section 3.7 - References

1. N. van Bel, A. T. Das, B. Berkhout, In Vivo SELEX of Single-Stranded Domains in the HIV-1 Leader RNA. *Journal of Virology*. **88**, 1870–1880 (2014).

2. S. C. Keane *et al.*, RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*. **348**, 917–921 (2015).

3. K. Lu *et al.*, NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*. **334**, 242–245 (2011).

4. E. Pollom *et al.*, Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. *PLoS Pathog.* **9**, e1003294 (2013).

5. N. Goonetilleke *et al.*, The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J. Exp. Med.* **206**, 1253–1272 (2009).

6. J. F. Salazar-Gonzalez *et al.*, Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* **206**, 1273–1289 (2009).

7. S. R. Starck *et al.*, Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science*. **336**, 1719–1723 (2012).

8. N. Stern-Ginossar *et al.*, Decoding human cytomegalovirus. *Science*. **338**, 1088–

1093 (2012).

9. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. **324**, 218–223 (2009).

10. B. P. Dolan, J. R. Bennink, J. W. Yewdell, Translating DRiPs: progress in understanding viral and cellular sources of MHC class I peptide ligands. *Cell. Mol. Life Sci.* **68**, 1481–1489 (2011).

11. L. C. Antón, J. W. Yewdell, Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors. *J Leukoc Biol*. **95**, 551–562 (2014).

12. E. A. Reits, J. C. Vos, M. Gromme, J. Neefjes, The major substrates for TAP in vivo are derived from newly synthesized proteins. *Nature*. **404**, 774–778 (2000).

13. U. Schubert *et al.*, Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature*. **404**, 770–774 (2000).

14. R. Muthuswami *et al.*, The HIV plus-strand transfer reaction: determination of replication-competent intermediates and identification of a novel lentiviral element, the primer over-extension sequence. *Journal of Molecular Biology*. **315**, 311–323 (2002).

15. L. Kleiman, tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life*. **53**, 107–114 (2002).

16. A. T. Das, S. KOKEN, B. ESSINK, J. VANWAMEL, B. Berkhout, Human-Immunodeficiency-Virus Uses Trna(Lys,3) as Primer for Reverse Transcription in Hela-Cd4(+) Cells. *FEBS Lett.* **341**, 49–53 (1994).

17. S. Auxilien, G. Keith, S. F. Le Grice, J. L. Darlix, Role of post-transcriptional modifications of primer tRNALys,3 in the fidelity and efficacy of plus strand DNA transfer during HIV-1 reverse transcription. *J Biol Chem*. **274**, 4412–4420 (1999).

18. P. Charneau, M. Alizon, F. Clavel, A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *Journal of Virology*. **66**, 2814–2820 (1992).

19. C. M. Fennessey *et al.*, Generation and characterization of a SIVmac239 clone corrected at four suboptimal nucleotides. *Retrovirology*. **12**, 49 (2015).

20. B. Berkhout, A. T. Das, On the primer binding site mutation that appears and disappears during HIV and SIV replication. *Retrovirology*. **12**, 75 (2015).

21. S. Auxilien, Role of Post-transcriptional Modifications of Primer tRNALys,3 in the Fidelity and Efficacy of Plus Strand DNA Transfer during HIV-1 Reverse Transcription. *Journal of Biological Chemistry*. **274**, 4412–4420 (1999).

22. R. Muthuswami *et al.*, The HIV plus-strand transfer reaction: determination of replication-competent intermediates and identification of a novel lentiviral element, the primer over-extension sequence. *Journal of Molecular Biology*. **315**, 311–323 (2002).

23. L. Diaz, J. J. DeStefano, Strand transfer is enhanced by mismatched nucleotides at the 3' primer terminus: a possible link between HIV reverse transcriptase fidelity and recombination. *Nucleic Acids Res.* **24**, 3086–3092 (1996).

24. H. Yu, M. F. Goodman, Comparison of HIV-1 and avian myeloblastosis virus reverse transcriptase fidelity on RNA and DNA templates. *J Biol Chem*. **267**, 10888–10896 (1992).

25. B. F. Keele *et al.*, Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 7552–7557 (2008).

26. J. F. Salazar-Gonzalez *et al.*, Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J. Exp. Med.* **206**, 1273–89. PMC2715054. (2009).

27. K. J. Bar *et al.*, Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. *PLoS Pathog.* **8**, e1002721 (2012).

28. M. K. Liu *et al.*, Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. *J. Clin. Invest.* **123**, 380–93. PMC3533301. (2013).

29. J. F. Salazar-Gonzalez *et al.*, Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. *Journal of Virology*. **82**, 3952–3970 (2008).

30. M. E. Abram, A. L. Ferris, W. Shao, W. G. Alvord, S. H. Hughes, Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *Journal of Virology*. **84**, 9864–9878 (2010).

Section 3.8 – Figures and Tables

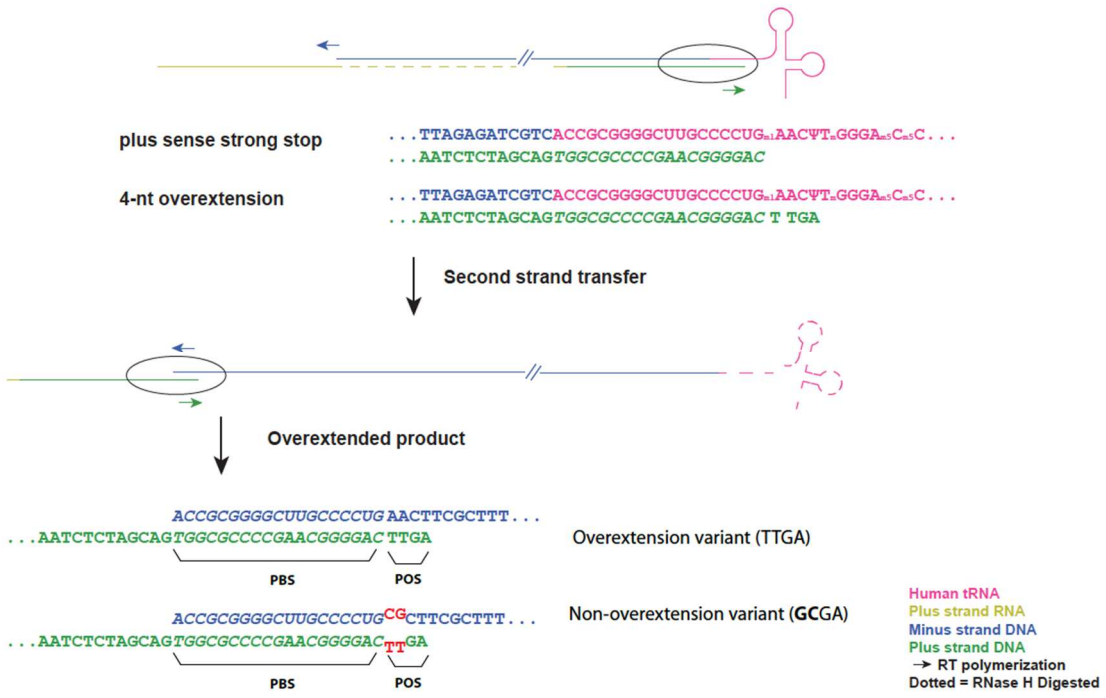


Figure 3-1 Graphical representation of reverse transcriptase overextension and strand transfer

During minus strand synthesis (blue arrow), RNA in RNA:DNA heteroduplexes is digested by reverse transcriptase's RNase H activity (dotted gold line). Short RNA sequences called the polypurine tracts are RNase-resistant and can prime plus strand synthesis (green line). During canonical plus strand (green) synthesis, 18 nucleotides of the host tRNA (pink) are reverse transcribed prior to strong stop, generating the plus sense primer binding sequence (PBS, italics). During overextension, 4 additional nucleotides—TTGA—are added to the nascent strand. After RNase digestion of the host tRNA, plus strand DNA intermediates must undergo strand transfer and anneal to the minus strand (bottom). Viruses can either match (overextension variant) or mismatch (non-overextension variant) the four overextended tRNA encoded nucleotides.

Table 3-1 Percent Overextension Variants

Subject ID	POS motif at Transmission	Total Number of Sequences	Percent Overextension Variants
CH0058	**C**TGA	36	8.3%
CH0131	**C**TGA	17	12%
CH0159	**C**TGA	16	6.3%
CH0164	**C**TGA	54	57%
CH0264	**CA**GA	136	2.9%
CH0752	**CC**GA	26	12%
CH0077	**GC**GA	36	11%
CH0185	**GC**GA	52	0.0%
CH0236	**GC**GA	367	8.2%
CH0470	**GC**GA	95	6.3%
CH1432	**GC**GA	20	25%

Percent single genome sequences that are overextension variants in longitudinally followed subjects infected with a single transmitted founder virus. Transmitted founder (TF) viruses from 11 subjects did not match TTGA (column 2). Between 16 and 367 sequences were available per subject (column 3). The percent of these sequences that were TTGA variants, or overextension variants, is indicated in column 4.

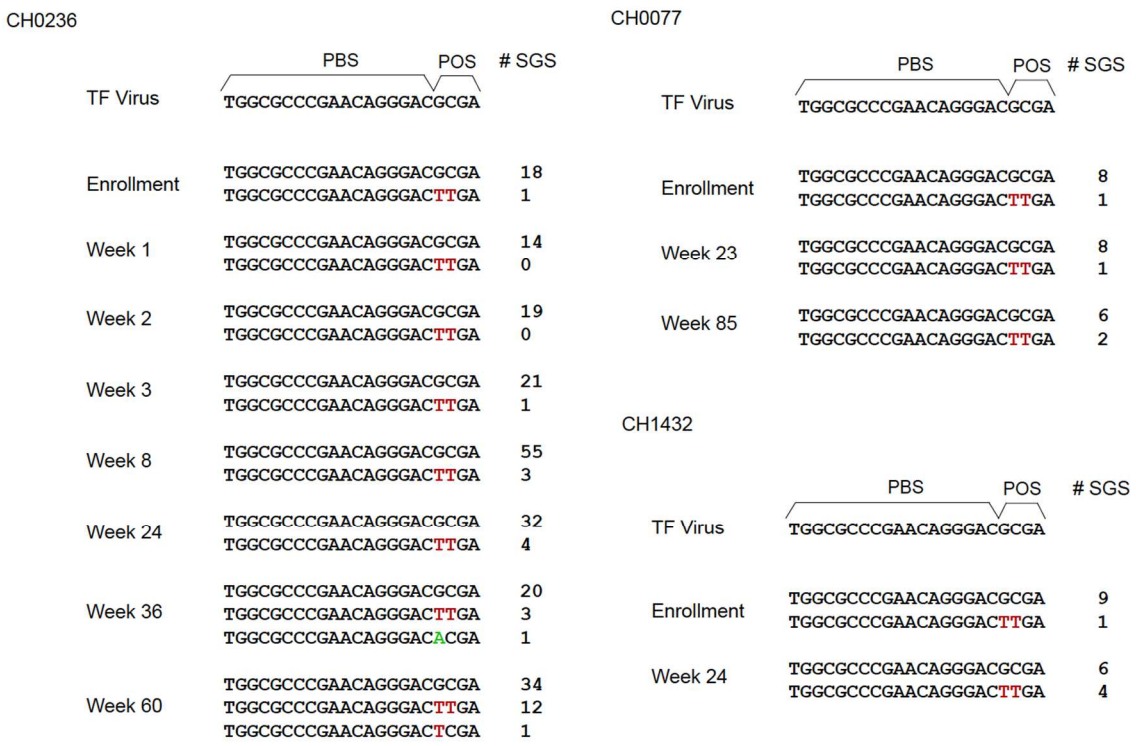


Figure 3-2 Accumulation of overextension variants in viral sequences from subjects CH0236, CH0077, and CH1432

CH0236, CH0077, and CH1432 were infected with non-overextension variants (GCGA). Depicted are nucleotide alignments of the PBS and POS variants observed over time, organized by timepoint. The number of single genome sequences (SGS) that were TF-matched non-overextension variants (top), overextension-variants (second), or neither (bottom) are indicated to the right. Sequences changes relative to the TF virus are highlighted.

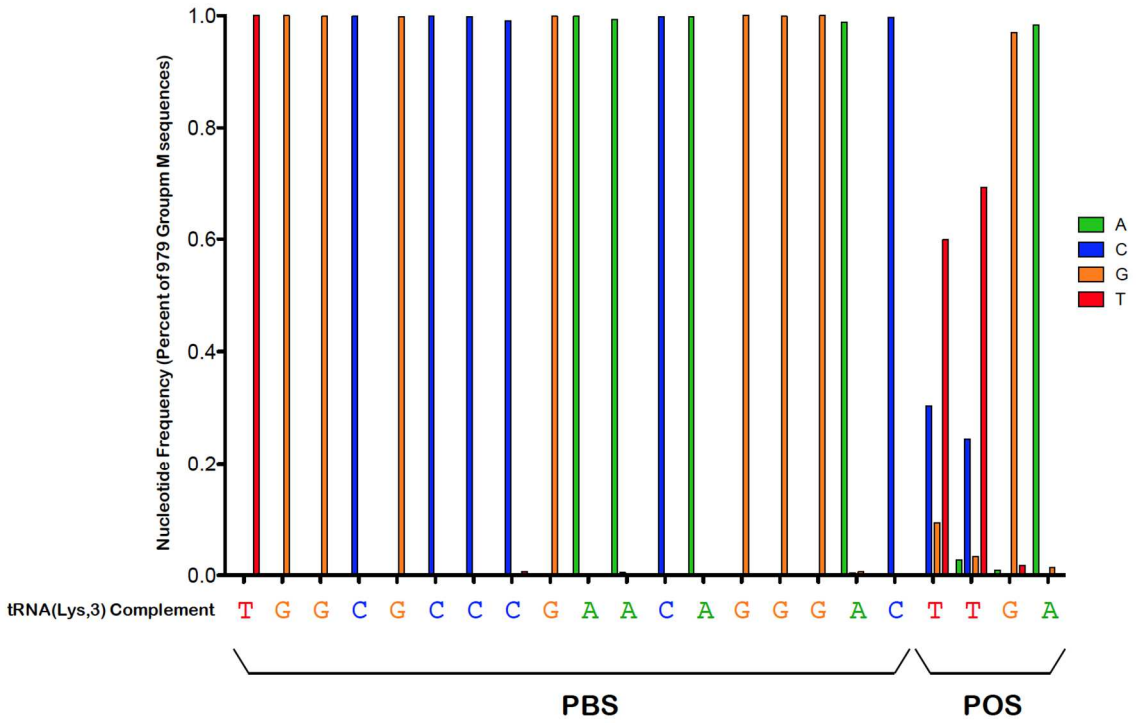


Figure 3-3 Nucleotide conservation of the PBS and POS relative to the reverse complement of tRNA(Lys,3)

Nucleotide frequency was plotted for each position with the color code indicated to the right. The complement to the tRNA(Lys,3) is shown at the bottom. >99% conservation is observed at sites throughout the PBS and >97% conservation is seen at positions 3 and 4 of the POS. Positions 1 and 2 of the POS exhibit polymorphism, with only 59.9% and 69.4% of sequences matching the tRNA complement, respectively.

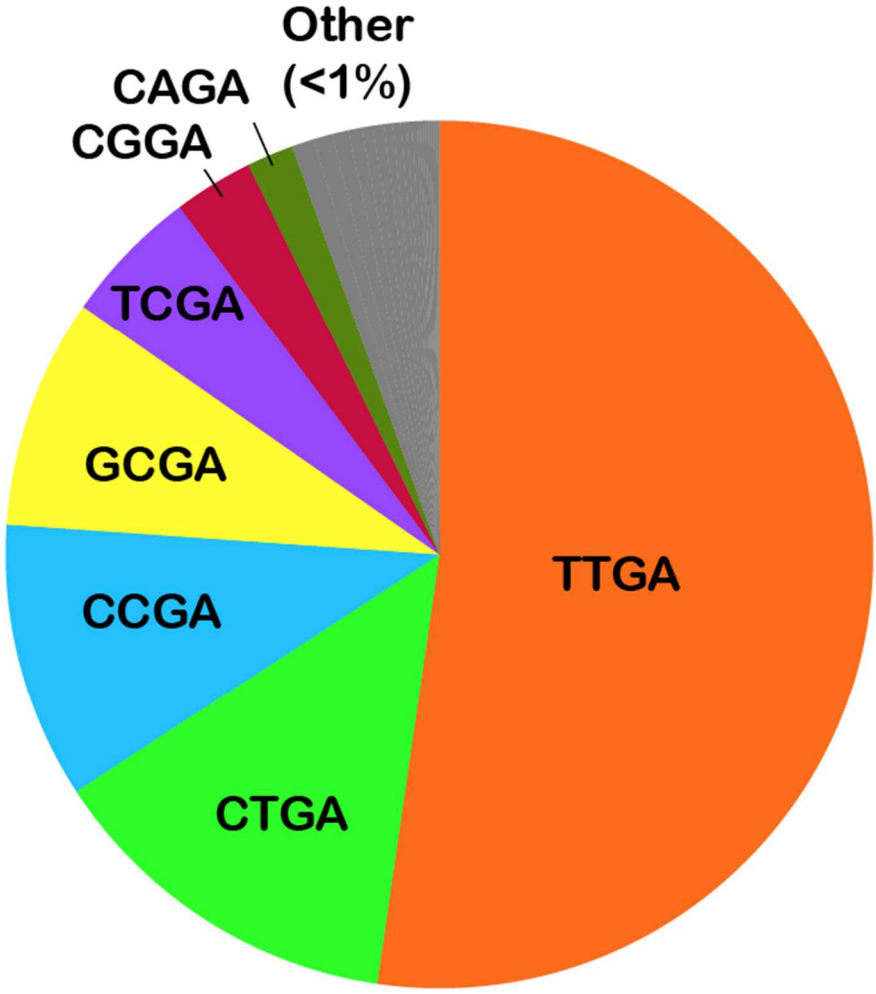


Figure 3-4 POS motifs present in published HIV-1 sequences

Relative abundance of each 4-nt motif is depicted in a pie chart. TTGA, the complement of the tRNA(Lys,3) within the POS, is only observed in 52.3% of sequences.

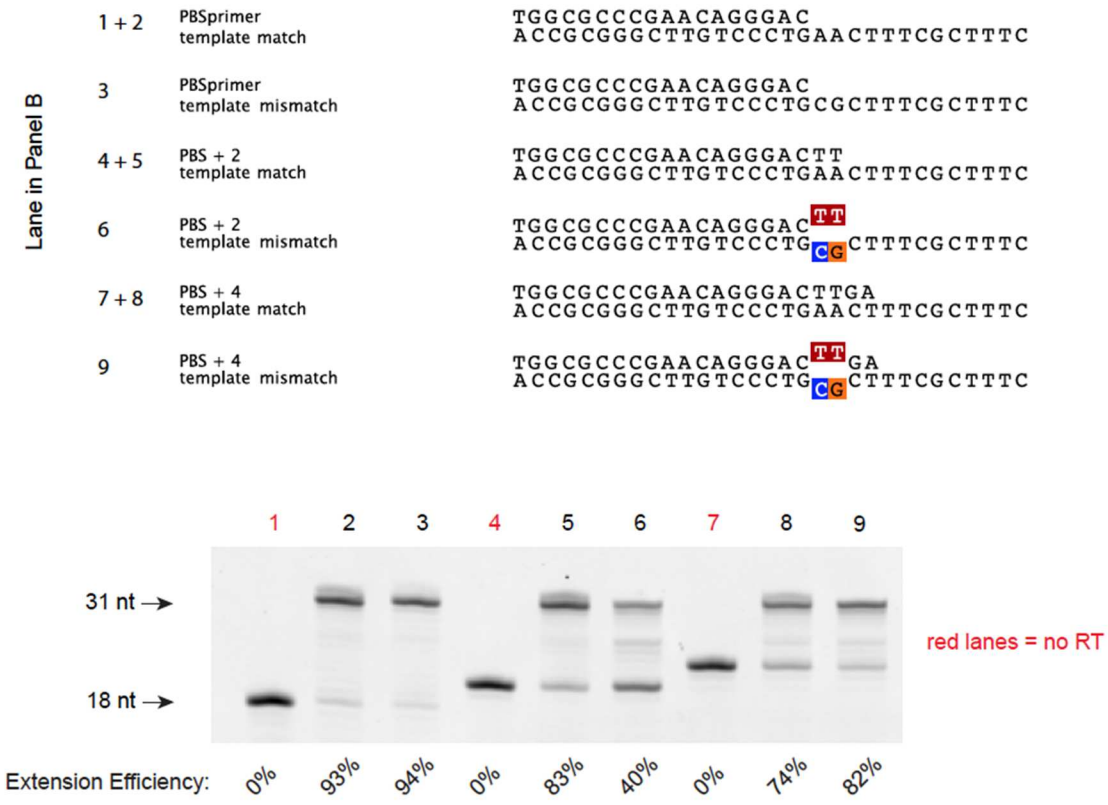


Figure 3-5 Effect of primer overextension 2-base pair mismatch on RT polymerization efficiency

To test the effect of base pair mismatch during primer overextension of non-overextension variants, we developed an *in vitro* RT extension assay. Primer template pairs that were tested are represented graphically (top). PBS+4-template mismatch duplexes (lane 9) recapitulate the mispairing that would be observed during primer overextension of non-overextension variants. RT polymerization results in extension of the primer (length: 18-22 nts) to 31 nts. No RT was added to lanes 1, 4, and 7.

Table 3-2 Percent SGS mutated away from TTGA

Subject ID	POS motif at Transmission	Total Number of Sequences	Percent Mutated
CH0325	TTGA	35	0.0%
CH0042	TTGA	122	0.0%
CH0569	TTGA	57	0.0%
CH0198	TTGA	34	2.9%
CH0040	TTGA	55	5.5%
CH0850	TTGA	135	1.5%
CH0256	TTGA	109	13.8%
CH0694	TTGA	166	9.6%

Percent single genome sequences mutated away from TTGA in longitudinally followed subjects infected with overextension variants

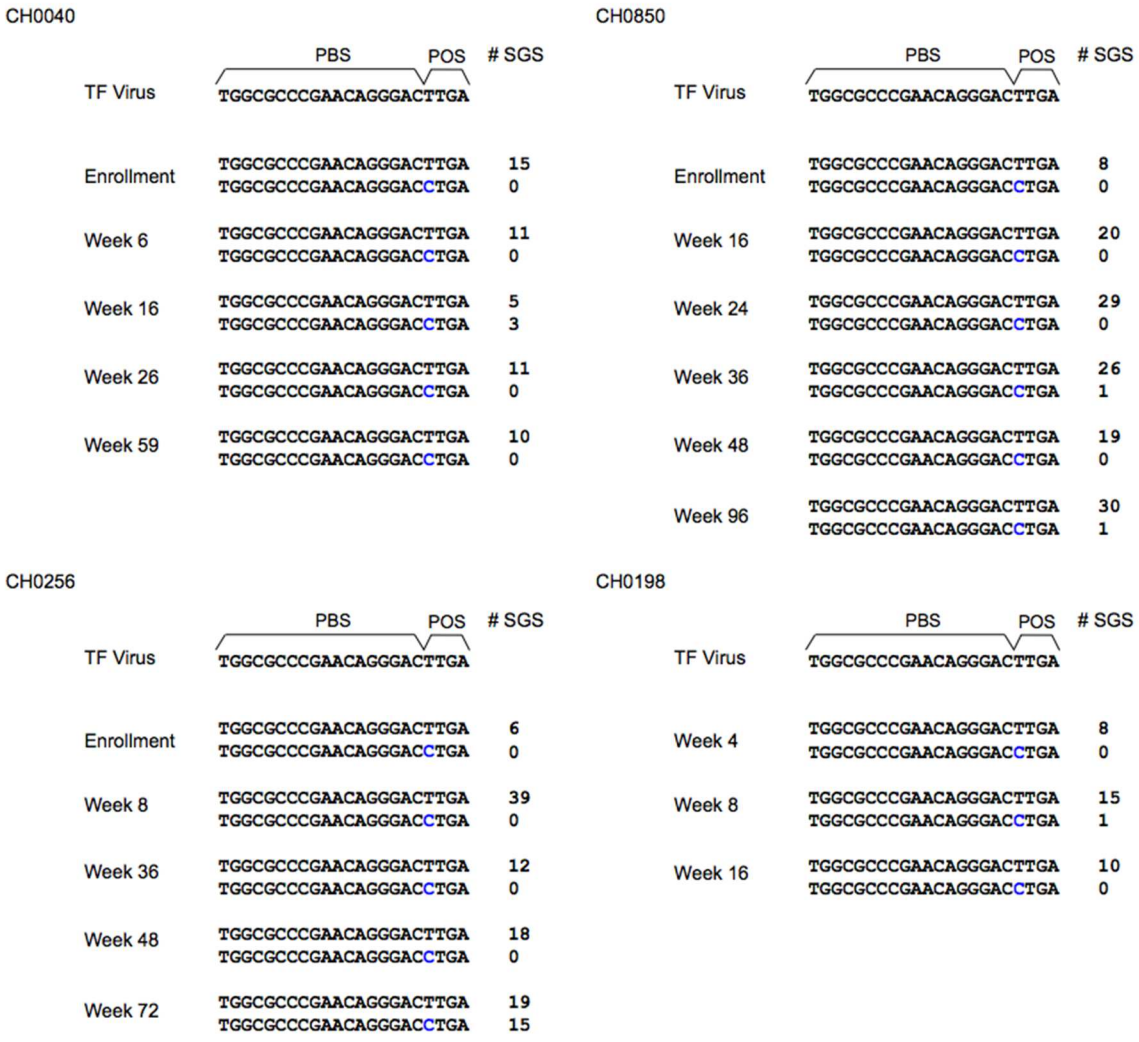


Figure 3-6 T-to-C transition at POS position 1 in subjects infected with TTGA variants

CH0040, CH0850, CH0256, and CH0198 were each infected with a TTGA POS variant. During infection, CTGA variants were detected in all subjects.

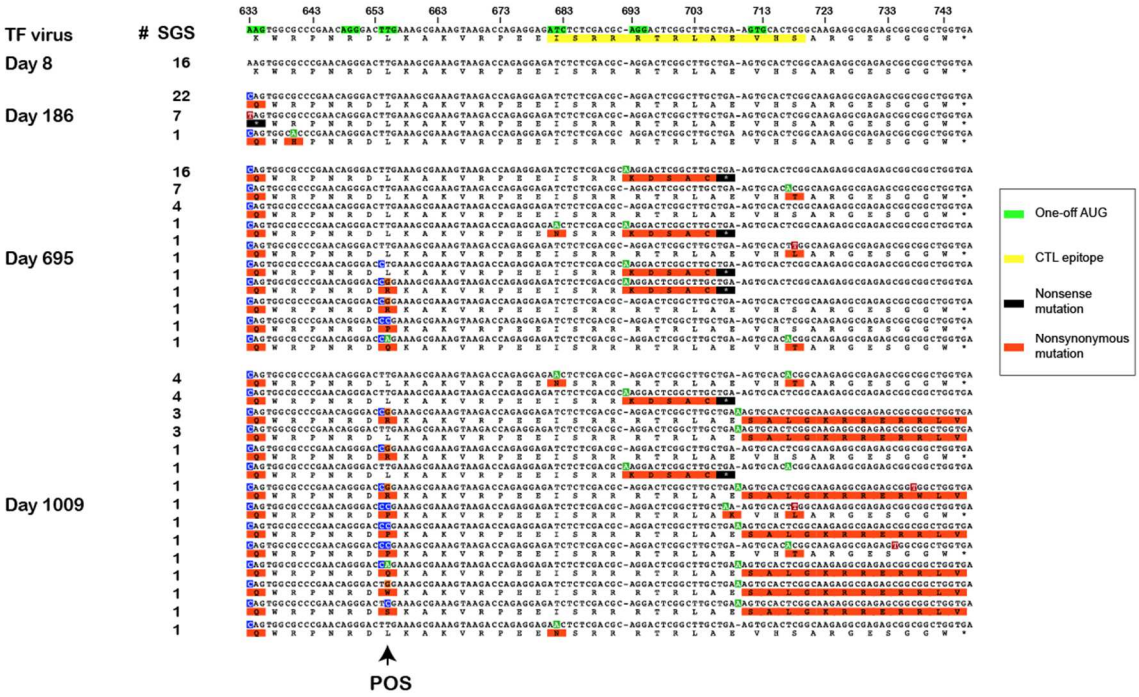


Figure 3-7 Mutation in the POS in concert with downstream CTL escape

Depicted is a longitudinal nucleotide/amino acid alignment in which identical sequences from a given timepoint are collapsed. Changes relative to the inferred transmitted founder variant (top) are highlighted both at the nucleotide and amino acid levels (key, right). Timepoint and number of identical single genome sequences (SGS) are indicated to the left. Numbering is relative to HXB2. Subject CH0694 T cells recognized two overlapping 5' leader peptides, IA9 and RS9, as measured by interferon-γ ELISPOT assays (Chapter 2, position 681-719, yellow box). Translation initiation for such 5' leader peptides occurs at one-off AUG-like codons, *e.g.* CUG or UUG (all one-off AUGs highlighted in green boxes). Changes within the POS at days 695 and 1009 mutated the embedded the one-off AUG (**TTG**A).

CHAPTER 4

STAGED INDUCTION OF HIV-1 GLYCAN-DEPENDENT BROADLY NEUTRALIZING ANTIBODIES

M. Bonsignori[1,2]*†, **E. F. Kreider[3]**†, D. Fera[4]†, R. R. Meyerhoff[1,2]†, T. Bradley[1,2]†, K. Wiehe[1,2], S. M. Alam[1,2], B. Aussedat[5], W. E. Walkowicz[5], K. K. Hwang[2], K. O. Saunders[2,6], R. Zhang[2], M. A. Gladden[2], A. Monroe[2], A. Kumar[2], S. M. Xia[2], M. Cooper[2], M. K. Louder,[7] K. McKee,[7] R. T. Bailer,[7] B. W. Pier,[4] C. A. Jette[4], G. Kelsoe [2,8], W. B. Williams[1,2], L. Morris[9], J. Kappes[10], K. Wagh[11], G. Kamanga[12]‡, M. S. Cohen[13], P. T. Hraber[11] ,D. C. Montefiori[2,6], A. Trama[2], H. X. Liao[1,2], T. B. Kepler[14], M. A. Moody[1,2,15], F. Gao[1,2], S. J. Danishefsky[5], J. R. Mascola[7], G. M. Shaw[3], B. H. Hahn[3], S. C. Harrison[4], B. T. Korber[10]*, B. F. Haynes[1,2]*

[1]Department of Medicine, Duke University School of Medicine, Duke University Medical Center, Durham, NC.
[2]Duke Human Vaccine Institute, Durham, NC.
[3]Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.
[4]Laboratory of Molecular Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA.
[5]Department of Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, New York, NY.
[6]Department of Surgery, Duke University School of Medicine, Duke University Medical Center, Durham, NC.
[7]Vaccine Research Center, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD.
[8]Department of Immunology, Duke University School of Medicine, Duke University Medical Center, Durham, NC.
[9]National Institute for Communicable Diseases, Johannesburg 2131, South Africa.
[10]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL.
[11]Los Alamos National Laboratory, Los Alamos, NM.
[12]University of North Carolina Project, Kamuzu Central Hospital, Lilongwe, Malawi.
[13]Departments of Medicine, Epidemiology and Microbiology and Immunology, University of North Carolina, Chapel Hill, NC.
[14]Departments of Microbiology, and Mathematics and Statistics, Boston University, Boston, MA.
[15]Department of Pediatrics, Duke University School of Medicine, Duke University Medical Center, Durham, NC.

* Correspondence to:

Mattia Bonsignori (mattia.bonsignori@dm.duke.edu)
Duke Human Vaccine Institute
DUMC 103020
2 Genome Court
Durham, NC 27710
Ph: 919-681-9739
Fax: +919-684-5230

Bette Korber (btk@lanl.gov)
Los Alamos National Laboratory

Los Alamos, NM 87506
Ph. 505-665-4453
Fax. 505-665-3493

Barton F. Haynes, M.D. (barton.haynes@dm.duke.edu)
Duke Human Vaccine Institute
DUMC 10320
2 Genome Court
Durham, NC 27710
Ph: 919-684-5279
Fx: 919-684-5230

† These authors contributed equally.

‡ Current address: FHI 360, Lilongwe, Malawi

One sentence summary: Stages of V3-glycan neutralizing antibody development are identified that explain the long duration required for their development.

Contribution

Under the supervision of BHH, BTK, Gerald H. Learn, and Yingying Li I performed virus sequence analysis and correlation of neutralization to virus evolution. I constructed various figures and wrote/edited the text.

All figures and supplemental materials appear at the end of this chapter

Section 4.1 – Abstract

A glycosylated site adjacent to the third variable loop (V3) of the HIV-1 envelope protein is an epitope for broadly neutralizing antibodies (bnAbs) and a desirable vaccine target. We identify here key events in the ontogeny of a V3-glycan bnAb. Two autologous neutralizing antibody lineages selected for virus escape mutations and consequently allowed initiation and affinity maturation of a V3-glycan bnAb lineage. The nucleotide substitution required to initiate the bnAb linage occurred at a low probability site for activation-induced cytidine deaminase activity. Cooperation of B-cell lineages and an improbable mutation critical for bnAb activity define the necessary events leading to V3-glycan bnAb development, explain why initiation of V3-glycan bnAbs is rare, and suggest an immunization strategy for inducing V3-glycan bnAbs.

Section 4.2 – Introduction

A vaccine to prevent HIV-1 infection should include immunogens that can induce broadly neutralizing antibodies (bnAbs) (*1, 2*). Of the five major targets for bnAbs, the glycan-rich apex of the HIV-1 envelope (Env) trimer and the base of the third variable loop (V3) are distinguished by the potency of antibodies directed against them (*3-8*). Although these antibodies have less breadth than those directed against the CD4 binding site (CD4bs) or the gp41 membrane-proximal region (MPER), one current goal of vaccine development is to elicit them in combination with other bnAb specificities to achieve broad coverage of transmitted/founder (TF) viruses (*1, 2*).

Mapping the co-evolution of virus and antibody lineages over time informs vaccine design by defining the succession of HIV-1 Env variants that evolve *in vivo* during the course of bnAb development (*9-11*). Antibody lineages with overlapping specificities can influence each other's affinity maturation by selecting for synergistic or antagonistic escape mutations: an example of such "cooperating" lineages is provided by two CD4bs-directed bnAbs that we characterized previously (*11, 12*). Thus, cooperating antibody lineages and their viral escape mutants allow identification of the specific Envs, among the myriads present in the infected individual, that stimulate bnAb development and that we wish to mimic in a vaccine.

Here we describe the co-evolution of an HIV-1 Env quasispecies and a memory B-cell lineage of gp120 V3-glycan directed bnAbs. To follow virus evolution, we sequenced

~1,200 HIV-1 *env* genes sampled over a 5 year period; to follow the antibody response, we identified natural heavy- and light-chain pairs of six antibodies from a bnAb lineage, designated DH270, and augmented this lineage by next generation sequencing (NGS). Structural studies defined the position of the DH270 Fab on gp140. We also found two B-cell lineages (DH272 and DH475) with neutralization patterns that likely selected for observed viral escape variants, which in turn stimulated the DH270 lineage to potent neutralization breadth. We found a mutation in the DH270 heavy chain that occurred early in affinity maturation at a disfavored activation-induced cytidine deaminase (AID) site and that was necessary for bnAb lineage initiation. This improbable mutation can explain the long period of antigenic stimulation needed for initial expansion of the bnAb B-cell lineage in this individual.

Section 4.3 – Results

Three N332 V3-glycan dependent antibody lineages

We studied an African individual (CH848) followed from the time of infection to development of plasma neutralization breadth 3.5 years later. Reduced plasma neutralization of N332A Env-mutated HIV-1 pseudoviruses and plasma neutralization fingerprinting demonstrated the presence of N332-sensitive broadly neutralizing antibodies (bnAbs) (table 4.S1)(*13*). To identify these antibodies, we screened memory B cells from weeks 205, 232, and 234 post-infection using memory B cell cultures (*14*) and antigen-specific sorting (*15*) and found three N332-sensitive lineages, designated DH270, DH272 and DH475. Their genealogy was augmented by NGS of memory B-cell cDNA from seven time points spanning week 11 to week 240 post-transmission.

DH270 antibodies were recovered from memory B cells at all three sampling times (weeks 205, 232, and 234) and expansion of the clone did not occur until week 186 (Fig. 4.1A and table 4.S2). Clonal expansion was concurrent with development of plasma neutralization breadth (table 4.S3), and members of the DH270 lineage also displayed neutralization breadth (Fig. 4.1B). The most potent DH270 lineage bnAb (DH270.6) was isolated using a fluorophore-labeled Man9 V3 glycopeptide that is a mimic of the V3-glycan bnAb epitope (*16*). The lineage derived from a V_H1~2*02 rearrangement that produced a CDRH3 of 20 amino acid residues paired with a light chain encoded by Vλ2~23 (fig. 4.S1). Neutralization assays and competition with V3-glycan bnAbs PGT125 and PGT128 confirmed lineage N332-dependence (fig. 4.S2).

The DH475 mAb was recovered from memory B cells at week 232 post-transmission by antigen-specific sorting using the fluorophore-labeled Man9-V3 glycopeptide (*16*). The earliest DH475 lineage V_HDJ_H rearrangements were identified with NGS at week 64 post-transmission (fig. 4.S3A and table 4.S2). Its heavy chain came from V_H3-23*01 (V_H mutation frequency = 10.1%) paired with a Vλ4-69*02 light chain (fig. 4.S3B).

The DH272 mAb came from cultured memory B cells obtained at week 205 post-transmission. DH272 lineage V_HDJ_H rearrangements were detected as early as 19 weeks post-transmission by NGS (fig 4.S3A and table 4.S2). The DH272 heavy chain used V_H1-2*02, as did DH270, but it paired with a Vκ 2-30 light chain. Its CDRH3 was 17 amino acids long; V_H mutation was 14.9%. DH272, an IgA isotype, had a 6-nt deletion in FRH3 (fig. 4.S3B).

For both DH272 and DH475 lineages, binding to CH848 TF Env gp120 depended on the N332 potential N-linked glycosylation (PNG) site (fig. 4.S3C). DH272 binding also depended on the N301 PNG site (fig. 4.S3C). Neither lineage had neutralization breadth (fig. 4.S3D).

Evolution of the CH848 virus quasispecies

We sequenced 1,223 HIV-1 3'-half single-genomes from virus in plasma collected at 26 time points over 246 weeks. Analysis of sequences from the earliest plasma sample indicated that CH848 had been infected with a single, subtype clade C founder virus, ~ 17 (CI 14-19) days prior to screening (figs. 4.S4 and 4.S5). By week 51 post-infection,

91% of the sequences had acquired an identical, 10-residue deletion in variable loop 1, a region that includes the PGT128-proximal residues 133-135 and 141 (figs. 4.S6 and 4.S7). Further changes accrued during the ensuing four years, including additional insertions and deletions (indels) in V1, mutations in the ^{324}GDIR^{327} motif within the V3 loop, deletion or shifting of N-linked glycosylation sites at positions 301 and 322, and mutations at PGT128-proximal positions in V1, V3, and C4, but none of these escape variants went to fixation during 4.5 years of follow-up (figs. 4.S6-S9 and Supplemental Text).

Ontogeny of DH270 lineage and acquisition of neutralization breadth

As with other V3-glycan bnAbs, viral neutralization clade specificity and intra-clade breadth of DH270 depended primarily on the frequency of the N332 glycosylation site within the relevant clade (Fig. 4.2A and Supplemental Text). Heterologous breadth and potency of DH270 lineage antibodies increased with accumulation of V_H mutations and although DH270.UCA did not neutralize heterologous HIV-1, five amino-acid substitutions in DH270.IA.4 (four in the heavy chain, one in the light chain) were sufficient to initiate the bnAb lineage and confer heterologous neutralization (Fig. 4.2B,C).

The capacity of the early DH270 lineage members to neutralize heterologous viruses correlated with the presence of short V1 loops (Fig. 4.2D). As the lineage evolved, it gained capacity to neutralize viruses with longer V1 loops, although with reduced potency (Fig. 4.2D and fig. 4.S10A-C). Neutralization of the same virus panel by V3

glycan bnAbs 10-1074, PGT121 and PGT128 followed the same inverse correlation between potency and V1 length (fig. 4.S10D-F).

Mutations in the DH270 antibody lineage that initiated heterologous neutralization

The likelihood of AID-generated somatic mutation has strong nucleotide-sequence dependence (*17, 18*). Moreover, we have recently shown for CD4bs bnAbs that V_H sites of high intrinsic mutability indeed determine many sites of somatic hypermutation (*11*). Like the VRC01-class CD4bs bnAbs, both DH270 and DH272 used V_H1~2*02 - although unlike the CD4bs bnAbs, V3 glycan bnAbs in general can use quite disparate V_H gene segments (*3, 19-23*) - and antibodies in both lineages have mutations at the same amino acid positions that correspond to sites of intrinsic mutability that we identified in the V_H1-2*02 CD4bs bnAbs (*11*) (fig. 4.S11A).

Presence of the canonical V_H1-2*02 allele in individual CH848 was confirmed by genomic DNA sequencing (fig. 4.S11B). Four nucleotide changes in the DH270 UCA conferred heterologous neutralization activity to the next intermediate antibody (IA4). The G92A and G102A nucleotide mutations in DH270.IA4 (and in DH272) occurred at "canonical" AID hotspots (DGYW); G164C (G164A for DH272) was in a "non-canonical" AID hotspot with a comparable level of mutability (*17*) (Fig. 4.3A). These high probability mutations encoded amino acid substitutions G31D, M34I and S55T (N for DH272). G31D and M31I became fixed in their respective lineages and S55T eventually became prevalent also in the DH272 lineage (Fig. 4.3B). In contrast, the G169C mutation in DH270.IA4, which encoded the G57R amino acid mutation, occurred

at a site with a very low predicted level of mutability (*17*) and generated a canonical cold spot (SYC) (Fig. 4.3A). Thus, once the rare G169C mutation had occurred, it had a high probability of being maintained, and indeed it was present in 523/758 (68%) DH270 lineage V_H sequences identified with NGS at week 186 post-transmission (Fig. 4.3C). The G169C mutation in DH270.IA4 also disrupted an overlapping AID hotspot at the adjacent guanine (170G). While there were no mutations in DH272 at G169, the overlapping AID hotspot at G170 did mutate, resulting in an initial G57V substitution (Fig. 4.3A,C). Thus, both the DH270 bnAb and DH272 autologous neutralizing lineages had mutations at Gly57, but the substitution in the DH270 lineage (G57R) was an improbable event whereas the substitution (G57V) in the DH272 lineage was much more likely. By week 111 post-transmission, all DH272 lineage V_HDJ_H transcripts sequenced by NGS harbored a mutation in the Gly57 codon, which resulted in the predominance of an encoded aspartic acid (Fig. 4.3B). In contrast, only 6/758 (0.8%) DH270 lineage transcripts isolated 186 weeks post-transmission had Val57 or Asp57; 48/758 (6.3%) retained Gly57, while over two-thirds, 514/758 (67.8%), had G57R (Fig. 4.3B).

Reversion of Arg57 to Gly abrogated DH270.IA4 neutralization of autologous and heterologous HIV-1 isolates (Fig. 4.3D). A DH270.IA4 R57V mutant, with the base change that would have occurred had the overlapping AID hotspot been used, also greatly reduced DH270.IA4 neutralization, confirming that Arg57, rather than the absence of Gly57 was responsible for the acquired neutralizing activity (Fig. 4.3D). Finally, the DH270.UCA G57R mutant neutralized both autologous and heterologous viruses, confirming that G57R alone could confer neutralizing activity on the DH270 germline

antibody (Fig. 4.3E). Thus, the improbable G169C mutation conferred reactivity against autologous virus and initiated acquisition of heterologous neutralization breadth in the DH270 lineage.

Autologous neutralizing antibody lineages that cooperated with DH270

Evidence for functional interaction among the three N332-dependent lineages came from the respective neutralization profiles against a panel of 90 autologous viruses from transmitted/founder to week 240 post-transmission (Fig. 4.4A). Both DH475 and DH272 neutralized autologous viruses isolated during the first year of infection that were resistant to most DH270 lineage antibodies (only DH270.IA1 and DH270.4 neutralized weakly) (Fig. 4.4A). DH475 neutralized viruses from week 15 through week 39 and DH272 neutralized the CH848 transmitted/founder and all viruses isolated up to week 51, when viruses that resisted DH475 and DH272 became strongly sensitive to all mature antibodies in the DH270 lineage (Fig. 4.4A).

The identification of specific mutations implicated in the switch of virus sensitivity was complicated by the high levels of mutations accumulated by virus Env over time (fig 4.S12). We identified virus signatures that defined the DH270.1 and DH272/DH475 immunotypes and introduced four of them, in various combinations, into the DH272/DH475-sensitive virus that was closest in sequence to the DH270.1-sensitive immunotype : a 10 amino-acid residue deletion in V1 (Δ134-143); a D185N mutation in V2, which introduced an N-linked glycosylation site; an N413Y mutation in V4, which disrupted an N-linked glycosylation site; and a 2 amino-acid residue deletion (Δ463-464)

in V5 (see Supplemental text). The large V1 deletion was critical for DH270.1 neutralization, with smaller contributions from the other changes; the V1 deletion increased virus resistance to DH475 (3.5-fold increase). V1-loop-mediated resistance to DH475 neutralization increased further when combined with the Δ463-464 V5 deletion (5-fold increase) (Fig. 4.4B). The V1 loop of the transmitted/founder virus (34 residues) was longer than the global mean of 28 (range 11 to 64) (*24*). As we found for heterologous neutralization, DH270 lineage antibodies acquired the ability to neutralize larger fractions of autologous viruses as maturation progressed by gaining activity for viruses with longer V1 loops, although at the expense of lower potency (fig. 4.S13A-C). This correlation was less clear for gp120 binding (fig. 4.S13D-F), however, suggesting that the V1 loop-length dependency of V3 glycan bnAb neutralization has a conformational component. Thus, DH475 cooperated with the DH270 bnAb lineage by selecting viral escape mutants sensitive to bnAb lineage members.

For DH272, the variants that we made did not implicate a specific cooperating escape mutation. The Δ134-143 (V1 deletion) mutated virus remained sensitive to DH272 neutralization; both combinations of the V1 deletion in our panel that were resistant to DH272 and sensitive to DH270.1 included D185N, which on its own also caused DH272 resistance but did not lead to DH270.1 sensitivity (Fig. 4.4C). Thus, we have suggestive, but not definitive, evidence that DH272 also participated in selecting escape mutants for the DH270 bnAb lineage.

Structure of DH270 lineage members

We determined crystal structures for the Fabs of DH270.UCA3, DH270.1, DH270.3, DH270.5 and DH270.6, as well as for DH272 (Fig. 4.5; table 4.S6). The variable domains of the DH270 antibodies all superposed well, indicating that antibody affinity maturation modulated the antibody-antigen interface without substantially changing the antibody conformation (Fig. 4.5A). Mutations accumulated at different positions for DH270 lineage bnAbs in distinct branches (fig. 4.S14), possibly accounting for their distinct neutralization properties. Because of uncertainty in the inferred sequence of the germline precursor (fig. 4.S15A,B), we also determined the structure of DH270.UCA1, which has a somewhat differently configured CDR H3 loop (fig. 4.S15C); reconfiguration of this loop during early affinity maturation could account for the observed increase (with respect to the UCA) in heterologous neutralization by several intermediates. DH272 had a CDRH3 configured differently from that of DH270 lineage members and a significantly longer CDRL1 (Fig. 4.5B), compatible with their distinct neutralization profiles.

We also compared the structures of DH270 lineage members with those of other N332-dependent bnAbs. All appear to have one long CDR loop that can extend through the network of glycans on the surface of the gp120 subunit and contact the "shielded" protein surface. The lateral surfaces of the Fab variable module can then interact with the reconfigured or displaced glycans to either side. PGT128 has a long CDRH2 (Fig. 4.5C), in which a 6-residue insertion is critical for neutralization breadth and potency (*5, 19*). PGT124 has a shorter and differently configured CDR H2 loop, but a long CDR H3 instead (Fig. 4.5D) (*25*).

Structure of the DH270 – HIV Env Complex

We determined a three-dimensional (3D) image reconstruction, from negative-stain electron microscopy (EM), of the DH270.1 Fab bound with a gp140 trimer (92Br SOSIP.664) (Fig. 4.5E,F and fig. 4.S15). The three DH270.1 Fabs project laterally, with their axes nearly normal to the threefold of gp140, in a distinctly more "horizontal" orientation than seen for PGT124 and PGT128 (Fig. 4.5G,H). This orientational difference is consistent with differences between DH270 and PGT124 or PGT128 in the lengths and configurations of their CDR loops, which required an alternative DH270 bnAb position when docked onto the surface of the Env trimer. We docked the BG505 SOSIP coordinates (*26*) and the Fab into the EM reconstruction, and further constrained the latter by the observed effects of mutations in the gp140 surface (fig 4.S16, 4.S17). Asp325 was essential for binding DH270.1; it is a potential partner for Arg57 on the Fab. Mutating Asp321 led to a modest loss in affinity; R327A had no effect (fig. 4.S18). These data further distinguish DH270 from PGT124 and PGT128. Mutating W101, Y105, D107, D115, Y116 or W117 in DH270.1 individually to alanine substantially reduced binding to the SOSIP trimer, as did pairwise mutation to alanines of S106 and S109. The effects of these mutations illustrate the critical role of the CDRH3 loop in binding with HIV-1 Env (fig. 4.S18).

Section 4.4 - Discussion

We can reconstruct from the data presented here a plausible series of events during the development of a V3-glycan bnAb in a natural infection. The DH272 and DH475 lineages neutralized the autologous TF and early viruses, and the resulting escape viruses were neutralized by the DH270 lineage. In particular, V1 deletions were necessary for neutralization of all but the most mature DH270 lineage antibodies. DH475 (and possibly DH272) escape variants stimulated DH270 affinity maturation, including both somatic mutations at sites of intrinsic mutability (*11*) and a crucial, improbable mutation at an AID coldspot within CDRH2 (G57R). The G57R mutation initiated expansion of the DH270 bnAb lineage. The low probability of this heterologous neutralization-conferring mutation and the complex lineage interactions that occurred likely explains why it took 4.5 years for the DH270 lineage to expand.

The CH848 viral population underwent a transition from a long V1 loop in the TF (34 residues) to short ones (16-17 residues) when escaping DH272/DH475 and facilitating expansion of DH270, to restoration of longer V1 loops later in infection as resistance to DH270 intermediates developed. Later DH270 antibodies adapted to viruses with longer V1 loops, allowing recognition of a broader spectrum of Envs and enhancing breadth. DH270.6 could neutralize heterologous viruses regardless of V1 loop length, but viruses with long loops tended to be less sensitive to it. Association of long V1 loops with reduced sensitivity was evident for three other V3 glycan bnAbs and may be a general feature of this class.

Since we cannot foresee the susceptibility to a particular bnAb lineage of each specific potential transmitted/founder virus to which vaccine recipients will be exposed, it will be important for a vaccine to induce bnAbs of multiple specificities to minimize transmitted/founder virus escape *(27, 28)*.

What stimulated the UCA of the DH270 bnAb lineage? The absence of detectable binding to the CH848 TF envelope as gp120, gp140 or membrane bound trimers (Fig. 4.6) raises two possibilities. One is that the lineage arose at the end of year 1, either from a primary response to viruses present at that time (e.g., with deletions in V1-V2) or from subversion of an antibody lineage initially elicited by some other antigen. The other is that some altered form of the CH848 TF envelope protein -- e.g. shed gp120, or a fragment of it -- exposed the V3 loop and the N301 and N332 glycans in a way that bound and stimulated the germline BCR, even though the native CH848 TF Env did not. The latter explanation is in keeping with our observation that the DH270 lineage intermediate antibodies did not recognize free glycans or cell surface membrane expressed gp160 trimers (Fig. 4.6). It is also consistent with the observation that a synthetic glycopeptide mimic of the V3-glycan bnAb gp120 epitope (Man_9-V3 glycopeptide) binds to the DH270 UCA (fig 4.S19A). The synthesis, antigenicity and ability of the Man_9-V3 glycopeptide to reflect the native Env V3-glycan bnAb epitope are described in detail by Alam and colleagues in the companion manuscript *(16)*. DH270 UCA also bound to the aglycone form of the same peptide (fig. 4.S19B). Similarly, the early intermediate antibodies (IA4, IA3, IA2) each bound to both the Man_9V3 glycopeptide and its aglycone form, and their binding was stronger to the aglycone V3

peptide than to the Man_9-V3 glycopeptide (fig. 4.S19B). Binding to the Man_9-V3 glycopeptide was low (>10μM) (fig. 4.S19A). DH270.1 bound the glycopeptide with higher affinity than did the aglycone ($K_{d,glycopeptide}$ = 331nM), correlated with an increased nucleotide mutation frequency (5.6%) (fig. 4.S19A,B). As mutations accumulated in the DH270 bnAb lineage, binding of the Man_9-V3 glycopeptide increased, culminating in a K_d of 188 nM in the most potent bnAb, DH270.6, which does not bind the aglycone-V3 peptide (fig. 4.S19A,B). Thus, both the Man_9-V3 glycopeptide and the aglycone-V3 peptide bound to the DH270.UCA and antibody binding was independent of glycans until the DH270 lineage had acquired a nucleotide mutation frequency of ~6%. These observations suggest that a denatured, fragmented or otherwise modified form of Env, with the V3-glycan epitope optimally exposed for binding the DH270 UCA, initiated the DH270 lineage (*29, 30*).

It will be important to define how often an improbable mutation such as G57R determines the time it takes for a bnAb lineage in an HIV-1 infected individual to develop, and how many of the accompanying mutations are necessary for potency or breadth rather than being non-essential mutations at AID mutational hotspots (*11, 31*). Mutations of the latter type might condition the outcome or modulate the impact of a key, improbable mutation, without contributing directly to affinity. Should the occurrence of an unlikely mutation be rate-limiting for breadth or potency in many other cases, a program of rational immunogen design will need to focus on modified envelopes most likely to select very strongly for improbable yet critical antibody nucleotide changes. A search for an Env that might select for the critical G57R mutation in DH270 UCA or

IA4-like antibodies yielded Env 10.17 from week 135 of infection (fig. 4.S20A,B), which derived from the only autologous virus Env that DH270.IA4 could bind.

The following proposal for a strategy to induce V3 glycan bnAbs recreates the events that led to bnAb induction in CH848: start by priming with a ligand that binds the bnAb UCA, such as the synthetic glycopeptide mimic of the V3-glycan bnAb gp120 epitope, then boost with an Env that can select G57R CDR H2 mutants, followed by Envs with progressive V1 lengths (fig. 4.S20C). We hypothesize that more direct targeting of V3-glycan UCAs and intermediate antibodies can accelerate the time of V3-glycan bnAb development in the setting of vaccination.

Finally, our study describes a general strategy for the design of vaccine immunogens that can select specific antibody mutations thereby directing antibody lineage maturation pathways.

Section 4.5 – References

1. D. R. Burton, J. R. Mascola, Antibody responses to envelope glycoproteins in HIV-1 infection. Nature immunology 16, 571-576 (2015).

2. J. R. Mascola, B. F. Haynes, HIV-1 neutralizing antibodies: understanding nature's pathways. Immunol. Rev. 254, 225-244 (2013).

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

4. L. M. Walker, S. K. Phogat, P. Y. Chan-Hui, D. Wagner, P. Phung, J. L. Goss, T. Wrin, M. D. Simek, S. Fling, J. L. Mitcham, J. K. Lehrman, F. H. Priddy, O. A. Olsen, S. M. Frey, P. W. Hammond, S. Kaminsky, T. Zamb, M. Moyle, W. C. Koff, P. Poignard, D. R. Burton, Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science 326, 285-289 (2009).

5. K. J. Doores, L. Kong, S. A. Krumm, K. M. Le, D. Sok, U. Laserson, F. Garces, P. Poignard, I. A. Wilson, D. R. Burton, Two classes of broadly neutralizing antibodies within a single lineage directed to the high-mannose patch of HIV envelope. J. Virol. 89, 1105-1118 (2015).

6. D. Sok, K. J. Doores, B. Briney, K. M. Le, K. L. Saye-Francisco, A. Ramos, D. W. Kulp, J. P. Julien, S. Menis, L. Wickramasinghe, M. S. Seaman, W. R. Schief, I. A.

Wilson, P. Poignard, D. R. Burton, Promiscuous glycan site recognition by antibodies to the high-mannose patch of gp120 broadens neutralization of HIV. Science translational medicine 6, 236ra263 (2014).

7. D. Sok, U. Laserson, J. Laserson, Y. Liu, F. Vigneault, J. P. Julien, B. Briney, A. Ramos, K. F. Saye, K. Le, A. Mahan, S. Wang, M. Kardar, G. Yaari, L. M. Walker, B. B. Simen, E. P. St John, P. Y. Chan-Hui, K. Swiderek, S. H. Kleinstein, G. Alter, M. S. Seaman, A. K. Chakraborty, D. Koller, I. A. Wilson, G. M. Church, D. R. Burton, P. Poignard, The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies. PLoS pathogens 9, e1003754 (2013).

8. H. Mouquet, L. Scharf, Z. Euler, Y. Liu, C. Eden, J. F. Scheid, A. Halper-Stromberg, P. N. Gnanapragasam, D. I. Spencer, M. S. Seaman, H. Schuitemaker, T. Feizi, M. C. Nussenzweig, P. J. Bjorkman, Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. Proc. Natl. Acad. Sci. U. S. A. 109, E3268-3277 (2012).

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

10. H. X. Liao, R. Lynch, T. Zhou, F. Gao, S. M. Alam, S. D. Boyd, A. Z. Fire, K. M. Roskin, C. A. Schramm, Z. Zhang, J. Zhu, L. Shapiro, J. C. Mullikin, S. Gnanakaran, P. Hraber, K. Wiehe, G. Kelsoe, G. Yang, S. M. Xia, D. C. Montefiori, R. Parks, K. E. Lloyd, R. M. Scearce, K. A. Soderberg, M. Cohen, G. Kamanga, M. K. Louder, L. M. Tran, Y. Chen, F. Cai, S. Chen, S. Moquin, X. Du, M. G. Joyce, S. Srivatsan, B.

Zhang, A. Zheng, G. M. Shaw, B. H. Hahn, T. B. Kepler, B. T. Korber, P. D. Kwong, J. R. Mascola, B. F. Haynes, Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496, 469-476 (2013).

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

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

13. M. Pancera, T. Zhou, A. Druz, I. S. Georgiev, C. Soto, J. Gorman, J. Huang, P. Acharya, G. Y. Chuang, G. Ofek, G. B. Stewart-Jones, J. Stuckey, R. T. Bailer, M. G. Joyce, M. K. Louder, N. Tumba, Y. Yang, B. Zhang, M. S. Cohen, B. F. Haynes, J.

R. Mascola, L. Morris, J. B. Munro, S. C. Blanchard, W. Mothes, M. Connors, P. D. Kwong, Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature 514, 455-461 (2014).

14. M. Bonsignori, K. K. Hwang, X. Chen, C. Y. Tsao, L. Morris, E. Gray, D. J. Marshall, J. A. Crump, S. H. Kapiga, N. E. Sam, F. Sinangil, M. Pancera, Y. Yongping, B. Zhang, J. Zhu, P. D. Kwong, S. O'Dell, J. R. Mascola, L. Wu, G. J. Nabel, S. Phogat, M. S. Seaman, J. F. Whitesides, M. A. Moody, G. Kelsoe, X. Yang, J. Sodroski, G. M. Shaw, D. C. Montefiori, T. B. Kepler, G. D. Tomaras, S. M. Alam, H. X. Liao, B. F. Haynes, Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. J. Virol. 85, 9998-10009 (2011).

15. E. S. Gray, M. A. Moody, C. K. Wibmer, X. Chen, D. Marshall, J. Amos, P. L. Moore, A. Foulger, J. S. Yu, B. Lambson, S. Abdool Karim, J. Whitesides, G. D. Tomaras, B. F. Haynes, L. Morris, H. X. Liao, Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. J. Virol. 85, 7719-7729 (2011).

16. S. M. Alam, B. Aussedat, Y. Vohra, R. R. Meyerhoff, E. M. Cale, W. E. Walkowicz, N. A. Radakovich, K. Anasti, L. Armand, R. J. Parks, M. G. Joyce, M. Pancera, A. Druz, I. S. Georgiev, T. Von Holle, A. Eaton, M. K. Louder, R. T. Bailer, L. Morris, S. A. Karim, M. Cohen, H. X. Liao, D. C. Montefiori, P. K. Park, A. Fernández-Tejada, K. Wiehe, T. B. Kepler, K. O. Saunders, J. Sodroski, P. D. Kwong, J. R.

Mascola, M. Bonsignori, M. A. Moody, S. Danishefsky, B. F.Haynes, Mimicry of an HIV Broadly Neutralizing Antibody Epitope with a Synthetic Glycopeptide. Submitted

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

18. We accessed the SF5 mutability model dataset at http://clip.med.yale.edu/shm/download.php.

19. R. Pejchal, K. J. Doores, L. M. Walker, R. Khayat, P. S. Huang, S. K. Wang, R. L. Stanfield, J. P. Julien, A. Ramos, M. Crispin, R. Depetris, U. Katpally, A. Marozsan, A. Cupo, S. Maloveste, Y. Liu, R. McBride, Y. Ito, R. W. Sanders, C. Ogohara, J. C. Paulson, T. Feizi, C. N. Scanlan, C. H. Wong, J. P. Moore, W. C. Olson, A. B. Ward, P. Poignard, W. R. Schief, D. R. Burton, I. A. Wilson, A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. Science 334, 1097-1103 (2011).

20. L. Kong, J. H. Lee, K. J. Doores, C. D. Murin, J. P. Julien, R. McBride, Y. Liu, A. Marozsan, A. Cupo, P. J. Klasse, S. Hoffenberg, M. Caulfield, C. R. King, Y. Hua, K. M. Le, R. Khayat, M. C. Deller, T. Clayton, H. Tien, T. Feizi, R. W. Sanders, J. C. Paulson, J. P. Moore, R. L. Stanfield, D. R. Burton, A. B. Ward, I. A. Wilson,

Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. Nature structural & molecular biology 20, 796-803 (2013).

21. J. P. Julien, D. Sok, R. Khayat, J. H. Lee, K. J. Doores, L. M. Walker, A. Ramos, D. C. Diwanji, R. Pejchal, A. Cupo, U. Katpally, R. S. Depetris, R. L. Stanfield, R. McBride, A. J. Marozsan, J. C. Paulson, R. W. Sanders, J. P. Moore, D. R. Burton, P. Poignard, A. B. Ward, I. A. Wilson, Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. PLoS pathogens 9, e1003342 (2013).

22. M. Pancera, Y. Yang, M. K. Louder, J. Gorman, G. Lu, J. S. McLellan, J. Stuckey, J. Zhu, D. R. Burton, W. C. Koff, J. R. Mascola, P. D. Kwong, N332-Directed broadly neutralizing antibodies use diverse modes of HIV-1 recognition: inferences from heavy-light chain complementation of function. PloS one 8, e55701 (2013).

23. P. L. Moore, E. S. Gray, C. K. Wibmer, J. N. Bhiman, M. Nonyane, D. J. Sheward, T. Hermanus, S. Bajimaya, N. L. Tumba, M. R. Abrahams, B. E. Lambson, N. Ranchobe, L. Ping, N. Ngandu, Q. Abdool Karim, S. S. Abdool Karim, R. I. Swanstrom, M. S. Seaman, C. Williamson, L. Morris, Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. Nat. Med. 18, 1688-1692 (2012).

24. LANL HIV Sequence Database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html)

25. F. Garces, D. Sok, L. Kong, R. McBride, H. J. Kim, K. F. Saye-Francisco, J. P. Julien, Y. Hua, A. Cupo, J. P. Moore, J. C. Paulson, A. B. Ward, D. R. Burton, I. A. Wilson, Structural evolution of glycan recognition by a family of potent HIV antibodies. Cell 159, 69-79 (2014).

26. J. H. Lee, N. de Val, D. Lyumkis, A. B. Ward, Model Building and Refinement of a Natively Glycosylated HIV-1 Env Protein by High-Resolution Cryoelectron Microscopy. Structure 23, 1943-1951 (2015).

27. M. Bonsignori, D. C. Montefiori, X. Wu, X. Chen, K. K. Hwang, C. Y. Tsao, D. M. Kozink, R. J. Parks, G. D. Tomaras, J. A. Crump, S. H. Kapiga, N. E. Sam, P. D. Kwong, T. B. Kepler, H. X. Liao, J. R. Mascola, B. F. Haynes, Two distinct broadly neutralizing antibody specificities of different clonal lineages in a single HIV-1-infected donor: implications for vaccine design. J. Virol. 86, 4688-4692 (2012).

28. K. Wagh, T. Bhattacharya, C. Williamson, A. Robles, M. Bayne, J. Garrity, M. Rist, C. Rademeyer, H. Yoon, A. Lapedes, H. Gao, K. Greene, M. K. Louder, R. Kong, S. A. Karim, D. R. Burton, D. H. Barouch, M. C. Nussenzweig, J. R. Mascola, L. Morris, D. C. Montefiori, B. Korber, M. S. Seaman, Optimal Combinations of Broadly Neutralizing Antibodies for Prevention and Treatment of HIV-1 Clade C Infection. PLoS Pathog. 12:e1005520 (2016).

29. M. Kuraoka, A. G. Schmidt, T. Nojima, F. Feng, A. Watanabe, D. Kitamura, S. C. Harrison, T. B. Kepler, G. Kelsoe, Complex Antigens Drive Permissive Clonal Selection in Germinal Centers. Immunity 44, 542-552 (2016).

30. L. Hangartner, R. M. Zinkernagel, H. Hengartner, Antiviral antibody responses: the two extremes of a wide spectrum. Nature reviews. Immunology 6, 231-243 (2006).

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

32. G. D. Tomaras, N. L. Yates, P. Liu, L. Qin, G. G. Fouda, L. L. Chavez, A. C. Decamp, R. J. Parks, V. C. Ashley, J. T. Lucas, M. Cohen, J. Eron, C. B. Hicks, H. X. Liao, S. G. Self, G. Landucci, D. N. Forthal, K. J. Weinhold, B. F. Keele, B. H. Hahn, M. L. Greenberg, L. Morris, S. S. Karim, W. A. Blattner, D. C. Montefiori, G. M. Shaw, A. S. Perelson, B. F. Haynes, Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. J. Virol. 82, 12449-12463 (2008).

33. E. E. Giorgi, B. Funkhouser, G. Athreya, A. S. Perelson, B. T. Korber, T. Bhattacharya, Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. BMC bioinformatics 11, 532 (2010).

34. G. M. Shaw, E. Hunter, HIV transmission. Cold Spring Harb Perspect Med 2, a006965 (2012).

35. W. B. Williams, H. X. Liao, M. A. Moody, T. B. Kepler, S. M. Alam, F. Gao, K. Wiehe, A. M. Trama, K. Jones, R. Zhang, H. Song, D. J. Marshall, J. F. Whitesides,

K. Sawatzki, A. Hua, P. Liu, M. Z. Tay, K. E. Seaton, X. Shen, A. Foulger, K. E. Lloyd, R. Parks, J. Pollara, G. Ferrari, J. S. Yu, N. Vandergrift, D. C. Montefiori, M. E. Sobieszczyk, S. Hammer, S. Karuna, P. Gilbert, D. Grove, N. Grunenberg, M. J. McElrath, J. R. Mascola, R. A. Koup, L. Corey, G. J. Nabel, C. Morgan, G. Churchyard, J. Maenza, M. Keefer, B. S. Graham, L. R. Baden, G. D. Tomaras, B. F. Haynes, HIV-1 VACCINES. Diversion of HIV-1 vaccine-induced immunity by gp41-microbiota cross-reactive antibodies. Science 349, aab1253 (2015).

36. D. C. Montefiori, Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. Curr Protoc Immunol Chapter 12, Unit 12 11 (2005).

37. H. X. Liao, M. C. Levesque, A. Nagel, A. Dixon, R. Zhang, E. Walter, R. Parks, J. Whitesides, D. J. Marshall, K. K. Hwang, Y. Yang, X. Chen, F. Gao, S. Munshaw, T. B. Kepler, T. Denny, M. A. Moody, B. F. Haynes, High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. J. Virol. Methods 158, 171-179 (2009).

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

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

40. T. B. Kepler, S. Munshaw, K. Wiehe, R. Zhang, J. S. Yu, C. W. Woods, T. N. Denny, G. D. Tomaras, S. M. Alam, M. A. Moody, G. Kelsoe, H. X. Liao, B. F. Haynes,

Reconstructing a B-Cell Clonal Lineage. II. Mutation, Selection, and Affinity Maturation. Frontiers in immunology 5, 170 (2014).

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

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

43. R. Bransteitter, P. Pham, P. Calabrese, M. F. Goodman, Biochemical analysis of hypermutational targeting by wild type and mutant activation-induced cytidine deaminase. J. Biol. Chem. 279, 51612-51621 (2004).

44. M. S. Seaman, H. Janes, N. Hawkins, L. E. Grandpre, C. Devoy, A. Giri, R. T. Coffey, L. Harris, B. Wood, M. G. Daniels, T. Bhattacharya, A. Lapedes, V. R. Polonis, F. E. McCutchan, P. B. Gilbert, S. G. Self, B. T. Korber, D. C. Montefiori, J. R. Mascola, Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. J. Virol. 84, 1439-1452 (2010).

45. X. Wu, Z. Y. Yang, Y. Li, C. M. Hogerkorp, W. R. Schief, M. S. Seaman, T. Zhou, S. D. Schmidt, L. Wu, L. Xu, N. S. Longo, K. McKee, S. O'Dell, M. K. Louder, D. L. Wycuff, Y. Feng, M. Nason, N. Doria-Rose, M. Connors, P. D. Kwong, M. Roederer, R. T. Wyatt, G. J. Nabel, J. R. Mascola, Rational design of envelope identifies

broadly neutralizing human monoclonal antibodies to HIV-1. Science 329, 856-861 (2010).

46. J. F. Salazar-Gonzalez, M. G. Salazar, B. F. Keele, G. H. Learn, E. E. Giorgi, H. Li, J. M. Decker, S. Wang, J. Baalwa, M. H. Kraus, N. F. Parrish, K. S. Shaw, M. B. Guffey, K. J. Bar, K. L. Davis, C. Ochsenbauer-Jambor, J. C. Kappes, M. S. Saag, M. S. Cohen, J. Mulenga, C. A. Derdeyn, S. Allen, E. Hunter, M. Markowitz, P. Hraber, A. S. Perelson, T. Bhattacharya, B. F. Haynes, B. T. Korber, B. H. Hahn, G. M. Shaw, Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J. Exp. Med. 206, 1273-1289 (2009).

47. B. F. Keele, E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A. Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J. M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F. Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T. Bhattacharya, B. T. Korber, B. H. Hahn, G. M. Shaw, Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. 105, 7552-7557 (2008).

48. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948 (2007).

49. M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649 (2012).

50. P. Hraber, B. Korber, K. Wagh, E. E. Giorgi, T. Bhattacharya, S. Gnanakaran, A. S. Lapedes, G. H. Learn, E. F. Kreider, Y. Li, G. M. Shaw, B. H. Hahn, D. C. Montefiori, S. M. Alam, M. Bonsignori, M. A. Moody, H. X. Liao, F. Gao, B. F. Haynes, Longitudinal Antigenic Sequences and Sites from Intra-Host Evolution (LASSIE) Identifies Immune-Selected HIV Variants. Viruses 7, 5443-5475 (2015).

51. J. L. Kirchherr, X. Lu, W. Kasongo, V. Chalwe, L. Mwananyanda, R. M. Musonda, S. M. Xia, R. M. Scearce, H. X. Liao, D. C. Montefiori, B. F. Haynes, F. Gao, High throughput functional analysis of HIV-1 env genes without cloning. J. Virol. Methods 143, 104-111 (2007).

52. E. P. Go, A. Herschhorn, C. Gu, L. Castillo-Menendez, S. Zhang, Y. Mao, H. Chen, H. Ding, J. K. Wakefield, D. Hua, H. X. Liao, J. C. Kappes, J. Sodroski, H. Desaire, Comparative Analysis of the Glycosylation Profiles of Membrane-Anchored HIV-1 Envelope Glycoprotein Trimers and Soluble gp140. J. Virol. 89, 8245-8257 (2015).

53. D. Fera, A. G. Schmidt, B. F. Haynes, F. Gao, H. X. Liao, T. B. Kepler, S. C. Harrison, Affinity maturation in an HIV broadly neutralizing B-cell lineage through reorientation of variable domains. Proc. Natl. Acad. Sci. U. S. A. 111, 10275-10280 (2014).

54. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt A 276, 307-326 (1997).

55. A. J. McCoy, Solving structures of protein complexes by molecular replacement with Phaser. Acta crystallographica. Section D, Biological crystallography 63, 32-41 (2007).

56. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. Journal of applied crystallography 40, 658-674 (2007).

57. T. Zhou, J. Zhu, X. Wu, S. Moquin, B. Zhang, P. Acharya, I. S. Georgiev, H. R. Altae-Tran, G. Y. Chuang, M. G. Joyce, Y. D. Kwon, N. S. Longo, M. K. Louder, T. Luongo, K. McKee, C. A. Schramm, J. Skinner, Y. Yang, Z. Yang, Z. Zhang, A. Zheng, M. Bonsignori, B. F. Haynes, J. F. Scheid, M. C. Nussenzweig, M. Simek, D. R. Burton, W. C. Koff, J. C. Mullikin, M. Connors, L. Shapiro, G. J. Nabel, J. R. Mascola, P. D. Kwong, Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. Immunity 39, 245-258 (2013).

58. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography 66, 213-221 (2010).

59. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. Acta crystallographica. Section D, Biological crystallography 60, 2126-2132 (2004).

60. K. Cowtan, P. Main, Miscellaneous algorithms for density modification. Acta crystallographica. Section D, Biological crystallography 54, 487-493 (1998).

61. I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson, D. C. Richardson, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35, W375-383 (2007).

62. R. W. Sanders, R. Derking, A. Cupo, J. P. Julien, A. Yasmeen, N. de Val, H. J. Kim, C. Blattner, A. T. de la Pena, J. Korzun, M. Golabek, K. de Los Reyes, T. J. Ketas, M. J. van Gils, C. R. King, I. A. Wilson, A. B. Ward, P. J. Klasse, J. P. Moore, A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. PLoS pathogens 9, e1003618 (2013).

63. G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38-46 (2007).

64. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612 (2004).

65. S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59, 307-321 (2010).

66. D. C. Nickle, L. Heath, M. A. Jensen, P. B. Gilbert, J. I. Mullins, S. L. Kosakovsky Pond, HIV-specific probabilistic models of protein evolution. PloS one 2, e503 (2007).

67. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 20, 289-290 (2004).

68. S. Gnanakaran, M. G. Daniels, T. Bhattacharya, A. S. Lapedes, A. Sethi, M. Li, H. Tang, K. Greene, H. Gao, B. F. Haynes, M. S. Cohen, G. M. Shaw, M. S. Seaman, A. Kumar, F. Gao, D. C. Montefiori, B. Korber, Genetic signatures in the envelope glycoproteins of HIV-1 that associate with broadly neutralizing antibodies. PLoS computational biology 6, e1000955 (2010).

69. T. Bhattacharya, M. Daniels, D. Heckerman, B. Foley, N. Frahm, C. Kadie, J. Carlson, K. Yusim, B. McMahon, B. Gaschen, S. Mallal, J. I. Mullins, D. C. Nickle, J. Herbeck, C. Rousseau, G. H. Learn, T. Miura, C. Brander, B. Walker, B. Korber, Founder effects in the assessment of HIV polymorphisms and HLA allele associations. Science 315, 1583-1586 (2007).

70. G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, WebLogo: a sequence logo generator. Genome Res. 14, 1188-1190 (2004).

71. F. Garces, J. H. Lee, N. de Val, A. T. de la Pena, L. Kong, C. Puchades, Y. Hua, R. L. Stanfield, D. R. Burton, J. P. Moore, R. W. Sanders, A. B. Ward, I. A. Wilson, Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. Immunity 43, 1053-1063 (2015).

72. A. C. Siepel, A. L. Halpern, C. Macken, B. T. Korber, A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. AIDS Res. Hum. Retroviruses 11, 1413-1416 (1995).

73. L. Kong, A. Torrents de la Pena, M. C. Deller, F. Garces, K. Sliepen, Y. Hua, R. L. Stanfield, R. W. Sanders, I. A. Wilson, Complete epitopes for vaccine design derived from a crystal structure of the broadly neutralizing antibodies PGT128 and 8ANC195 in complex with an HIV-1 Env trimer. Acta crystallographica. Section D, Biological crystallography 71, 2099-2108 (2015).

74. P. Wang, W. Kim. L. B. Pickens, X. Gao, Y. Tang, Heterologous expression and manipulation of three tetracycline biosynthetic pathways. Angew Chem Int Ed Engl 51, 11136-11140 (2012).

Section 4.6 – Acknowledgements and Contributions

Acknowledgments. The authors thank Michael C. Mangiapani, Robert J. Parks, Callie Vivian, Krissey Lloyd, Daniel M. Kozink, Florence Perrin, Abby J. Cooper, Andrew Foulger, Giovanna Hernandez, Jamie Pritchett, Sabrina Arora, Thaddeus Gurley, Lawrence Armand, Amanda Martelli and Amy Wang for technical assistance; John Whitesides and Dawn Marshall of the DHVI Flow Core Facility for single-cell sorting; Gerald H. Learn, Timothy G. Decker, and Yingying Li for single genome sequencing and analysis; the beam line staff at Advanced Photon Source 24-ID-C and 24-ID-E for support during data collection and Harvard Medical School electron microscopy staff for training and assistance with EM data; and Kelly Soderberg for project management. The authors thank the HIV Vaccine Trials Network (HVTN) for the continued use of NGS pre-vaccination samples from the HVTN082 and HVTN204 trials (*31*). E.F.K is supported by a Ruth L. Kirschstein National Research Service Award F30 AI112426. D.F. is supported by a F32 fellowship (1F32AI116355-01) from the NIH. R.R.M is supported by a Medical Scientist Training Program (MSTP) training grant T32GM007171 and by a Ruth L. Kirschstein National Research Service Award F30-AI122982-0, NIAID. K.O.S. is supported by a NIAID extramural project grant R01-AI120801. This work was funded by UM-1- AI100645 from NIH, NIAID, Division of AIDS, Duke Center for HIV/AIDS Vaccine Immunology-Immunogen Discover (CHAVI-ID) to B.F.H.

Authors contributions. MB, BHH, SCH, BTK and BFH coordinated and designed the study, evaluated data, wrote and edited manuscript and figures; MB, EFK, DF, TB, KW, KOS, AM, BTK and SCH produced figures; MB directed memory B cell cultures and functional primary screenings, performed intrinsic mutability analysis; EFK isolated viruses, analyzed data, wrote and edited text; DF, BWP, CAJ and SCH performed structural analyses; RRM isolated antibodies; TB performed site directed mutagenesis and NGS, and analyzed data; KW analyzed NGS and antibody clonal lineage data, wrote and edited text; BA, WEW and SJD synthesized the Man_9V3 glycopeptide; SMA analyzed data and provided intellectual contributions; KKH, MAG and AM executed memory B cell cultures and functional screenings of purified antibodies, tabulated and graphed data; KOS performed glycan analyses, made figures; RZ, AT, KOS, HXL produced plasmids, expressed recombinant proteins; AK, FG produced pseudoviruses and made pseudovirus mutants; SMX, MC, MKL, KMK, RTB, JRM and DCM performed neutralization studies; GK, JRM, GMS provided guidance and intellectual contributions; WBW provided V_H1-2*02 NGS data from HIV-1-negative individuals; LM directed neutralization and epitope mapping studies on CH848 plasma; JK produced transfected cell lines; KW, PTH, BTK performed virus signature analyses, analyzed virus evolution, GK and MSC supervised clinical sites; TBK performed and supported computational analysis of antibody clones; MAM directed antigen-specific memory B cell sorting; BFH conceived the study, evaluated all data, wrote and edited the paper. All authors approved the paper.

Section 4.7 – Figures

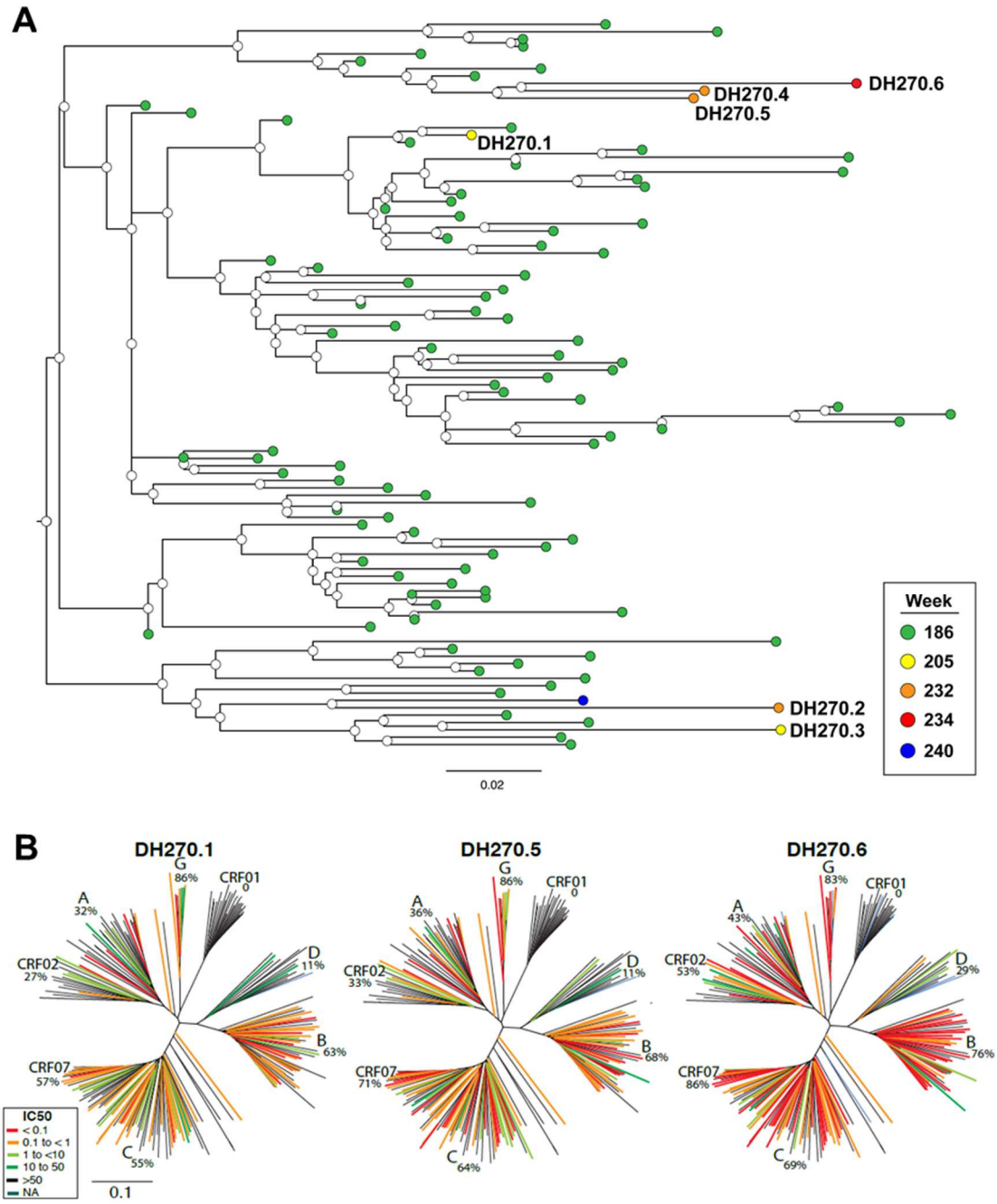


Figure 4-1 DH270 lineage with time of appearance and neutralization by selected members

(**A**) Phylogenetic relationship of 6 mAbs and 93 NGS V_HDJ_H sequence reads in the DH270 clonal lineage. External nodes (filled circles) represent V_HDJ_H nucleotide sequences of either antibodies retrieved from cultured and sorted memory B cells (labeled) or a curated dataset of NGS V_HDJ_H rearrangement reads (unlabeled). Coloring is by time of isolation. Samples from week 11, 19, 64, 111, 160, 186 and 240 were tested and time-points from which no NGS reads within the lineage were retrieved are reported in table 4.S2. The majority of NGS sequence reads were retrieved from week 186 (green circles; see details in Methods). Internal nodes (open circles) represent inferred ancestral intermediate sequences. Units for branch-length estimates are nucleotide substitution per site. (**B**) Neutralization dendrograms display single mAb neutralization of a genetically diverse panel of 207 HIV-1 isolates. Coloring is by IC_{50}.

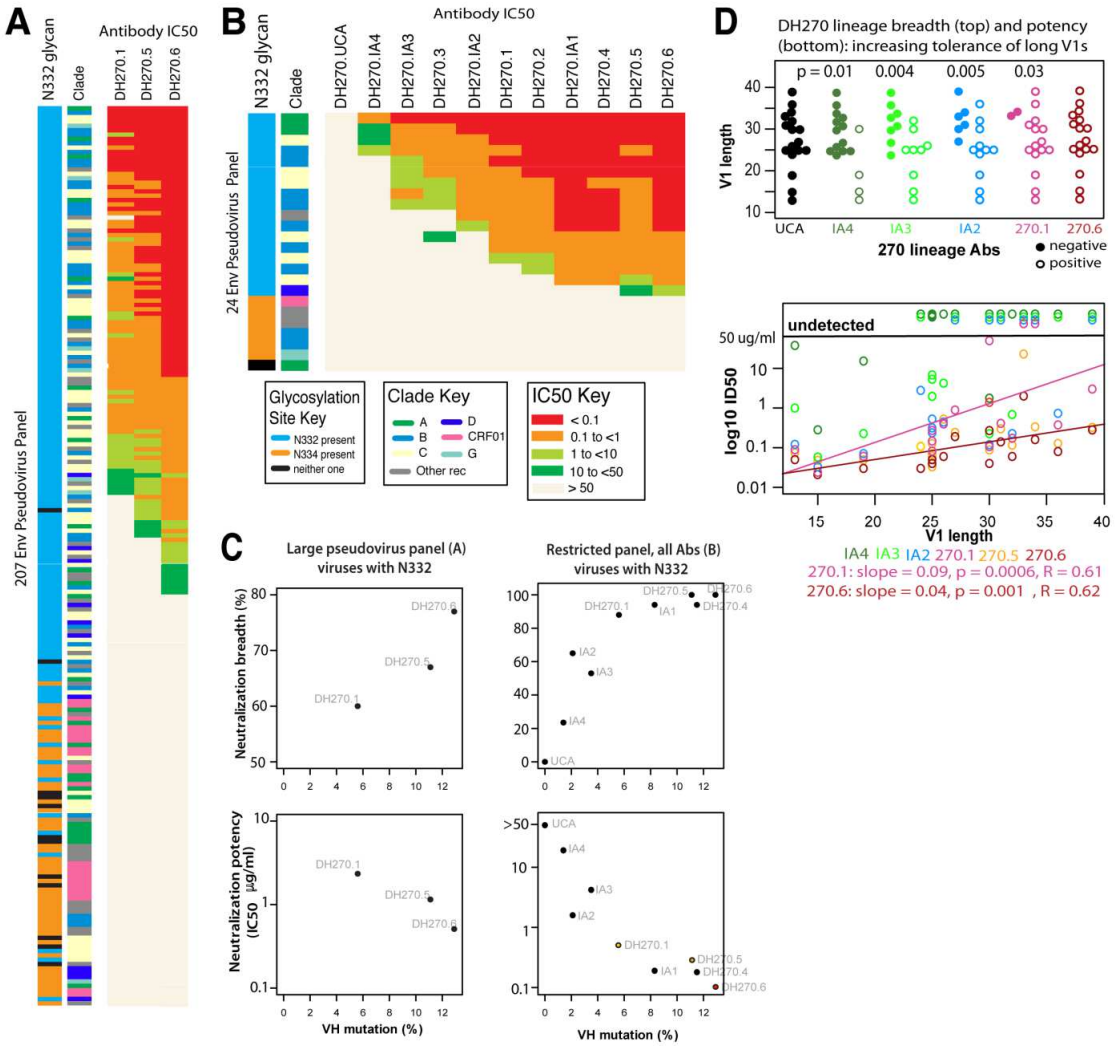


Figure 4-2 Heterologous breadth in the DH270 lineage

(**A**) Neutralizing activity of DH270.1, DH270.5 and DH270.6 bnAbs (columns) for 207 tier 2 heterologous viruses (rows). Coloring is by neutralization IC_{50} (µg/ml). Keys to the presence of an N332 N-glycosylation site and clade are on the left. (**B**). Heterologous neutralization of all DH270 lineage antibodies for a 24-virus panel. Coloring follows panel A. See fig. S1 (**C**) Co-variation between V_H mutation frequencies (*x*-axis), neutralization breadth (*y*-axis, top panels) and potency (*y*-axis, bottom panels) of individual antibodies against N332 viruses from the larger (left) and smaller (right) pseudovirus panels. (**D**) Correlation between viral V1 loop length and DH270 lineage antibody neutralization. Top panel: neutralization of 17 viruses (with N332 and sensitive to at least one DH270 lineage antibody) by selected DH270 lineage antibodies from UCA to mature bnAbs (*x*-axis). Viruses are identified by their respective V1 loop lengths (*y*-axis); for each virus, neutralization sensitivity is indicated by a clear circle and resistance by a solid circle. The *p*-value is a Wilcoxon rank sum comparison of V1 length

distributions between sensitive and resistant viruses. Bottom panel: regression lines (IC_{50} for neutralization vs. V1 loop length) for DH270.1 and DH270.6, with a p-value based on Kendall's tau.

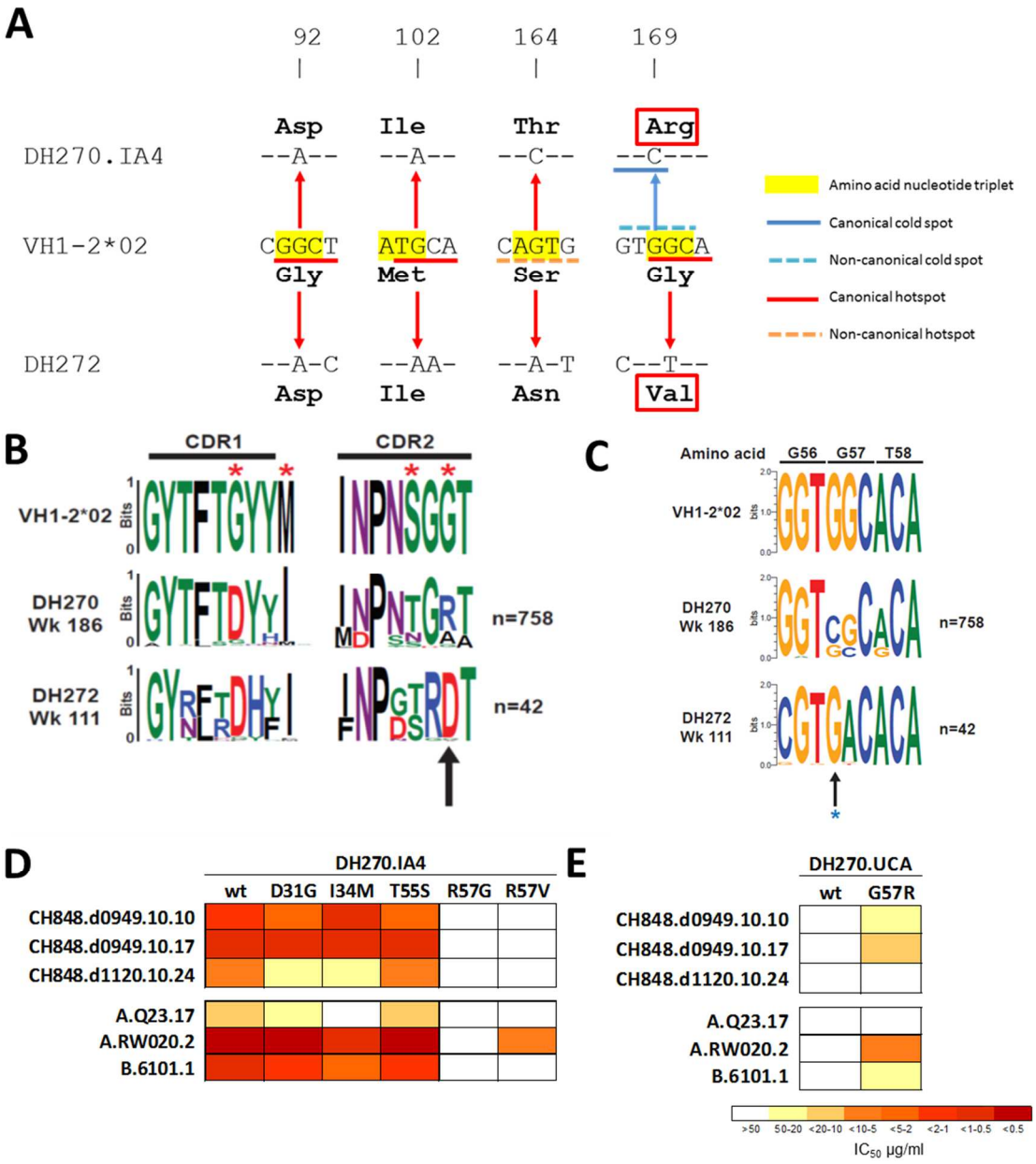


Figure 4-3 A single disfavored mutation early during DH270 clonal development conferred neutralizing acitivty to the V3 glycan bnAb DH270 precursor antibodies

(A) Nucleotide (nt) alignment of DH270.IA4 and DH272 to V_H1-2*02 sequence at the four V_H positions that mutated from DH270.UCA to DH270.IA4. The codons that encode the amino acids (aa) affected by the mutations are highlighted in yellow. AID hotspots are indicated by red and orange lines (canonical and non-canonical, respectively); AID cold spots by blue lines (solid: canonical; dotted: non-canonical)(*16*). At position 169, DH270.IA4 retained positional conformity with DH272 but not identity conformity (red

boxes). (**B**) Sequence logo plot of aa mutated from germline (top) in NGS reads of the DH270 (middle) and DH272 (bottom) lineages at weeks 186 and 111 post-transmission, respectively. Red asterisks indicate aa mutated in DH270.IA4. The black arrow indicates lack of identity conformity between the two lineages at aa position 57. (**C**) Sequence logo plot of nucleotide mutations (position 165-173) in the DH270 and DH272 lineages at weeks 186 and 111 post-transmission, respectively. The arrow indicates position 169. (**D**) Effect of reversion mutations on DH270.IA4 neutralization. Coloring is by IC_{50}. (**E**) Effect of G57R mutation on DH270.UCA autologous (top) and heterologous (bottom) neutralizing activity.

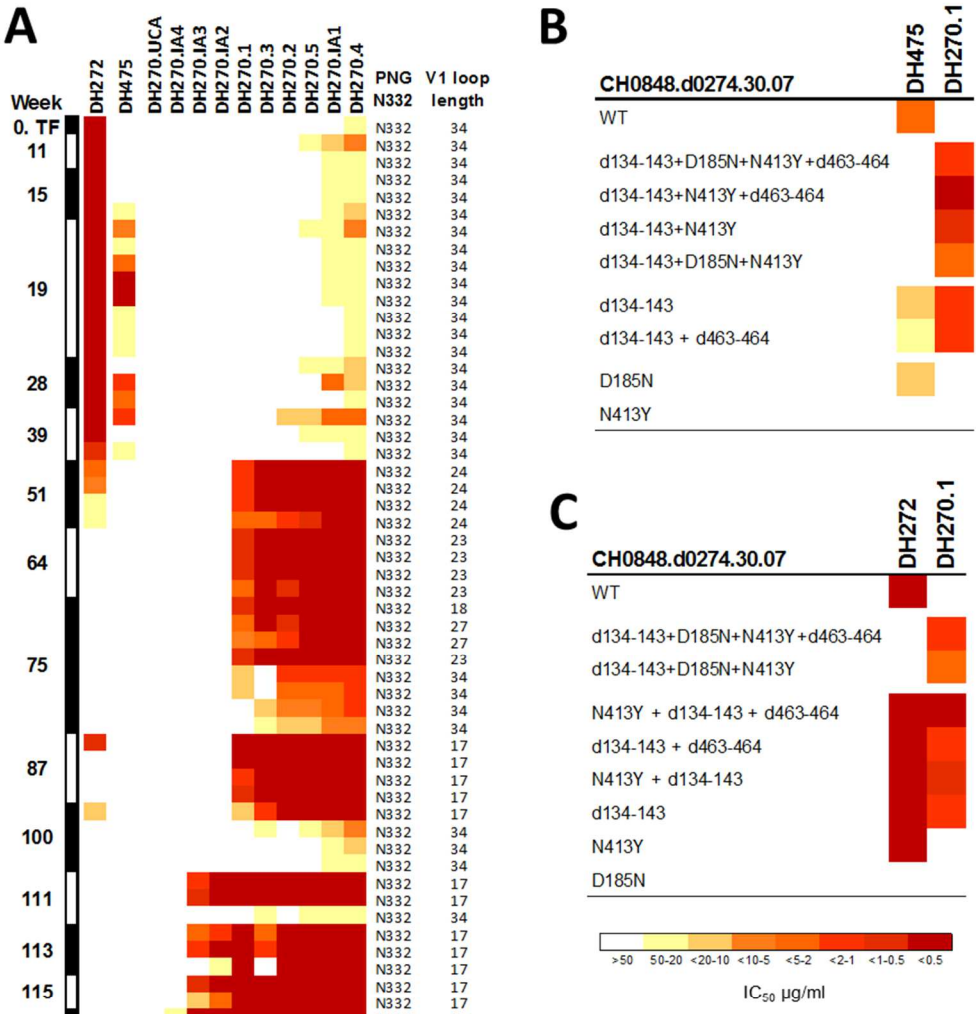

A
Week
0. TF
11
15
19
28
39
51
64
75
87
100
111
113
115
119
123
128
136
160
186
205
231
234
236
240
246
DH272
DH475
DH270.UCA
DH270.IA4
DH270.IA3
DH270.IA2
DH270.1
DH270.3
DH270.2
DH270.5
DH270.IA1
DH270.4
PNG N332
V1 loop length
B
CH0848.d0274.30.07
DH475
DH270.1
WT
d134-143+D185N+N413Y+d463-464
d134-143+N413Y+d463-464
d134-143+N413Y
d134-143+D185N+N413Y
d134-143
d134-143 + d463-464
D185N
N413Y
C
CH0848.d0274.30.07
DH272
DH270.1
WT
d134-143+D185N+N413Y+d463-464
d134-143+D185N+N413Y
N413Y + d134-143 + d463-464
d134-143 + d463-464
N413Y + d134-143
d134-143
N413Y
D185N
>50
50-20
<20-10
<10-5
<5-2
<2-1
<1-0.5
<0.5
IC50 µg/ml

Figure 4-4 Cooperation among DH270, DH272, and DH475 N332 dependent V3 glycan nAb lineages

(A) Neutralizing activity of DH272, DH475 and DH270 lineage antibodies against CH848 viruses. Heat map summarizing neutralization data from 90 pseudoviruses (row) and 10 DH270 lineage mAbs plus DH272 and DH475 (columns). Neutralization potency (IC_{50}) is shown in different colors from white (>50μg/ml) to dark red (<0.5μg/ml), as indicated in the bar. The Env pseudoviruses were generated for the CH848 TF virus and variants from weeks 11 to 246 post-transmission and were assayed in the TZM-bl cell-based neutralization assay. For each pseudovirus, presence of an N332 PNG site and V1 loop length are indicated on the right. (**B, C**) Susceptibility to DH270.1 and to (B) DH475 or (C) DH272 of autologous viruses bearing selected immunotype-specific mutations. Heat map as described for panel A.

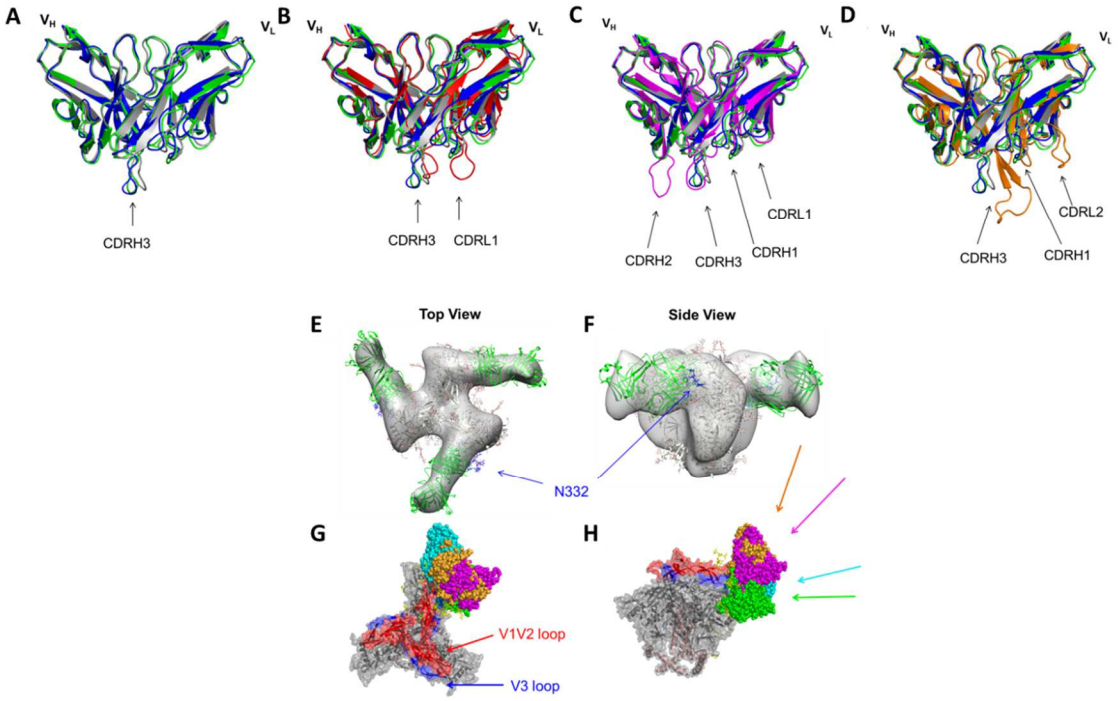


Figure 4-5 Fab crystal structures and 3D reconstruction of DH270.1 bound with the 92BR SOSIP.664 trimer

Superposition of backbone ribbon diagrams for DH270 lineage members: UCA1 (gray), DH270.1 (green), and DH270.6 (blue): (**A**) alone, (**B**) with the DH272 cooperating antibody (red), (**C**) with PGT 128 (magenta), and (**D**) with PGT124 (orange). Major differences in CDR regions are indicated by an arrow. (**E**) Top and (**F**) side views of a fit of the DH270.1 Fab (green) and the BG505 SOSIP trimer (gray) into a map obtained from negative-stain EM. (**G**) Top and (**H**) side views of the BG505 trimer (PDB ID: 5ACO)(*25*) (gray, with V12 and V3 loops highlighted in red and blue, respectively) bound with PGT122 (PDB ID: 4NCO)(*20*) (orange), PGT128 (PDB ID: 3TYG)(*18*) (magenta), PGT135 (PDB ID: 4JM2)(*19*) (cyan) and DH270.1 (green), superposed to illustrate the different positions of the several Fabs on gp140. The arrows indicate the direction of the principal axis of each of the bnAb Fabs; the color of each arrow is the color that represents the corresponding bnAb.

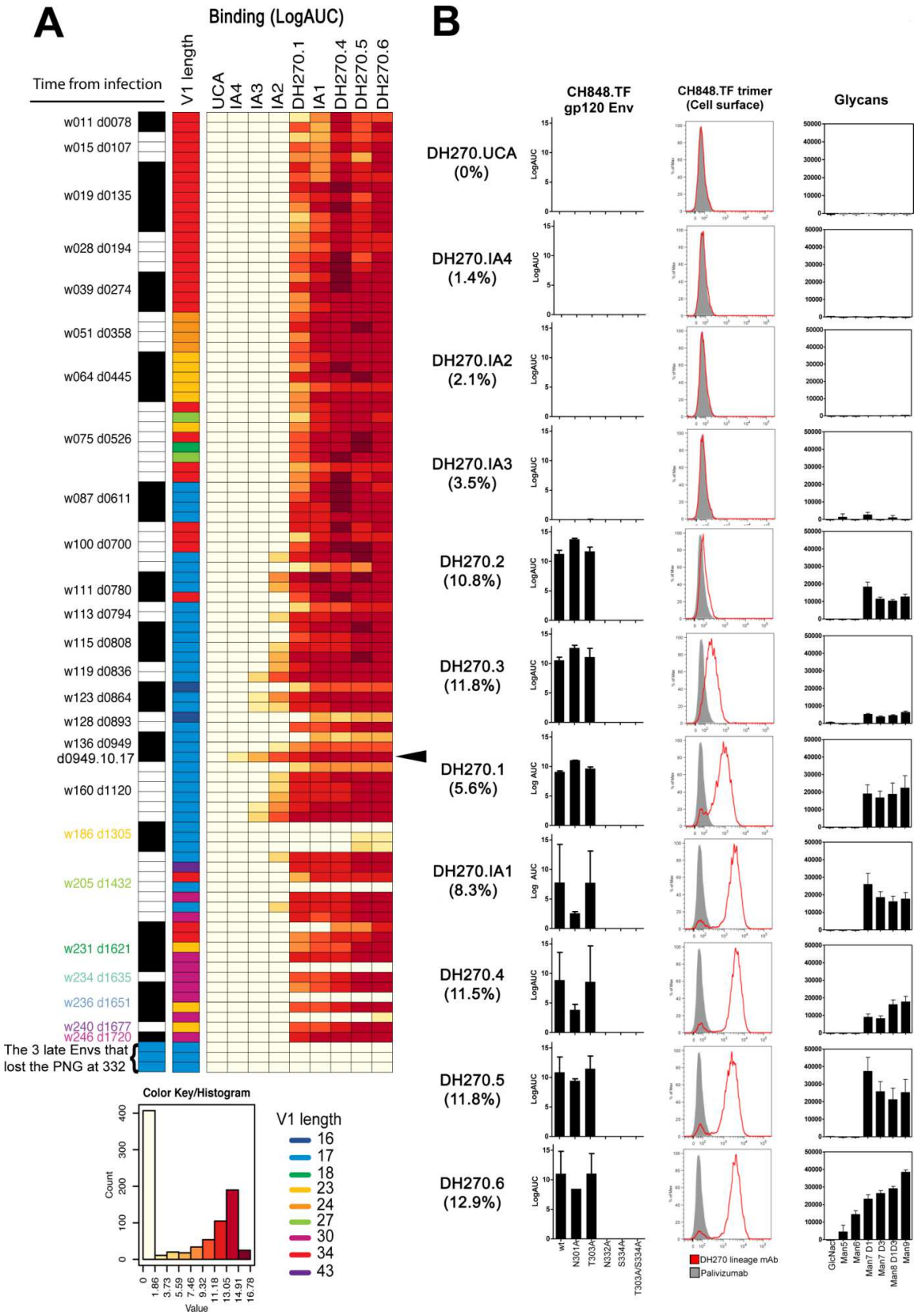

A
Binding (LogAUC)
Time from infection
V1 length
UCA
IA4
IA3
IA2
DH270.1
IA1
DH270.4
DH270.5
DH270.6
w011 d0078
w015 d0107
w019 d0135
w028 d0194
w039 d0274
w051 d0358
w064 d0445
w075 d0526
w087 d0611
w100 d0700
w111 d0780
w113 d0794
w115 d0808
w119 d0836
w123 d0864
w128 d0893
w136 d0949
d0949.10.17
w160 d1120
w186 d1305
w205 d1432
w231 d1621
w234 d1635
w236 d1651
w240 d1677
w246 d1720
The 3 late Envs that
lost the PNG at 332
Color Key/Histogram
Count
Value
V1 length
16
17
18
23
24
27
30
34
43
B
CH848.TF
gp120 Env
CH848.TF trimer
(Cell surface)
Glycans
DH270.UCA
(0%)
DH270.IA4
(1.4%)
DH270.IA2
(2.1%)
DH270.IA3
(3.5%)
DH270.2
(10.8%)
DH270.3
(11.8%)
DH270.1
(5.6%)
DH270.IA1
(8.3%)
DH270.4
(11.5%)
DH270.5
(11.8%)
DH270.6
(12.9%)
LogAUC
wt
N301A
T303A
N332A
S334A
T303A/S334A
DH270 lineage mAb
Palivizumab
GlcNac
Man5
Man6
Man7 D1
Man7 D3
Man8 D1D3
Man9

Figure 4-6 DH270 lineage antibody binding to autologous CH0848 Env components

(A) DH270 lineage antibodies (column) were tested by ELISA for binding to 120 CH848 autologous gp120 Env glycoproteins (rows) isolated from time of infection (CH848.TF) to 245 weeks post-infection. CH848 Envs are grouped based on week of isolation. Day of isolation is also shown. The first column indicates presence or absence of an N-linked glycosylation site in position 332 with the color scheme of Figure 1. The second column shows V1 length of each CH848 virus. The binding of individual antibodies is expressed as log area under the curve (LogAUC) in a heat map color-coded as indicated. (**B**) Left: Binding to CH848.TF mutants with disrupted N301 and/or N332 glycan sites. Results are expressed as LogAUC. Middle: Binding to CH848 Env trimer expressed on the cell surface of CHO cells. Results are expressed as maximum percentage of binding and are representative of duplicate experiments. DH270 antibodies are shown in red. Palivizumab is the negative control (gray area). Right: Binding to free glycans measured on a microarray. Results are the average of background-subtracted triplicate measurements and are expressed in RU.

Section 4.8 – Materials and Methods

Study Design

The CH848 donor, from which the DH270, DH272 and DH475 antibody lineages were isolated, is an African male enrolled in the CHAVI001 acute HIV-1 infection cohort (*32*) and followed for 5 years, after which he started antiretroviral therapy. During this time viral load ranged from 8,927 to 442,749 copies/ml (median = 61,064 copies/ml), and CD4 counts ranged from 288 to 624 cells/mm^3 (median = 350 cells/ mm^3). The time of infection was estimated by analyzing the sequence diversity in the first available sample using the Poisson Fitter tool (*33*) as described in (*10*). Results were consistent with a single founder virus establishing the infection (*34*).

MAbs DH270.1 and DH270.3 were isolated from cultured memory B cells isolated 205 weeks post-transmission (*14*). DH270.6 and DH475 mAbs were isolated from Man9-V3 glycopeptide-specific memory B cells collected 232 and 234 weeks post-transmission, respectively, using direct sorting. DH270.2, DH270.4 and DH270.5 mAbs were isolated from memory B cells collected 232 weeks post-transmission that bound to Consensus C gp120 Env but not to Consensus C N332A gp120 Env using direct sorting

Experimental methods

Detailed description of experimental methods is provided in Supplementary Materials and

<u>Flow Cytometry, Memory B Cell Cultures and mAb Isolation</u>

A total of 30,700 memory B cells from individual CH848 were isolated from PBMC collected 205 weeks post-transmission using magnetic-activated cell sorting as described in (*14*). Memory B cells were cultured at limiting dilution at a calculated concentration of 2 cells/well for 2 weeks as described in (*11*) using irradiated CD40L L cells (7,500 cGy) as feeder cells at a concentration of 5,000 cells/well; culture medium was refreshed 7 days after plating. Cell culture supernatants were screened for neutralization of autologous CH848.TF virus using the tzm-bl neutralization assay (*14, 36*) and for binding to CH848.TF gp120 Env, CH848.TF gp140 Env, Consensus C gp120 Env and consensus C N332A gp120 Env. Concurrently, cells from each culture were transferred in RNAlater (Qiagen) and stored at -80°C until functional assays were completed.

MAbs DH270.1 and DH270.3 were isolated from cultures that bound to CH848.TF gp120 Env and Consensus C gp120 but did not bind to C N332A gp120 Env. DH272 was isolated from a culture that neutralized 99% CH848.TF virus infectivity. DH272 dependency to N332-linked glycans was first detected on the transiently transfected recombinant antibody tested at higher concentration and confirmed in the purified recombinant antibody (*37*). From the stored RNAlater samples, mRNA of cells from these cultures was extracted and retrotranscribed as previously described (*14*).

DH270.6 and DH475 mAbs were isolated from Man9-V3 glycopeptide-specific memory B cells collected 232 and 234 weeks post-transmission, respectively, using direct sorting *(16)*. Briefly, biotinylated Man9–V3 peptides were tetramerized via streptavidin that was conjugated with either AlexaFluor 647 (AF647; ThermoScientific) or Brilliant Violet 421 (BV421) (Biolegend) dyes. Peptide tetramer quality following conjugation

was assessed by flow cytometry to a panel of well-characterized HIV-1 V3 glycan antibodies (PGT128, and 2G12) and linear V3 antibodies (F39F) attached to polymer beads. PBMCs from donor CH848 were stained with LIVE/DEAD Fixable Aqua Stain (ThermoScientific), anti-human IgM (FITC), CD3 (PE-Cy5), CD235a (PE-Cy5), CD19 (APC-Cy7), and CD27 (PE-Cy7) (BD Biosciences); anti-human antibodies against IgD (PE); anti-human antibodies against CD10 (ECD), CD38 (APC-AF700), CD19 (APC-Cy7), CD16 (BV570), CD14 (BV605) (Biolegend); and Man9GlcNac2 V3 tetramer in both AF647 and BV421. PBMCs that were Aqua Stain -, CD14-, CD16-, CD3-, CD235a-, positive for CD19+, and negative for surface IgD were defined as memory B cells; these cells were then gated for Man9–V3+ positivity in both AF647 and BV421, and were single-cell sorted using a BD FACS Aria II into 96-well plates containing 20 μl of reverse transcriptase buffer (RT).

DH270.2, DH270.4 and DH270.5 mAbs were isolated from memory B cells collected 232 weeks post-transmission that bound to Consensus C gp120 Env but not to Consensus C N332A gp120 Env using direct sorting. Reagents were made using biotinylated Consensus C gp120 Env and Consensus C N332A gp120 Env by reaction with streptavidin that was conjugated with either AlexaFluor 647 (AF647; ThermoScientific) or Brilliant Violet 421 (BV421) (Biolegend) dyes, respectively. Env tetramer quality following conjugation was assessed by flow cytometry to a panel of well-characterized HIV-1 V3 glycan antibodies (PGT128, and 2G12) and linear V3 antibodies (F39F) attached to polymer beads. PBMCs were stained as outlined for DH475 and DH270.6, however these cells were then gated for Consensus C gp120 positivity and Consensus C

N332A gp120 negativity in AF647 and BV421, respectively, and were single cell sorted and processed as outlined for DH475 and DH270.6.

For all antibodies, cDNA synthesis, PCR amplification, sequencing and V(D)J rearrangement analysis were conducted as previously described (*11*). Reported mutation frequency is calculated as frequency of nucleotide mutations in the V gene region of antibody sequence. CDRH3 lengths reported are defined as the number of residues after the invariant Cys in FR3 and before the invariant Trp in FR4.

Antibody production

Immunoglobulin genes of mAbs DH270.1 through DH270.6, DH272 and DH475 were amplified from RNA from isolated cells, expression cassettes made, and mAbs expressed as described (*12*). Inference of unmutated common ancestor (UCA) and intermediate antibodies DH270.IA1 through DH270.IA4 was conducted using methods previously described (*38*).

Heavy chain plasmids were co-transfected with appropriate light chain plasmids at an equal ratio in Expi 293 cells using ExpiFectamine 293 transfection reagents (Thermo Fisher Scientific) according to the manufacturer's protocols. We used the enhancer provided with the kit, transfected cultures were incubated at 37°C 8% CO2 for 2-6 days, harvested, concentrated and incubated overnight with Protein A beads at 4°C on a rotating shaker before loading the bead mixture in columns for purification; following PBS/NaCl wash, eluate was neutralized with trizma hydrochloride and antibody

concentration was determined by Nanodrop. Purified antibodies were tested in SDS-Page Coomassie and western blots, and stored at 4°C.

Next-generation sequencing

PBMC-extracted RNA from weeks 11, 19, 64, 111, 160, 186, and 240 post-infection were used to generate cDNA amplicons for next-generation sequencing (Illumina Miseq) as described previously (*35*). Briefly, RNA isolated from PBMCs was separated into two equal aliquots before cDNA production; cDNA amplification and NGS were performed on both aliquots as independent samples (denoted A and B). Reverse transcription (RT) was carried out using human IgG, IgA, IgM, Igκ and Igλ primers as previously described (*37*). After cDNA synthesis, IgG isotype IGHV1 and IGHV3 genes were amplified separately from weeks 11, 19, 64, 111, 160, and 186. IGHV1-IGHV6 genes were amplified at week 240. A second PCR step was performed to add Nextera index sequencing adapters (Illumina) and libraries were purified by gel extraction (Qiagen) and quantified by quantitative PCR using the KAPA SYBR FAST qPCR kit (KAPA Biosystems). Each replicate library was sequencing using the Illumina Miseq V3 2x 300bp kit.

NGS reads were computationally processed and analyzed as previously described (*35*). Briefly, forward and reverse reads were merged with FLASH (*39*) with average read length and fragment read length parameters set to 450 and 300, respectively. Reads were quality filtered using FASTX (http://hannonlab.cshl.edu/fastx_toolkit/) for sequences with a minimum of 50 percent of bases with a Phred quality score of 20 or greater

(corresponding to 99% base call accuracy). Primer sequences were discarded and only unique nucleotide sequences were retained. To mitigate errors introduced during PCR amplification, reads detected in sample A and B with identical nucleotide V_HDJ_H rearrangement sequences were delineated as replicated sequences. The total number of unique reads per sample and total number of replicated sequences ("Overlap") across samples for each time point is listed in **table S2**. We used replicated sequences to define presence of antibody clonal lineages at any time-point.

We identified clonally-related sequences to DH270, DH272 and DH475 from the longitudinal NGS datasets by the following procedure. First, the CDR H3 of the probe-identified clonal parent sequence was BLASTed (E-value cutoff =.01) against the pooled sample A and B sequence sets at each timepoint to get a candidate set of putative clonal members ("candidate set"). Next we identified replicated sequences across samples A and B in the candidate set. We then performed a clonal kinship test with the Cloanalyst software package (http://www.bu.edu/computationalimmunology/research/software/) (*38, 40*) as previously described (*35*) on replicated sequences. Clonally-related sequences within Sample A and B (including non-replicated sequences) were identified by performing the same clonal kinship test with Cloanalyst on the candidate set prior to identifying replicated sequences.

Clonal lineage reconstruction was performed on the NGS replicated sequences and probe-identified sequences of each clone using the Cloanalyst software package. A maximum of 100 sequences were used as input for inferring phylogenetic trees of clonal lineages. Clonal sequence sets were sub-sampled down to 100 sequences by collapsing to

one sequence within a 2 or 9 base pair difference radius for the DH272 and DH270 clones, respectively.

The pre-vaccination NGS samples that were analyzed in fig. S11A were obtained from HIV-1 uninfected participants of the HVTN082 and HVTN204 trials as previously described (*35*).

Sequence Analysis of Antibody Clonal Lineages

Unmutated common ancestors (UCA) and ancestral intermediate sequences were computationally inferred with the Cloanalyst software package (*38*). Cloanalyst uses Bayesian inference methods to infer the full unmutated V(D)J rearrangement thereby including a predicted unmutated CDR3 sequence. For lineage reconstructions when only cultured or sorted sequences were used as input, the heavy and light chain pairing relationship was retained during the inference of ancestral sequences. UCA inferences were performed each time a new member of the DH270 clonal lineage was experimentally isolated and thus several versions of the DH270 UCA were produced and tested. UCA1 and UCA3 were used for structural determination. UCA4 (referred to as DH270.UCA throughout the text), which was inferred using the most observed DH270 clonal members and had the lowest uncertainty of UCAs inferred (as quantified by the sum of the error probability over all base positions in the sequence), was used for binding and neutralization studies. Subsequently, the DH270 UCA was also re-inferred when NGS data became available. We applied a bootstrapping procedure to infer the UCA with the NGS data included, resampling clonal lineage trees 10 times with 100 input

NGS sequences each. The UCA4 amino acid sequence was recapitulated by 7 out of 10 UCA inferences of the resampled NGS trees confirming support for UCA4.

Each inference of V(D)J calls is associated with a probability. The probability of the DH270 lineage to use the V_H1-2 family gene was 99.99% and that of using allele 02 (V_H1-2*02) was 98.26%. Therefore, there was a 0.01% probability that the family was incorrectly identified and a 1.74% probability that the allele was incorrectly identified. Therefore, we sequenced genomic DNA of individual CH848. As previously reported, positional conformity is defined as sharing a mutation at the same position in the V gene segment and identity conformity as sharing the same amino acid substitution at the same position (*11*).

We refer to the widely established AID hot and cold spots (respectively WRCY and SRY and their reverse-complements) as "canonical" and to other hot and cold spots defined by Yaari et al. as "non-canonical" (*17, 41-43*).

Sequencing of germline variable region from genomic DNA.

Genomic DNA was isolated from donor CH848 from PBMCs 3 weeks after infection (QIAmp DNA Blood mini kit; Qiagen). IGVH1-2 and IGVL2-23 sequences were amplified using 2 independent primer sets by PCR. To ensure amplification of non-rearranged variable sequences, both primer sets reverse primers aligned to sequences present in the non-coding genomic DNA downstream the V-recombination site. The forward primer for set 1 resided in the IGVH1-2 and IGVL2-23 leader sequences and upstream of the leader in set 2. The PCR fragments were cloned into a pcDNA2.1

(TOPO-TA kit; Life technologies) and transformed into bacteria for sequencing of individual colonies. The following primers were used: VH1-2_1_S: tcctcttcttggtggcagcag; VH1-2_2_S: tacagatctgtcctgtgccct; VH1-2_1_tmAS: ttctcagccccagcacagctg; VH1-2_2_TmAS: gggtggcagagtgagactctgtcaca; VL2-23_2_S: agaggagcccaggatgctgat; VL2-23_1_S: actctcctcactcaggacaca; VL2-23_1_AS: tctcaaggccgcgctgcagca; VL2-23_2_AS: agctgtccctgtcctggatgg.

We identified two variants of V_H1-2*02: the canonical sequence and a variant that encoded a V_H that differed by 9 amino acids. Of these 9 amino acids, only 1 was shared among DH270 antibodies whereas 8 amino acids were not represented in DH270 lineage antibodies (fig. S11B). The V_H1-2*02 variant isolated from genomic DNA did not encode an arginine at position 57. We conclude that between the two variants of V_H1-2*02 identified from genomic DNA from this individual, the DH270 lineage is likely derived from the canonical V_H1-2*02 sequence.

Direct binding ELISA

Direct-binding ELISAs were performed as described (*11*). Briefly, 384-well plates were blocked for 1 h at room temperature (RT) or overnight at 4°C (both procedures were previously validated); primary purified antibodies were tested at a starting concentrations of 100μg/ml, serially three-fold diluted and incubated for 1 h at RT; HRP-conjugated human IgG antibody was added at optimized concentration of 1:30,000 in assay diluent for 1 hour and developed using TMB substrate; plates were read at 450 nm in a SpectraMax 384 PLUS reader (Molecular Devices, Sunnyvale, CA); results are reported as logarithm area under the curve (LogAUC) unless otherwise noted.

For biotinylated avi-tagged antigens, plates were coated with streptavidin (2µg/ml); blocked plates were stored at -20°C until used and biotinylated avi-tagged antigens were added at 2 µg/ml for 30 minutes at RT.

Competition ELISAs were performed using 10µl of primary purified monoclonal antibody, starting at 100µg/ml and diluted in a two-fold concentration, incubated for 1 h at RT. Ten µl of biotinylated target Mab was added at the EC_{50} determined by a direct binding of biotinylated-Mab for one hour at RT. After background subtractions, percent inhibition was calculated as follows: 100-(test Ab triplicate mean/no inhibition control mean)*100.

Assessment of virus neutralization

Antibody and plasma neutralization was measured in TZM-bl cell-based assays performed as described previously (*36*). Neutralization breadth of DH270.1, DH270.5 and DH270.6 was assessed using the 384-well plate declination of the assay using an updated panel of 207 geographically and genetically diverse Env-pseudoviruses representing the major circulating genetic subtypes and recombinant forms as described (*44, 45*). The data were calculated as a reduction in luminescence units compared with control wells, and reported as IC_{50} in µg/ml (*36*).

Single genome sequencing and pseudovirus production

3' half genome single genome sequencing of HIV-1 from longitudinally collected plasma was performed as previously described (*46, 47*). Sequence alignment was

performed using ClustalW (version 2.11) (*48*) and was adjusted manually using Geneious 8 (version 8.1.6)(*49*). Env amino acid sequences were then aligned and evaluated for sites under selection using code derived from the Longitudinal Antigenic Sequences and Sites from Intra-host Evolution (LASSIE) tool (*50*). Using both LASSIE-based analysis and visual inspection, 100 representative *env* genes were selected for pseudovirus production. CMV promoter-ligated *env* genes were prepared and used to generate pseudotyped viruses as previously described (*51*).

Generation of cell surface-expressed CH848 Env trimer CHO cell line.

The membrane-anchored CH848 TF Env trimer was expressed in CHO-S cells. Briefly, the CH848 env sequence was codon-optimized and cloned into an HIV-1-based lentiviral vector. A heterologous signal sequence from CD5 was inserted replacing that of the HIV-1 Env. The proteolytic cleavage site between gp120 and gp41 was altered, substituting serine residues for Arg508 and Arg511, the tyrosine at residues 712 was changed to alanine (Y712A), and the cytoplasmic tail was truncated by replacing the Lys808 codon with a sequence encoding (Gly)3 (His)6 followed immediately by a TAA stop codon. This env-containing sequences was inserted into the vector immediately downstream of the tetracycline (tet)-responsive element (TRE), and upstream of an internal ribosome entry site (IRES) and a contiguous puromycin (puro)-T2A-EGFP open reading frame (generating K4831), as described previously for the JRFL and CH505 Envs (*52*).

CHO-S cells (Invitrogen) modified to constitutively express the reverse tet transactivator (rtTA) were transduced with packaged vesicular stomatitis virus (VSV) G glycoprotein-pseudotyped CH848 Env expression vector. Transduced cells were incubated in culture medium containing 1μg/ml of doxycycline (dox) and selected for 7 days in medium supplemented with 25 μg/ml of puromycin, generating the Env expressor-population cell line termed D831. From D831, a stable, high-expressor clonal cell line was derived, termed D835. The integrity of the recombinant *env* sequence in the clonal cell lines was confirmed by direct (without cloning) sequence analysis of PCR amplicons.

Cell surface-expressed trimeric CH848 Env binding.

D831 Selected TRE2.CH848.JF-8.IRS6A Chinese Hamster Ovary Cells were cultured in DMEM/F-12 supplemented with HEPES and L-glutamine (Thermo Fischer, Cat#11330057) 10% heat inactivated fetal bovine serum [FBS] (Thermo Fischer, Cat#10082147) and 1% Penicillin-Streptomycin (Thermo Fischer, Cat#15140163) and harvested when 70-80% confluent by trypsinization. A total 75,000 viable cells/well were transferred in 24-well tissue culture plates. After a 24-to-30-hour incubation at 37°C/5% CO2 in humidified atmosphere, CH848 Envs expression was induced with 1μg/mL doxycycline (Sigma-Aldrich, Cat#D9891) treatment for 16-20 hours. Cells were then washed in Stain buffer [PBS/2% FBS] and incubated at 4°C for 30 minutes. Stain buffer was removed from cells and 0.2ml/well of DH270 lineage antibodies, palivizumab (negative control) or PGT128 (positive control) were added at optimal concentration of

5μg/mL for 30 minutes at 4°C. After a 2X wash, cells were stained with 40 ul of APC-conjugated mouse anti-Human IgG (BD Pharmigen, Cat#562025) per well (final volume 0.2 ml/well) for 30 minutes at 4°C. Unstained cells were used as further negative control. Cells were washed 3X and gently dissociated with 0.3ml/well PBS/5mM EDTA for 30 minutes at 4°C, transferred into 5mL Polystyrene Round-Bottom Tubes (Falcon, Cat#352054), fixed with 0.1mL of BD Cytofix/Cytoperm Fixation solution (BD Biosciences, Cat#554722) and kept on ice until analyzed using a BD LSRFortessa Cell Analyzer. Live cells were gated through Forward/Side Scatter exclusion, and then gated upon GFP+ and APC.

Oligomannose Arrays.

Oligomannose arrays were printed with glycans at 100, 33, and 10 μM (Z Biotech). Arrays were blocked for 1h in Hydrazide glycan blocking buffer. Monoclonal antibodies were diluted to 50 μg/mL in Hydrazide Glycan Assay Buffer, incubated on an individual subarray for 1 h, and then washed 5 times PBS supplemented with 0.05% tween-20 (PBS-T). Subarrays that received biotinylated Concanavalin A were incubated with streptavidin-Cy3 (Sigma), whereas all other wells were incubated with anti-IgG-Cy3 (Sigma) for 1h while rotating at 40 rpm covered from light. The arrays were washed 5 times with 70 μL of PBS-T and then washed once with 0.01X PBS. The washed arrays were spun dry and scanned with a GenePix 4000B (Molecular Devices) scanner at wavelength 532 nm using GenePix Pro7 software. The fluorescence within each feature was background subtracted using the local method in GenePix Pro7 software (Molecular

Devices). To determine glycan specific binding, the local background corrected fluorescence of the print buffer alone was subtracted from each feature containing a glycan.

Affinity measurements

Antibody binding kinetic rate constants (k_a, k_d) of the Man_9-V3 glycopeptide and its aglycone form *(16)* were measured by Bio-layer Interferometry (BLI, ForteBio Octet Red96) measurements. The BLI assay was performed using streptavidin coated sensors (ForteBio) to capture either biotin-tagged Man_9-V3 glycopeptide or Aglycone-V3 peptide. The V3 peptide immobilized sensors were dipped into varying concentrations of antibodies following blocking of sensors in BSA (0.1%). Antibody concentrations ranged from 0.5 to 150 μg/mL and non-specific binding interactions were subtracted using the control anti-RSV Palivizumab (Synagis) mAb. Rate constants were calculated by global curve fitting analyses to the Bivalent Avidity model of binding responses with a 10 min association and 15 min dissociation interaction time. The dissociation constant (K_d) values without avidity contribution were derived using the initial components of the association and dissociation rates (k_{a1} and k_{d1}) respectively. Steady-state binding K_d values for binding to Man_9-V3 glycopeptide with avidity contribution were derived using near steady-state binding responses at varying antibody concentrations (0.5-80μg/mL) and using a non-linear 4-paramater curve fitting analysis.

HIV-1 Env site-directed mutagenesis

Deletion Mutant of CH0848.d0274.30.07 *env* gene was constructed using In Fusion HD EcoDry Cloning kit (Clontech, Mountain View, CA) as per manufacturer instructions. Quick Change II Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA) was used to introduce point mutations. All final env mutants were confirmed by sequencing.

Antibody site-directed mutagenesis

Site-directed mutagenesis of antibody genes was performed using the Quikchange II lightening multi-site-directed mutagenesis kit following manufacturer's protocol (Agilent). Mutant plasmid products were confirmed by single-colony sequencing. Primers used for introducing mutations were: DH270_IA4_D31G: cccagtgtatatagtagccggtgaaggtgtatcca; DH270.IA4 I34M: tcgcacccagtgcatatagtagtcggtgaaggtgt; DH270.IA4 T55S: gatggatcaaccctaactctggtcgcacaaactat; DH270.IA4 R57G: tgtgcatagtttgtgccaccagtgttagggttgat; DH270.IA4 R57V: cttctgtgcatagtttgtgacaccagtgttagggttgatc; DH270.UCA G57R: atcaaccctaacagtggtcgcacaaactatgcaca.

Env glycoprotein expression

The codon-optimized CH848-derived env genes were generated by de novo synthesis (GeneScript, Piscataway, NJ) or site-directed mutagenesis in mammalian expression

plasmid pcDNA3.1/hygromycin (Invitrogen, Grand Island, NY) as described (*10*), and stored at −80°C until use.

Expression and purification of DH270 lineage members

The heavy- and light-chain variable and constant domains of the DH270 lineage Fabs were cloned into the pVRC-8400 expression vector using Not1 and Nhe1 restriction sites and the tissue plasminogen activator signal sequence. The C terminus of the heavy-chain constructs contained a noncleavable 6x histidine tag. Site-directed mutagenesis was carried out, using manufacturer's protocols (Stratagene), to introduce mutations into the CDR regions of DH270.1.

Fabs were expressed and purified as described previously (*53*).

Crystallization, structure determination, and refinement

All His-tagged Fabs were crystallized at 20-25 mg/mL. Crystals were grown in 96-well format using hanging drop vapor diffusion and appeared after 24–48 h at 20 °C. Fab crystals were obtained in the following conditions: 2.5 M ammonium sulfate and 100 mM sodium acetate, pH 5.0 for DH272; 1.5M ammonium sulfate and 100mM sodium acetate pH 4.0 for UCA1; 20% PEG 4K, 100 mM sodium acetate, pH 5 and 100 mM magnesium sulfate for UCA3; 10% PEG 8K, 100 mM PIPES pH 6 and 1M NaCl for DH270.1; 1.4M lithium sulfate and 100 mM sodium acetate, pH 4.5 for DH270.3; 40% PEG 400 and 100 mM sodium citrate, pH 4.0 for DH270.5; and 30% PEG 4K, 100 mM PIPES pH 6, 1M

NaCl for DH270.6. All crystals were harvested and cryoprotected by the addition of 20–25% glycerol to the reservoir solution and then flash-cooled in liquid nitrogen.

Diffraction data were obtained at 100 K from beam lines 24-ID-C and 24-ID-E at the Advanced Photon Source using a single wavelength. Datasets from individual crystals (multiple crystals for UCA1, DH270.1 and DH270.5) were processed with HKL2000 (*54*). Molecular replacement calculations for the free Fabs were carried out with PHASER (*55, 56*), using I3.2 from the CH103 lineage [Protein Data Bank (PDB) ID 4QHL](*53*) or VRC01 from the VRC01/gp120 complex [Protein Data Bank (PDB) ID 4LST](*57*) as the starting models. Subsequent structure determinations were performed using DH270 lineage members as search models. The Fab models were separated into their variable and constant domains for molecular replacement.

Refinement was carried out with PHENIX (*58*), and all model modifications were carried out with Coot (*59*). During refinement, maps were generated from combinations of positional, group B-factor, and TLS (translation/libration/screw) refinement algorithms. Secondary-structure restraints were included at all stages for all Fabs; noncrystallographic symmetry restraints were applied to the DH270.1 and UCA3 Fabs throughout refinement. The resulting electron density map for DH270.1 was further improved by solvent flattening, histogram matching, and non-crystallographic symmetry averaging using the program DM (*60*). Phase combination was disabled in these calculations. After density modification, restrained refinement was performed using Refmac in Coot. Structure validations were performed periodically during refinement

using the MolProbity server (*61*). The final refinement statistics are summarized in **Table S4**.

Design of the 92BR SOSIP.664 construct

To generate the clade B HIV-1 92BR SOSIP.664 expression construct we followed established SOSIP design parameters (*62*). Briefly, the 92BR SOSIP.664 trimer was engineered with a disulfide linkage between gp120 and gp41 by introducing A501C and T605C mutations (HxB2 numbering system) to covalently link the two subunits of the heterodimer (*62*). The I559P mutation was included in the heptad repeat region 1 (HR1) of gp41 for trimer stabilization, and part of the hydrophobic membrane proximal external region (MPER), in this case residues 664–681 of the Env ectodomain, was deleted (*62*). The furin cleavage site between gp120 and gp41 ($_{508}$REKR$_{511}$) was altered to $_{506}$RRRRRR$_{511}$ to enhance cleavage (*62*). The resulting, codon-optimized 92BR SOSIP.664 *env* gene was obtained from GenScript (Piscataway, NJ) and cloned into pVRC-8400 as described above for Fabs using Nhe1 and NotI.

Purification of Envs for analysis by biolayer interferometry and negative stain EM

SOSIP.664 constructs were transfected along with a plasmid encoding the cellular protease furin at a 4:1 Env:furin ratio in HEK 293F cells. Site-directed mutagenesis was performed using manufacturer's protocols (Stratagene) for mutations in the V3 region and glycosylation sites. The cells were allowed to express soluble SOSIP.664 trimers for 5–7 days. Culture supernatants were collected and cells were removed by centrifugation

at 3800 x g for 20 min, and filtered with a 0.2 µm pore size filter. SOSIP.664 proteins were purified by flowing the supernatant over a lectin (*Galanthus nivalis*) affinity chromatography column overnight at 4°C. The lectin column was washed with 1x PBS and proteins were eluted with 0.5 M methyl-α-D-mannopyranoside and 0.5 M NaCl. The eluate was concentrated and loaded onto a Superdex 200 10/300 GL column (*GE Life Sciences*) prequilibrated in a buffer of 10 mM Hepes, pH 8.0, 150 mM NaCl and 0.02% sodium azide for EM, or in 2.5mM Tris, pH 7.5, 350 mM NaCl, 0.02% sodium azide for binding analysis, to separate the trimer-size oligomers from aggregates and gp140 monomers.

Electron Microscopy

Purified 92BR SOSIP.664 trimer was incubated with a five molar excess of DH270.1 Fab at 4 °C for 1 hour. A 3 µL aliquot containing ~0.01 mg/ml of the Fab - 92BR SOSIP.664 complex was applied for 15 s onto a carbon coated 400 Cu mesh grid that had been glow discharged at 20 mA for 30 s, followed by negative staining with 2% uranyl formate for 30 s. Samples were imaged using a FEI Tecnai T12 microscope operating at 120kV, at a magnification of 52,000x that resulted in a pixel size of 2.13 Å at the specimen plane. Images were acquired with a Gatan 2K CCD camera using a nominal defocus of 1500 nm at 10° tilt increments, up to 50°. The tilts provided additional particle orientations to improve the image reconstructions.

Negative Stain Image Processing and 3D Reconstruction

Particles were picked semi-automatically using EMAN2 (*63*) and put into a particle stack. Initial, reference-free, two-dimensional (2D) class averages were calculated and particles corresponding to complexes (with three Fabs bound) were selected into a substack for determination of an initial model. The *initial* model was calculated in EMAN2 using 3-fold symmetry and EMAN2 was used for subsequent refinement using 3-fold symmetry. In total, 5,419 particles were included in the final reconstruction for the 3D average of 92BR SOSIP.664 trimer complex with DH270.1. The resolution of the final model was determined using a Fourier Shell Correlation (FSC) cut-off of 0.5.

Model fitting into the EM reconstructions

The cryo-EM structure of PGT128-liganded BG505 SOSIP.664 (PDB ID: 5ACO) (*26*) and crystal structure of DH270.1 were manually fitted into the EM density and refined by using the UCSF Chimera 'Fit in map' function (*64*).

Biolayer Interferometry

Kinetic measurements of Fab binding to Envs were carried out using the Octet QK^{e} system (*ForteBio*); 0.2 mg/mL of each His-tagged Fab was immobilized onto an anti-Human Fab-CH1 biosensor until it reached saturation. The SOSIP.664 trimers were tested at concentrations of 200 nM and 600 nM in duplicate. A reference sample of buffer alone was used to account for any signal drift that was observed during the experiment. Association and dissociation were each monitored for 5 min. All experiments were conducted in the Octet instrument at 30 °C in a buffer of 2.5 mM Tris, pH 7.5, 350 mM

NaCl and 0.02% sodium azide with agitation at 1,000 rpm. Analyses were performed using nonlinear regression curve fitting using the Graphpad Prism software, version 6.

Protein Structure Analysis and Graphical Representations.

The Fabs and their complexes analyzed in this study were superposed by least squares fitting in Coot. All graphical representations with protein crystal structures were made using PyMol.

Definition of immunological virus phenotypes and virus signature analysis

The maximum likelihood trees depicting the heterologous virus panel and the full set of Env sequences for the subject CH848 were created using the Los Alamos HIV database PhyML interface (*65*). HIV substitution models (*66*) were used and the proportion of invariable sites and the gamma parameters were estimated from the data. Illustrations were made using the Rainbow Tree interface that utilizes Ape (*67*). The analysis that coupled neutralization data with the within-subject phylogeny based on Envs that were evaluated for neutralization sensitivity was performed using LASSIE (*50*). Signature analysis was performed using the methods fully described in (*68, 69*).

Statistical analysis

Statistical analysis was performed using R (http://www.R-project.org). Heat maps and logo plots (*70*) were generated using the Los Alamos HIV database web interfaces (www.hiv.lanl.gov, version Dec. 2015, HEATMAP and Analyze Align).

Accession Numbers

Coordinates and structure factors for UCA1, UCA3, DH270.1, DH270.3, DH270.5, DH270.6, and DH272 have been deposited in the Protein Data Bank with accession code T.B.D., T.B.D., T.B.D., T.B.D., T.B.D., T.B.D., and T.B.D., respectively.

Section 4.9 – Supplementary Text

Evolution of the CH848 virus quasispecies

Figure S6 shows that simultaneously with the first detection of the DH270 lineage at week 186, four distinctive phylogenetic lineages emerged in individual CH848 and each clade functionally defined a distinct immunological resistance profiles (i.e. immunotype). **Figure S9** shows the position of the four virus immunotypes on the full CH848 tree. The first profile was a set of Envs that all shifted the PNG site at N332 to 334 (**fig. S6**, open circles). Despite this mutation being associated with complete resistance to all antibodies in the DH270 lineage, this mutation was detected at relatively low frequency in CH848 (7-33% per sample), and was only detected transiently, not observed among 50 sequences from the last time point sampled (week 246). This suggested a balance where immune escape was countered by a cost in virological fitness. As noted above, the PNG site N332 was required for heterologous neutralization; the limited exposure *in vivo* may not have been adequate for selection of antibodies that could compensate for its loss. Env in the other 3 resistance profiles all retained N332, and all were embedded in clades that persisted through the last time point sampled. Envs

in the second profile were completely neutralization resistant, and their gp120s did not bind to the DH270 antibodies (**fig. S6**, triangles and **Additional Data table S3**). The gp120s of the other two sets could bind to DH270 gp120; one had weak neutralization sensitivity but only to later antibodies in the lineage (**fig. S6**, "X") the other had no detectable neutralization (**fig. S6**, "+"). The selection for resistance to the DH270 lineage as soon as it became established is clear: persistence of four divergent clades in CH848 Env protein, each with distinctive immunological resistance phenotypes, suggest multiple distinctive immune escape routes are explored and selected. This would allow continuing Env escape mutations to accrue in distinct frameworks, and exposure of the antibody to Env diversity in the context of these frameworks may select for antibodies that provide greater breadth.

Ontogeny of DH270 lineage and acquisition of neutralization breadth

Only one of 62 pseudoviruses tested that lack the PNG site at N332, the B clade virus 5768.04, was sensitive to DH270.5 and DH270.6 (**Additional Dataset S1**). Across the full M group HIV-1 virus isolate panel used in neutralization assays, the loss of the PNG N332 sites accounted for 70% of the observed neutralization resistance. The circulating recombinant form CRF01, commonly found in Asia, very rarely has this glycosylation site (3% of sequences in the Los Alamos database and 4% (1/23) in our test panel) and, as shown in **Fig. 2A**, DH270 lineage antibodies did not neutralize CRF01 strains.

Other V3-glycan bnAbs (10-1074, PGT121 and PGT128) shared this N332 glycan dependency but PGT121 and PGT128 were not as restrictive (**Additional Dataset S1**)(*5, 6, 8*). Antibody 10-1074 was similar to DH270.6 in that it more strictly required the N332 PNG site, and its neutralization potency correlated with that of DH270.6 (Pearson's $p = 8.0e^{-13}$, $r = 0.63$)(*8*).

As a consequence of the N332 PNG site requirement of V3 glycan bnAbs to neutralize, *in vitro* estimation of neutralizing breadth was impacted simply by the fraction of CRF01 viruses included in the panel. Therefore, to evaluate the impact of other virus characteristics relevant to DH270 lineage neutralizing, we excluded viruses lacking the N332 PNG from the analysis. In addition, reagents that target this epitope would be a poor choice for use in Asian populations where CRF01 is frequently observed.

CH848 Env signatures

We previously studied cooperation between lineages that occurred soon after infection, at a time when diversity in the autologous quasispecies was limited (*12*). In contrast, in CH848 the earliest autologous quasispecies transition in sensitivity to DH272/DH475 neutralization to DH270 lineage members occurred between week 39 and week 51, when multiple virus variants were circulating. Viral diversity made it impractical to test all the possible permutations or mutations from the transmitted founder virus. To select a smaller pool of candidate mutations, we sought the two most similar CH848 Env sequences at the amino acid level with opposite sensitivity to DH272/DH475 and DH270.1 neutralization around week 51 and identified clones CH0848.3.d0274.30.07

and CH0848.3.d0358.80.06 being the most similar (sim: 0.98713). Among the differences in amino acid sequences between these two clones, the four that we selected (Δ134-143 in V1); D185N in V2; N413Y in V4; Δ463-464 in V5) were the only ones consistently different among all clones with differential sensitivity to DH272 and DH270.1. We elected to use DH270.1 for these cooperating studies as the least mutated representative of DH270 antibodies that gained autologous neutralization at week 51. The D185N and N413Y mutations were also identified by the signature analysis shown in **fig. S12** and **Additional dataset S4**.

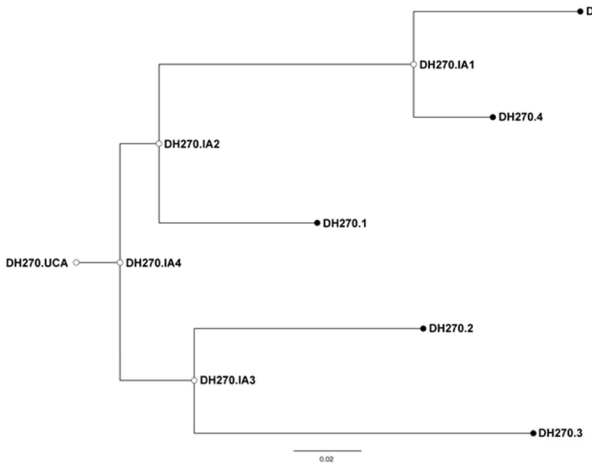
Among these four mutations, the large V1 deletion selected by DH475 was critical for DH270.1 neutralization. This V1 deletion removed a PNG site at position 137. While the hypervariable nature of the V1 loop (which evolves by insertion and deletion, resulting in extreme length heterogeneity, as well as extreme variation in number of PNG sites) complicated the interpretation of direct comparisons between two unrelated HIV-1 strains, it is worth noting that a PNG in this region specified as N137 in donor 17 was shown to be important for regulating affinity maturation of the PGT121 V3 glycan bnAb family, with some members of the lineage evolving to bind (PGT121-123) and others (PGT124) to accommodate or avoid this glycan (*71*).

Section 4.10 – Supplemental Figures and Tables

A

Antibody ID	V_H	D	J_H	Mutation frequency V_H	Mutation frequency V_HDJ_H nt	Mutation frequency V_HDJ_H aa	CDRH3 length	VL	JL	Mutation frequency V_L	Mutation frequency V_LJ_L nt	Mutation frequency V_LJ_L aa	CDRL3 length	Week of isolation	Insertions/ deletions
DH270.UCA	1-2*02	3-22*01	4*02	0.0%	0 (0%)	0 (0%)	20	2-23*02	2*01	0.0%	0 (0%)	0 (0%)	10	-	none
DH270.IA4	1-2*02	3-22*01	4*02	1.4%	4/381 (1.0%)	4/127 (3.1%)	20	2-23*02	2*01	0.7%	2/330 (0.6%)	1/110 (0.9%)	10	-	none
DH270.IA2	1-2*02	3-22*01	4*02	2.1%	7/381 (1.8%)	7/127 (5.5%)	20	2-23*02	2*01	2.1%	6/330 (1.8%)	3/110 (2.7%)	10	-	none
DH270.IA3	1-2*02	3-22*01	4*02	3.5%	16/381 (4.2%)	11/127 (8.7%)	20	2-23*02	2*01	1.4%	4/330 (1.2%)	2/110 (1.8%)	10	-	none
DH270.1	1-2*02	3-22*01	4*02	5.6%	21/381 (5.5%)	18/127 (14.2%)	20	2-23*02	2*01	5.2%	21/330 (6.3%)	11/110 (10.0%)	10	205	none
DH270.IA1	1-2*02	3-22*01	4*02	8.3%	33/381 (8.7%)	22/127 (17.3%)	20	2-23*02	2*01	6.9%	26/330 (7.9%)	14/110 (12.7%)	10	-	none
DH270.2	1-2*02	3-22*01	4*02	10.8%	46/381 (12.1%)	27/127 (21.3%)	20	2-23*02	2*01	3.8%	13/330 (3.9%)	7/110 (6.4%)	10	232	none
DH270.3	1-2*02	3-22*01	4*02	11.8%	48/381 (12.6%)	27/127 (21.3%)	20	2-23*02	2*01	8.3%	27/330 (8.2%)	16/110 (14.5%)	10	205	none
DH270.4	1-2*02	3-22*01	4*02	11.5%	44/381 (11.6%)	28/127 (22.0%)	20	2-23*02	2*01	8.0%	29/330 (8.8%)	16/110 (14.5%)	10	232	none
DH270.5	1-2*02	3-22*01	4*02	11.1%	45/381 (11.8%)	26/127 (20.5%)	20	2-23*02	2*01	11.5%	41/330 (12.4%)	22/110 (20.0%)	10	232	none
DH270.6	1-2*02	3-22*01	4*02	12.9%	47/381 (12.3%)	28/127 (22.0%)	20	2-23*02	2*01	7.6%	39/330 (11.8%)	17/110 (15.5%)	10	234	none

B



C

Amino Acid Sequence Alignment of DH270 Lineage Antibodies V_HDJ_H rearrangements

```
                 10        20        30        40        50        60        70        80        90        100       110       120
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
DH270.UCA        QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQKFQGRVTMTRDTSISTAYMELSRLRSDDTAVYYCARGGWIGLYYDSSGYPNFDYWGQGTLVTVSS
DH270.IA4       ------------------------------D--I--------------------T-R-------------------------------------------------------------------
DH270.IA3       ------------------------------D--I-------------A----TT-R----R---------------------R-------------------V-Y-------S----------
DH270.2         E----E--P-L-E-----------------D--I-------------A----TT-RSSF-RG---------E--V-------R------------KA-Y-A--V-Y------NS----------
DH270.3         ----------L-----------------LSD--V--L----------VA----T--R-ISPR---------T---MNV-----RG--------F-------S--V-Y-Y-----S-------S---
DH270.IA2       ----------M-----------------D--I--------------------T-R-------------------------T----------T-------------------------------
DH270.1         ----------M-------R----------D--I----------P--------ST-R--SP--------------------D-N--T-----M---TT-------S-T----------------
DH270.IA1       ----------M-N---------A-------DF-I----L-----------M--KT-R--N--N-------------G------R--T---------VT----SP-----Y-----H----------
DH270.4         E---------M-N---------A----G--DF-I----L---H--Q----M--KT-R--N--D------L------G------R--T---------VT----SP-----Y-----H------I----
DH270.5         ----------N---------AP------DF-I----L--------L--M--KT-R--QG-N-------------G------RS-T---------VT-A--SD-----Y-----H----------
DH270.6         ---------QM-N---------AP------DF-I--L--------Q----M--QT-R--T-RN------------G------RS-T-----I---TT----S------Y-----H------L----
```

D

Amino Acid Sequence Alignment of DH270 Lineage Antibodies V_LJ_L rearrangements

```
                 10        20        30        40        50        60        70        80        90        100       110
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
DH270.UCA        QSALTQPASVSGSPGQSITISCTGTSSDVGSYNLVSWYQQHPGKAPKLMIYEVSKRPSGVSNRFSGSKSGNTASLTISGLQAEDEADYYCCSYAGSSTVIFGGGTKLTVL
DH270.IA4        ........................Y.....................................................................................
DH270.IA3        ........................Y.....................W.....................................................................
DH270.2          ........................Y.....................I.....QN.....K.....................H.................S....
DH270.3          .FV.....................S.....................N.WA....D..A.............R.......N.F.S.STN.A.............
DH270.IA2        ........................Y.....................Y.....N.....................................................
DH270.1          ......................NY................V..YI....N.....................................IIF.........I
DH270.IA1        ........................Y...KFD.............Y.....N.W.....H.............................FG..A..VC.....V...
DH270.4          ........................Y..AKFD....F.........Y.....N.W.....H.............................FG..A..VC.....V...
DH270.5          LPV.....................IY...KFD......H.......YL....K.W.....H.....................V..........FG..AA.VC.....V...
DH270.6          T.L.....................KY....HD.......Y...V..Y.....N.........................R...........FG..A..VC.....V...
```

Figure 4.S1. Characteristics of DH270 lineage monoclonal antibodies. (A) Immunogenetics of DH270 lineage monoclonal antibodies; **(B)** Phylogenetic relationship of V_HDJ_H rearrangements of the unmutated common ancestor (DH270.UCA) and maturation intermediates DH270.IA1 through DH270.IA4 inferred from mature antibodies DH270.1 through DH270.5. DH270.6 was not included and clusters close to DH270.4 and DH270.5 as shown in **Figure 1**. DH270.6 addition did not modify the overall topology of the tree. Importantly, it did not affect the inference of DH270.UCA and DH270.IA4 (details in Supplemental Text). **(C)** Amino acid alignment of the V_HDJ_H rearrangements of the inferred UCA and intermediate antibodies and DH270.1 through DH270.6 mature antibodies. **(D)** Amino acid alignment of V_LJ_L rearrangements rearrangements of the inferred UCA and intermediate antibodies and DH270.1 through DH270.6 mature antibodies. For DH270.6, all experimental data presented in this manuscript were obtained using the light chain sequence reported here. The light chain sequence of DH270.6 was subsequently revised to amino acids Q and A in positions 1 and 3 (instead of T and L). This difference did not affect neutralization and binding of DH270.6.

	Neutralization IC_{50} µg/ml									
	AC13.8		**PVO.4**		**TRO.11**		**AC10.029**		**RHPA.4259**	
	wt	N332A	wt	N332A	wt	N332A	wt	N332A	wt	N332A
DH270.UCA	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
DH270.IA4	>50	>50	42	>50	43	>50	>50	>50	>50	>50
DH270.IA3	>50	>50	>50	>50	**0.2**	>50	>50	>50	**6.6**	>50
DH270.IA2	>50	>50	>50	>50	**0.1**	>50	>50	>50	**6.4**	>50
DH270.1	>50	>50	**0.2**	>50	**0.08**	>50	**1.9**	>50	**0.2**	>50
DH270.IA1	>50	>50	**0.07**	>50	**0.05**	32.4	**<0.02**	>50	**0.04**	>50
DH270.2	21	>50	**0.3**	>50	**0.06**	>50	**0.3**	>50	**0.1**	>50
DH270.3	>50	>50	23	>50	**0.3**	>50	43	>50	42	>50
DH270.4	15	>50	**0.1**	>50	**0.04**	14	**<0.02**	>50	**0.05**	>50
DH270.5	41	>50	**0.1**	>50	**0.07**	>50	**<0.02**	>50	**0.04**	>50
DH270.6	**1.4**	>50	**0.03**	>50	**0.02**	>50	**<0.02**	>50	**<0.02**	>50

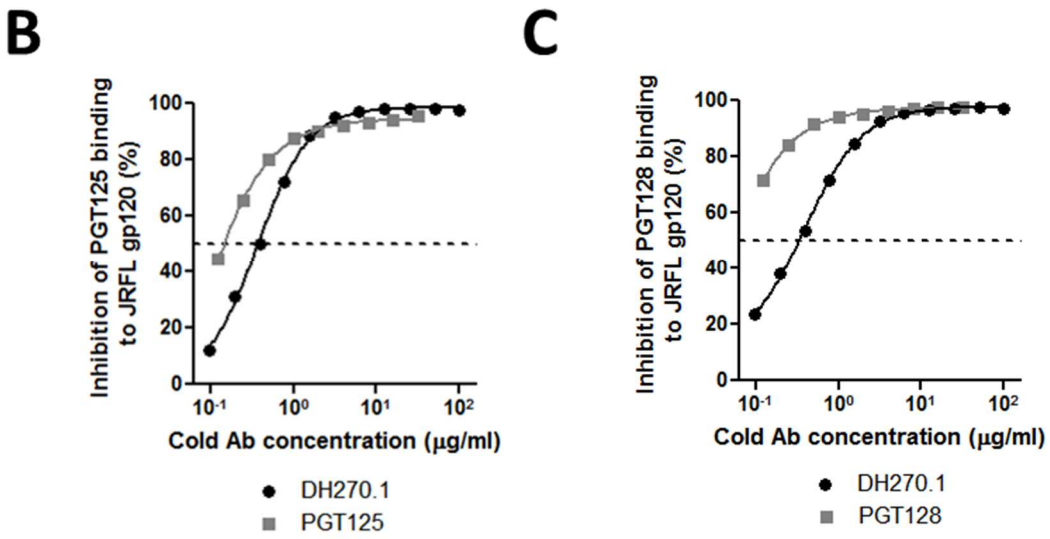
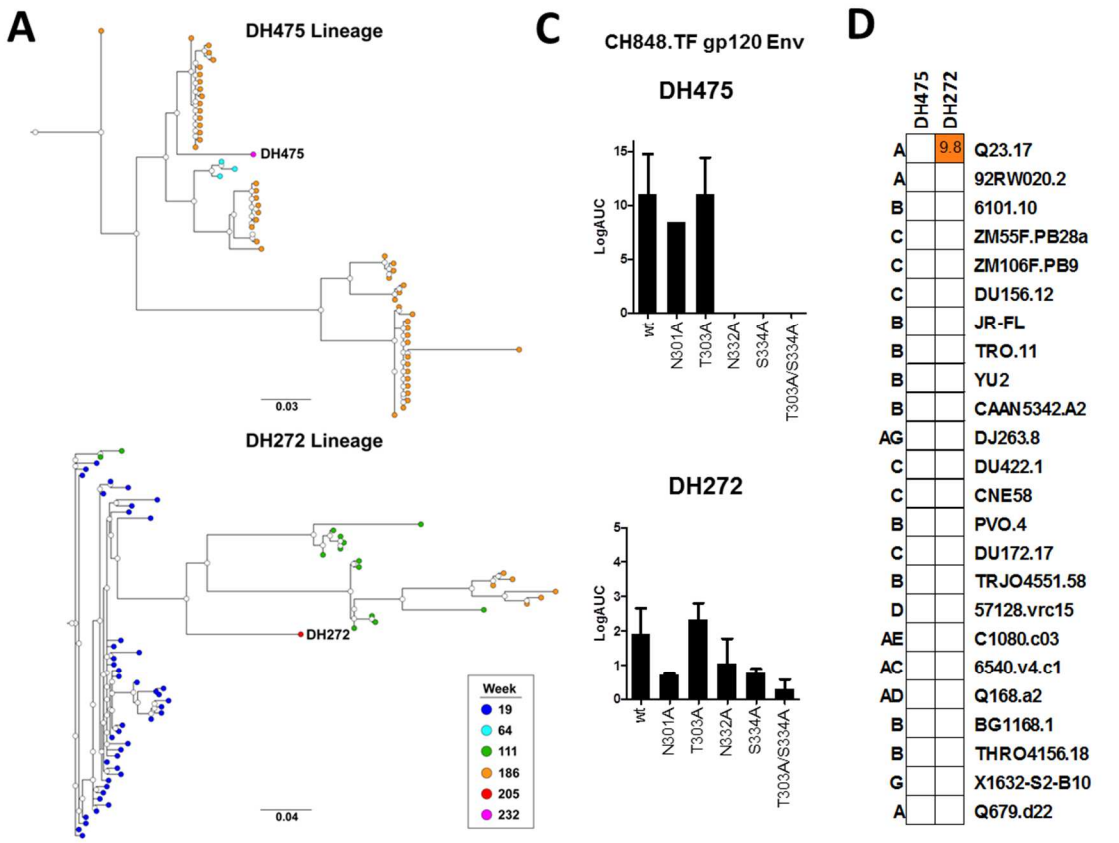


Figure 4.S2. DH270 lineage displays a N332-dependent V3 glycan bnAb functional profile. (A) DH270 antibody lineage neutralization of five HIV-1 pseudoviruses and

respective N332A mutants. Data are expressed as IC_{50} μg/ml. Positivity <10 μg/ml is shown in bold. (B,C) DH270.1 ability to compete gp120 Env binding of V3 glycan bnAbs PGT125 and PGT128. DH270.1 (cold Ab) inhibited binding of PGT125 (B) and PGT128 (C) to JRFL gp120 Env with IC_{50} = 0.4μg/ml for both (black line). Inhibition by cold PGT125 or PGT128 (grey line) was used as control (see Methods). Dotted line indicates 50% inhibition.



B

Antibody ID	V_H	D	J_H	V_H mutation frequency	CDRH3 length	V_L	J_L	κ/λ	V_L mutation frequency	CDRL3 length	Week of isolation	Insertions/ deletions
DH475	3-23*01	2-8*02	3*01	10.1%	19	4-69*02	1*01	λ	10.9%	9	232	none
DH272	1-2*02	3-3*01	4*02	14.9%	17	2-30*02	2*02	κ	5.1%	9	205	-6 FRH3

Figure 4.S3. DH475 and DH272 are strain-specific, N332-glycan dependent antibodies. (A) Phylogenetic trees of DH475 (top) and DH272 (bottom) clonal lineages. External nodes (filled circles) representing V_HDJ_H observed sequences retrieved from cultured and sorted memory B cells (labeled) or NGS antibody sequences (unlabeled) are colored according to time point of isolation. Internal nodes (open circles) represent inferred ancestral intermediate sequences. Branch length estimates units are nucleotide substitution per site. (B) Immunogenetics of DH475 and DH272 monoclonal antibodies; (C) Binding of DH475 (top) and DH272 (bottom) monoclonal antibodies to wild-type CH848TF gp120 Env (wild-type (wt), on the x-axis, and mutants with disrupted the 301 and/or 332 N-linked glycosylation sites. Results are expressed as LogAUC. (D) Heterologous neutralization profile of DH475 and DH272 monoclonal antibodies expressed as IC_{50} µg/ml on a multiclade panel of 24 viruses. White square indicates IC_{50} > 50 µg/ml, the highest antibody concentration tested. Clades are reported on the left and

virus identifiers on the right. DH475 neutralized no heterologous viruses and DH272 neutralized one Tier 1 heterologous virus.

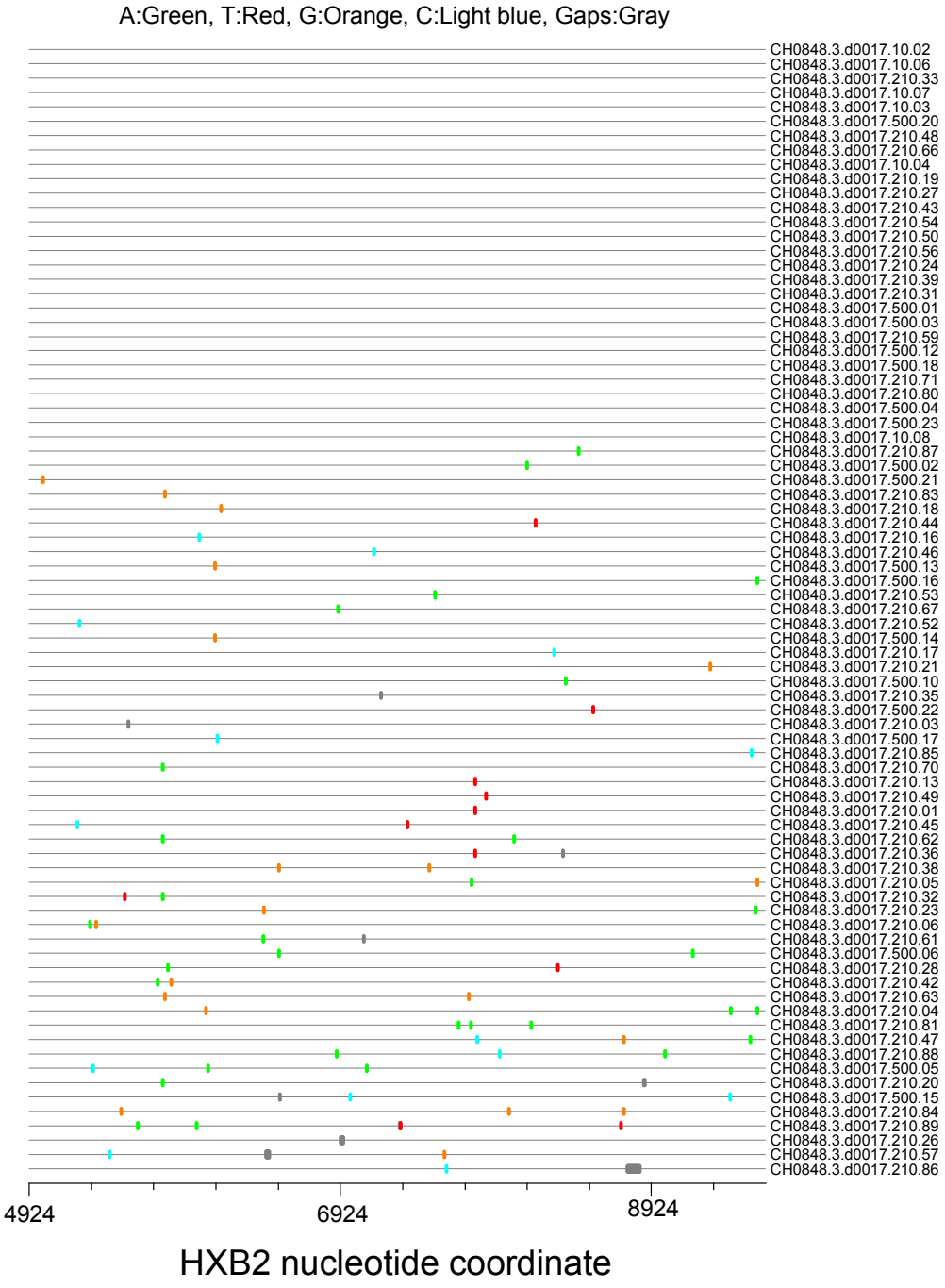


Figure 4.S4. CH848 was infected by a single transmitted founder virus. 79 HIV-1 3' half single genome sequences were generated from screening timepoint plasma. Depicted

is a nucleotide Highlighter plot (http://www.hiv.lanl.gov/content/sequence/HIGHLIGHT/HIGHLIGHT_XYPLOT/highlighter.html) demonstrating that this subject was infected with a single transmitted founder virus. Horizontal lines represent single genome sequences and tic marks denote nucleotide changes relative to the inferred TF sequence (key at top, nucleotide position relative to HXB2). Sequences exhibited random diversification from the TF virus as demonstrated by analysis using the Poisson Fitter v2 tool ($p_{\text{goodness of fit}} = 0.806$)(*33*). Transmission likely occurred approximately 17 days prior to screening (CI: 14-19).

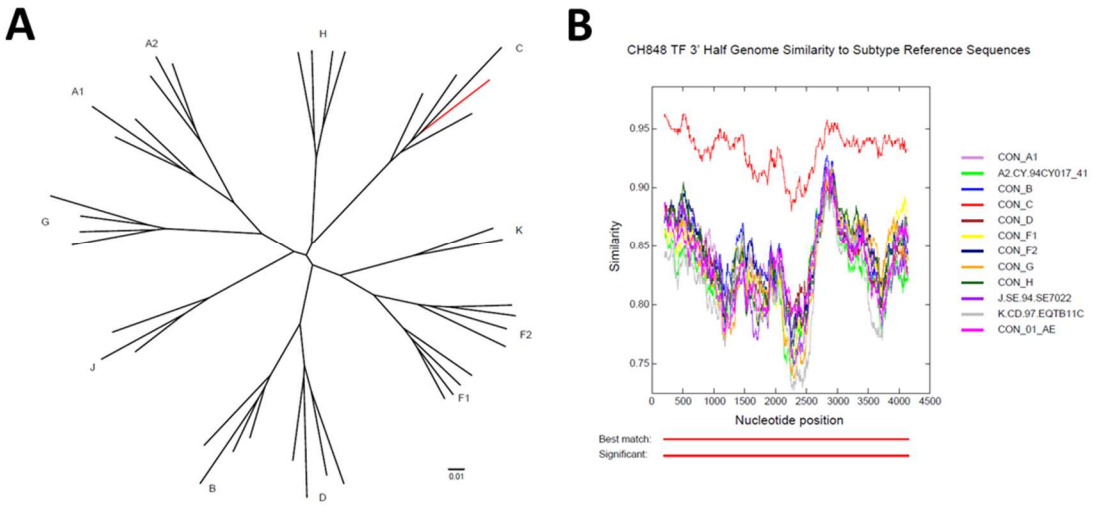


Figure 4.S5. CH848 was infected by a subtype C virus. (A) PhyML (*65*) was used to construct a maximum likelihood phylogenetic tree comparing the CH848 transmitted founder virus to representative sequences from subtypes A1, A2, B, C, D, F1, F2, G, H, and K (substitution model: GTR+I+G, scale bar bottom right). The CH848 TF sequence (red) clusters with subtype C viruses, consistent with the 98.6% prevalence of this subtype in Malawi, where CH848 was enrolled during acute seroconvertion (**www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp**). (B) The TF virus was typed using the Rega HIV-1 subtyping tool and the Recombination Identification Program (RIP)(*72*). Shown similarity plot was generated using RIP. Similarity to each subtype reference sequence is plotted on the *y*-axis and nucleotide position is plotted the *x*-axis (window size = 400 nt, significance threshold = 0.95, key to right). The two bars below the *x*-axis indicate which reference sequence is most similar to the CH848 TF sequence ("Best Match") and whether this similarity is statistically significant relative to the second best match ("Significant"). Both analyses demonstrated that the CH848 3' half genome represents a subtype C sequence with no evidence of inter-subtype recombination.

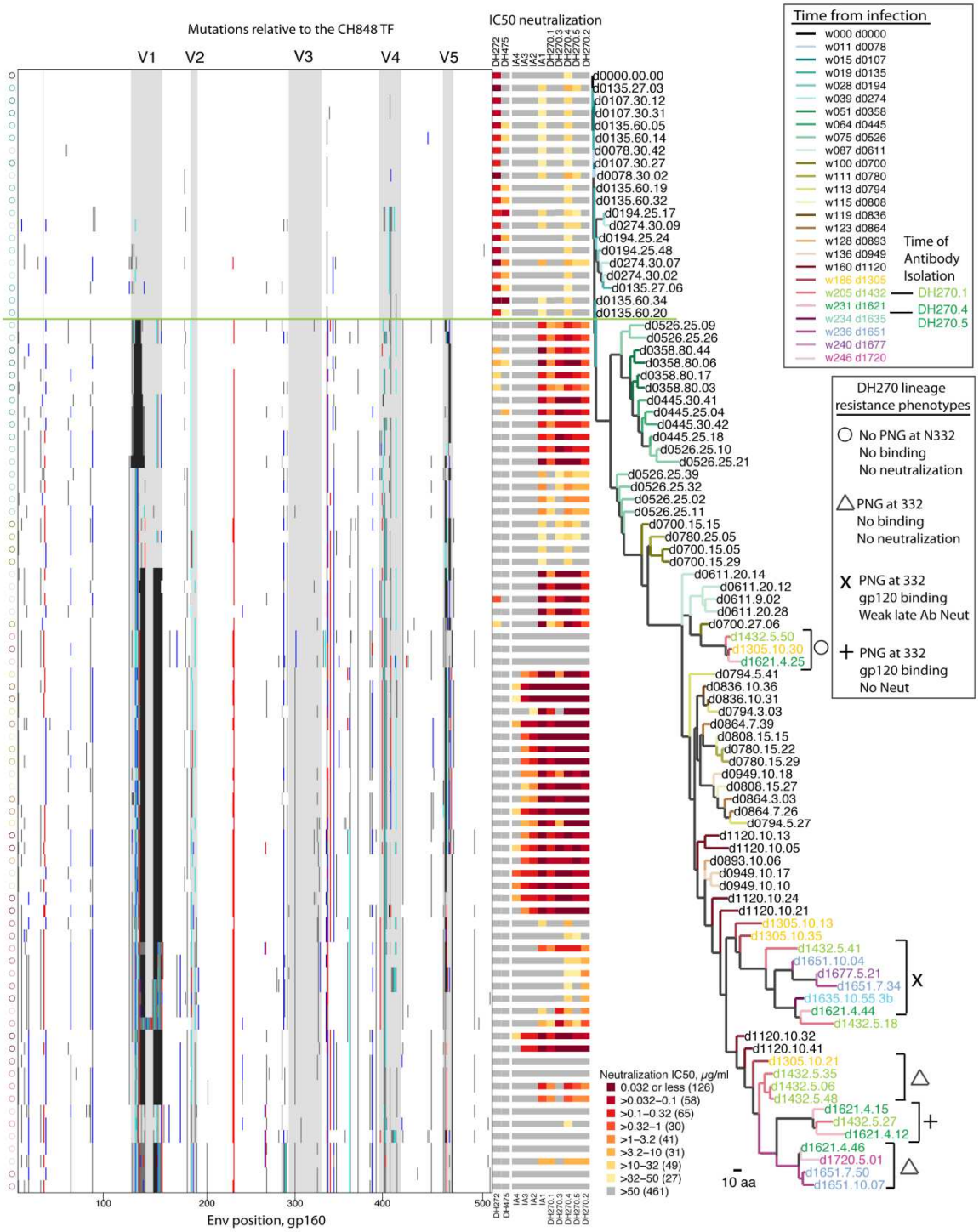


Figure 4.S6. Co-evolution of CH848 autologous virus and N332-dependent V3 glycan antibody lineages DH272, DH475 and DH270. Mutations relative to the CH848 transmitted founder virus in the alignment of CH848 sequences with accompanying neutralization data (Insertion/deletions = black. Substitutions: red = negative charge; blue

= positive charge; cyan = PNG sites)(*50*). The transition between DH272/DH475 sensitive and DH270 lineage sensitive virus immunotypes at day 356 (week 51) is indicated by the green line. Time points after week 186 (for which antibody isolation and NGS provided evidence of DH270 lineage presence) are highlighted. At this juncture four viral clades are established that co-exist and persist, each clade characterized by viruses with a distinct immunologically-defined DH270 resistance phenotypes (distinct patterns were defined by loss of N332 PNG site, gp120 binding and neutralization resistance). The sequences that represent these clades can be tracked through the full tree of all CH848 sequences (fig. S9).

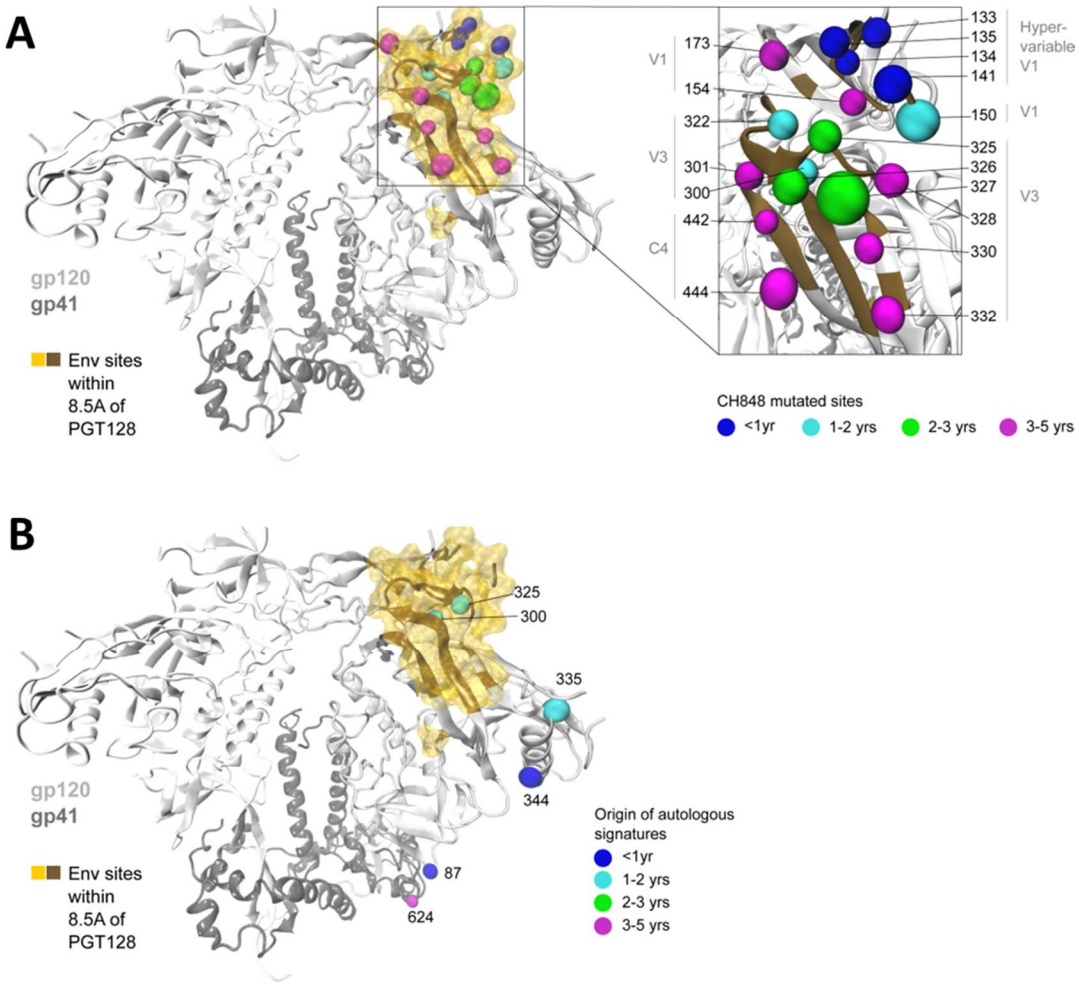


Figure 4.S7. Mutations in CH848 Env over time. (A) Variable positions that are close to the PGT128 epitope in a trimer structure (PDB ID: 4TVP)(*13*) are represented by spheres color-coded by the time post-infection when they first mutate away from the CH848 TF sequence. The PGT128 antibody structure (PDB ID: 5C7K)(*73*) was used as a surrogate for DH270, as a high resolution structure is not yet available for DH270. Env positions with either main chain, side chain or glycans within 8.5Å of any PGT128 heavy atom were considered to be proximal to PGT128. All such Env positions are shown in yellow surface and brown ribbon representations. Mutations in the V3 region do not begin to accrue until after the first year of infection when the DH270 lineage was first detected; V1 mutations arise early. (B) Same as (A) for mutating Env sites that were autologous antibody signatures of antibody sensitivity and resistance.

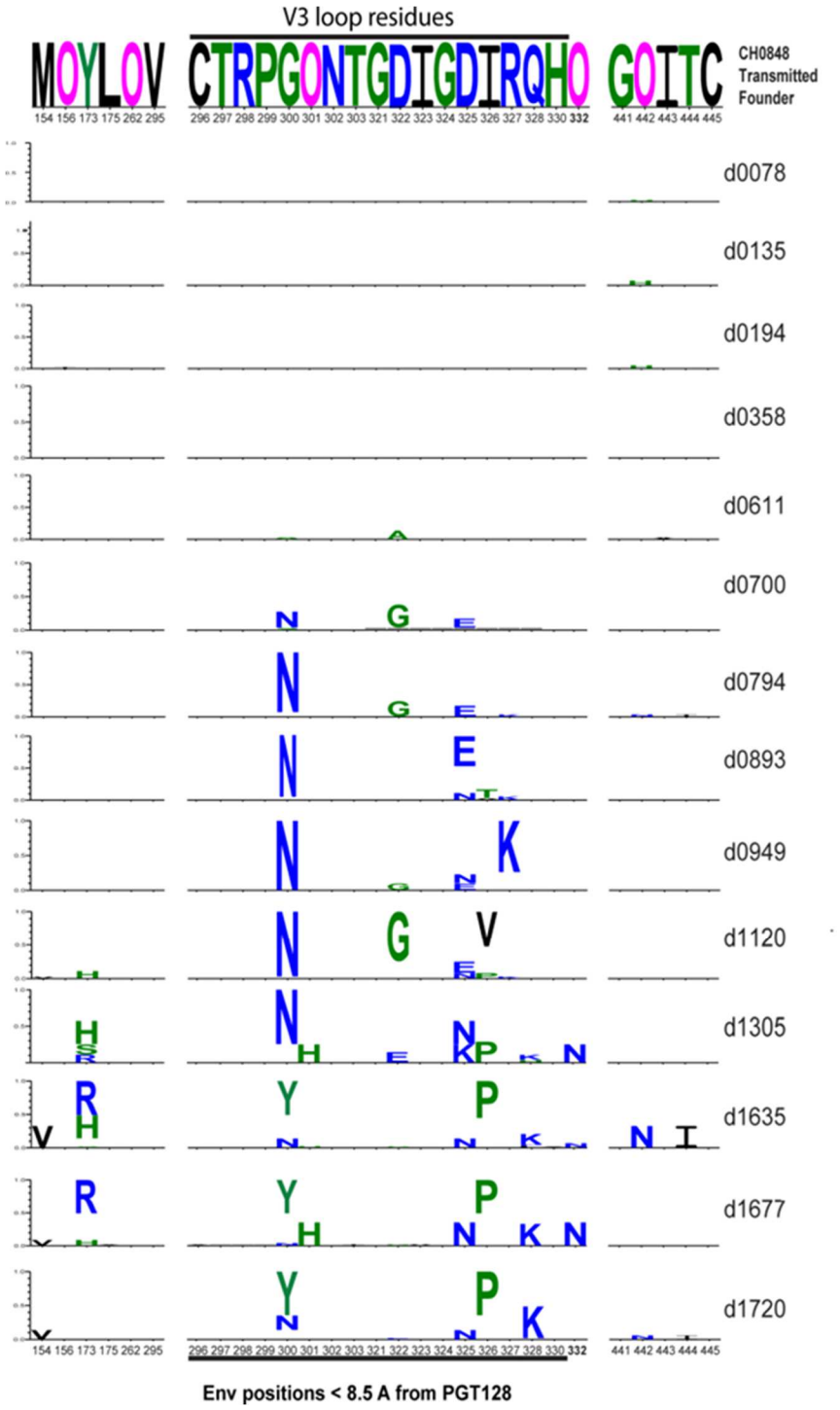


Figure 4.S8. Accumulation of amino acid mutations in CH848 virus over time. This figure shows all of the readily aligned positions near the contact site of V3 glycan antibodies in Fig. S7, (excluding amino acids that are embedded in the V1 hypervariable regions). The magenta O is a PNG site, whereas an N is an Asn that is not embedded in a glycosylation site. The logo plots represent the frequency of amino acids at each position,

and the TF amino acid is left blank to highlight the differences over time.

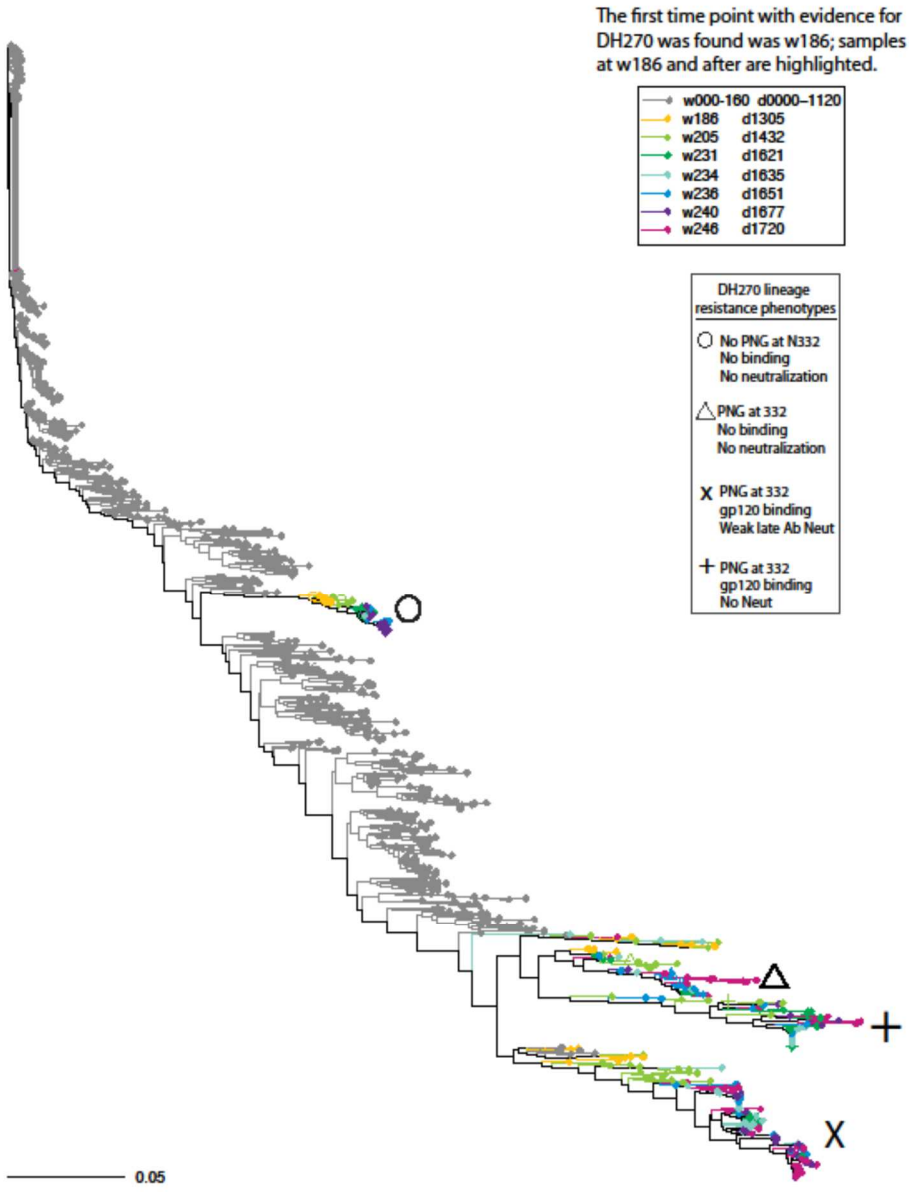


Figure 4.S9. Multiple viral lineages persisted during bnAb development. 1,223 Env protein sequences translated from single genome sequences were used to generate a Maximum Likelihood phylogenetic tree (*65*) rooted on the transmitted founder sequence. Sequences sampled prior to the development of Tier 2 heterologous breadth (week 186) are shaded in grey and sequences from after week 186 are highlighted using the color scheme from **fig. S6**. By week 186, the quasispecies resolved into at least four distinct lineages with different DH270 binding and neutralization profiles. Envs indicated with a circle simultaneously lack the PNG site (NXS/T) at HXB2 position 301 and shift the

PNG site at 332 to 334. Loss of critical V3 glycan bnAb-binding amino acids conferred complete resistance to DH270 lineage Abs, and by week 246 this virus was undetected. Binding and neutralization assays using representative members of each of these lineages revealed that certain variants escape DH270 lineage pressure by abrogating antibody binding (circle and triangle) while other maintaining binding while disrupting neutralization (X and cross).

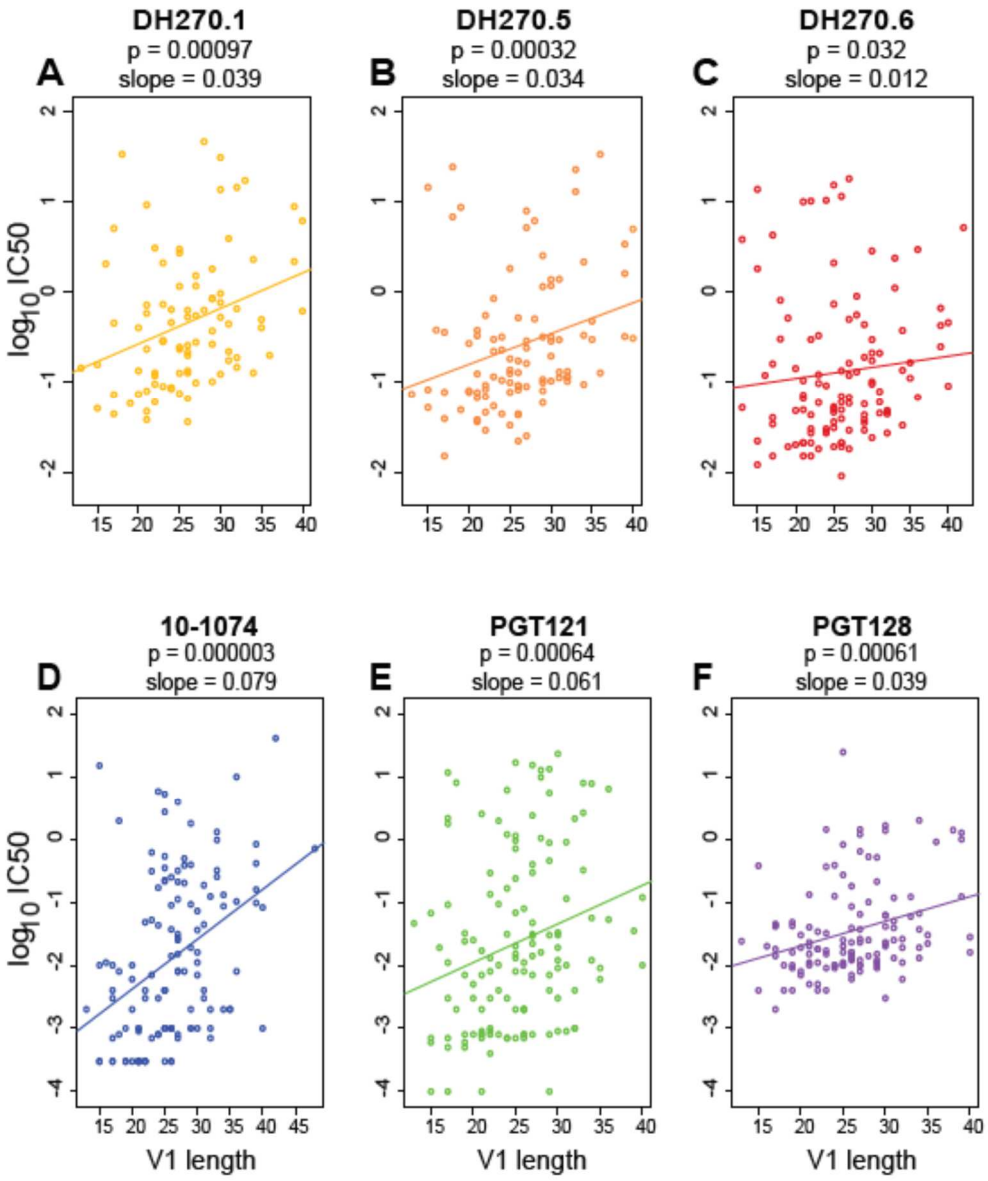
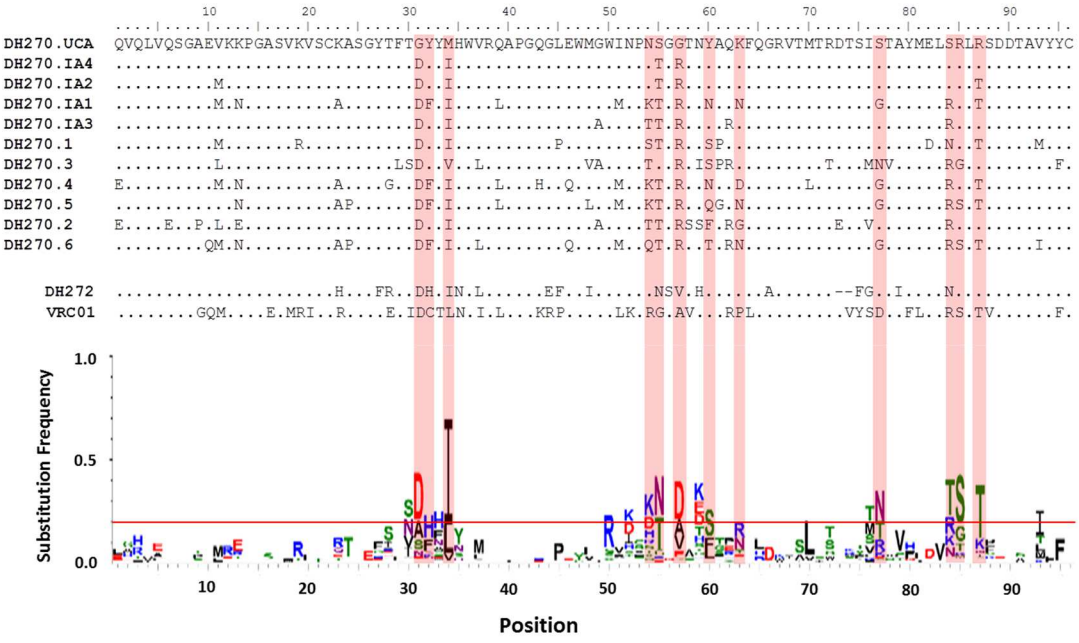


Fig. 4.S10. Inverse-correlation between the potency of V3 glycan broadly neutralizing antibodies and V1 length shown for the full panel of 207 viruses. **(**A-C**)** In Fig. 2D, the impact of V1 loop length on potency for the panel of 24 viruses is shown. The impact of V1 length on potency and breadth diminishes as the antibody develops. The same phenomenon is shown for each of the three DH270 lineage antibodies that were evaluated using the 207 virus heterologous panel. Correlation p-values are non-parametric two sided, Kendall's tau. Slopes are of the linear regression. The impact of V1 length is over 3 fold greater on DH270.1 than DH270.6, but still evident and statistically supported in DH270.6. In this figure you can also see the details of the increase in potency over the full set of Envs as the lineage evolves; the medians were provided in Fig. 2C, bottom left panel, to enable a comparison with the mutations in the antibody. The impact of V1

length on potency and breadth diminishes as the antibody develops. (D-E). Inverse-correlation between the potency of other V3 glycan antibodies 10-1074, PG121, PG128 and V1 length. 10-1074 was the most impacted by V1 loop length.

A



B

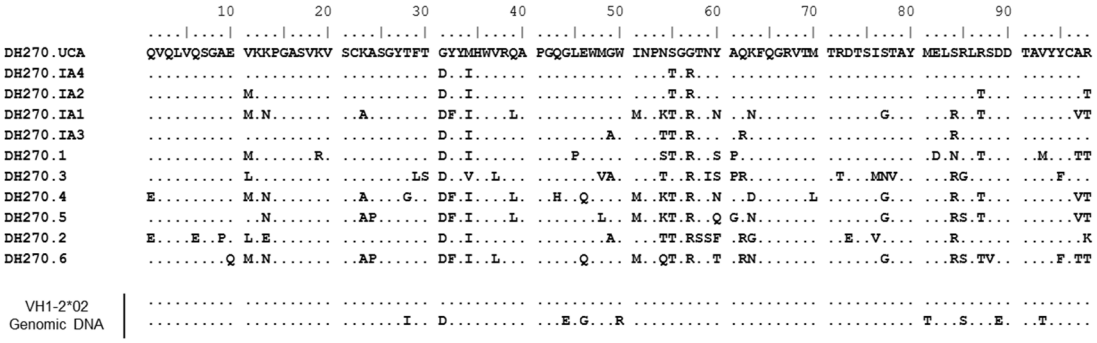


Figure 4.S11. Role of V_H1-2*02 intrinsic mutability in determining DH270 lineage antibody somatic hypermutation. (A) The sequence logo plot shows the frequency of V_H1-2*02 amino acid (aa) mutations from germline at each position, calculated from an alignment of 10,995 V_H1-2*02 reads obtained from 8 HIV-1 negative individuals by NGS (*35*). To distinguish genuine mutations from error introduced during PCR amplification, only sequences that replicated across two independent Illumina experiments were included in this analysis. Frequency of mutated aa at each position is shown in the logo plot. In this dataset, 20 positions mutated frequently (>20%, indicated

by the red line). The V_H aa sequences of DH270 lineage antibodies, DH272 and VRC01 are aligned on the top. In the DH270 lineage, 12 of these 20 positions were also frequently mutated (red vertical stripes). Mutations of the DH270 lineage antibodies were defined as "frequent" when they occurred in more than 25% of the isolated antibodies (inferred intermediate V_H sequences were excluded from this analysis) Of these 12 positions, 11 mutated to one of the two most frequent aa mutated in non-HIV-1 V_H1-2*02 sequences (identity conformity). G57R was the lone exception. DH272 mutated in 6 of these 12 positions and VRC01 mutated in 11 out of 12 positions. (B) V_H aa encoded by V_H1-2 sequences from genomic DNA aligned to DH270 lineage antibodies aa sequences (see "Sequencing of germline variable region from genomic DNA" in methods).

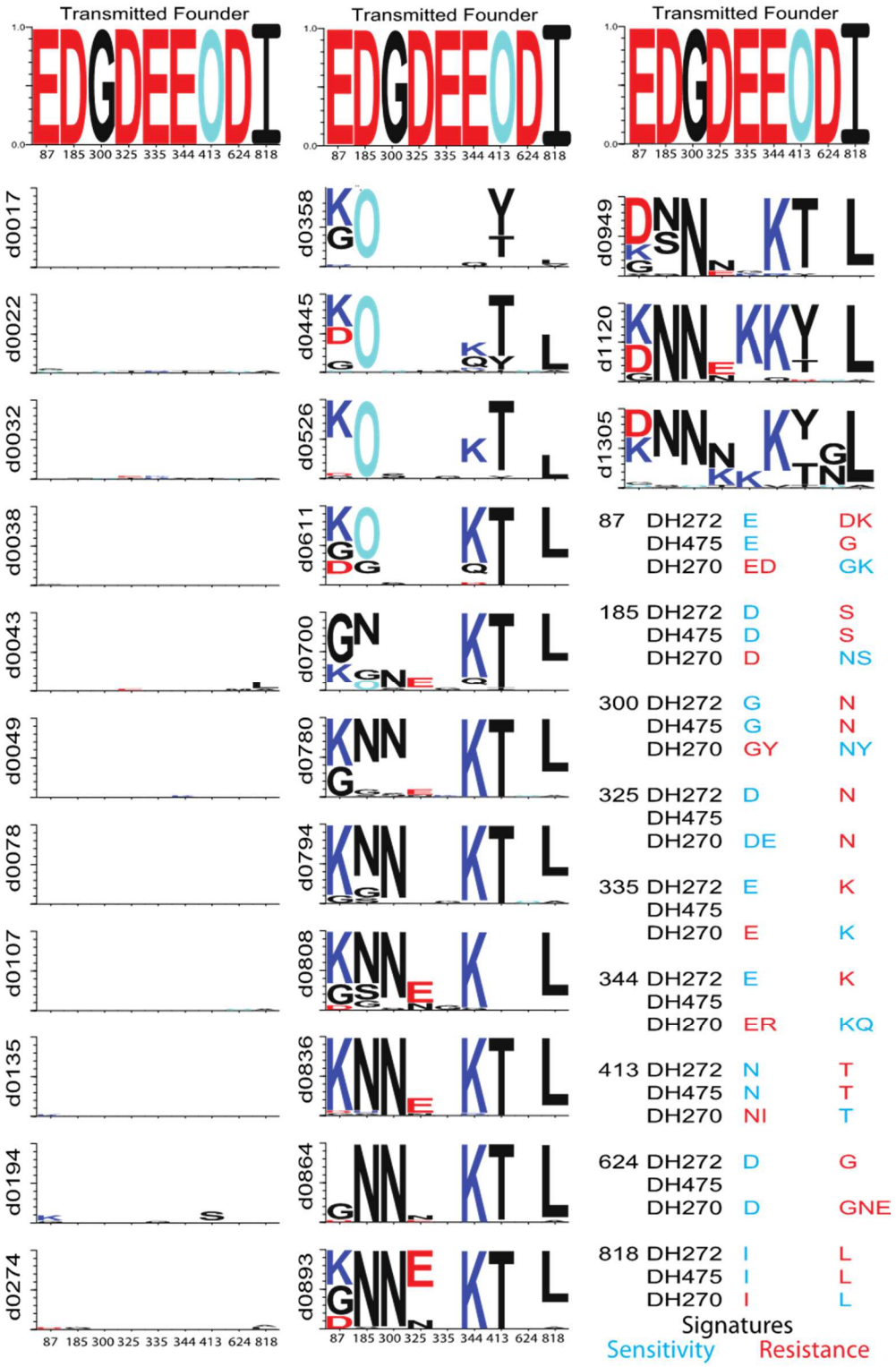


Figure 4.S12. Virus signature analysis. Logo plots represent the frequency of amino acids mutations in CH848 virus quasispecies from transmitted founder at indicated

positions over time. Red indicates a negatively charged amino acid, blue positive, black neutral; the light blue O is a PNG site. The signatures outlined in detail in Additional Dataset S4 are summarized in the bottom right column where a red amino acid is associated with resistance to the antibody on the right, a blue amino acid with sensitivity. Sites 325 and 624 are associated with mutual acquisition of resistance whereas the other seven are candidates for cooperativity.

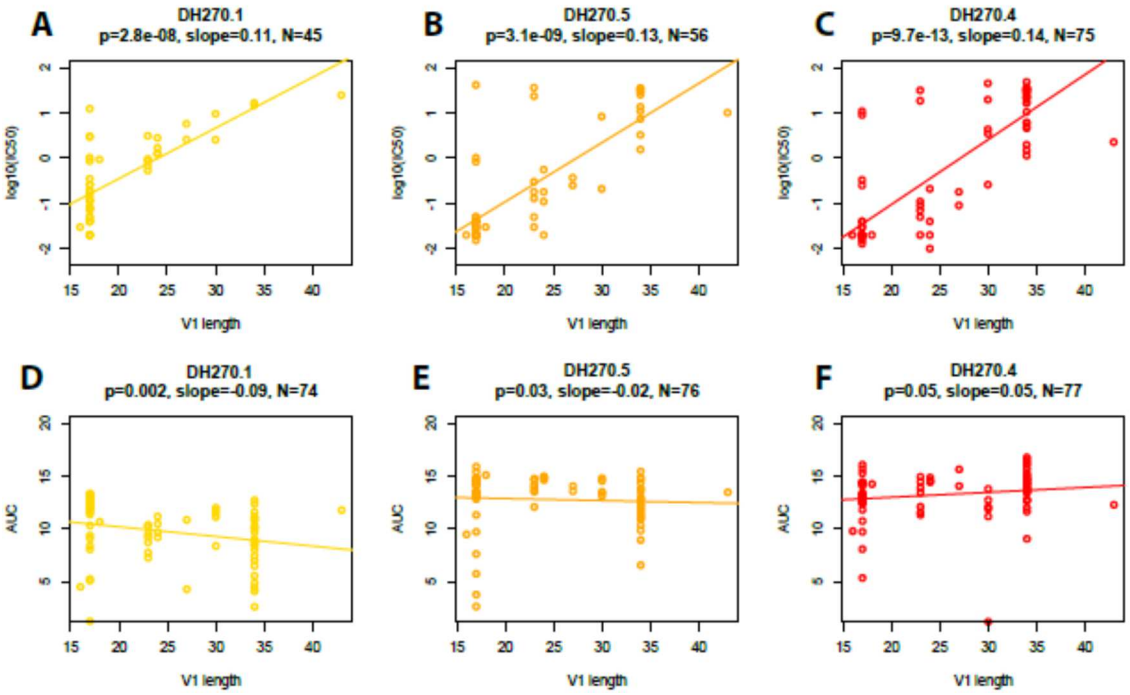


Figure 4.S13. Autologous Env V1 length associations with DH270 lineage neutralization (A-C) and gp120 binding (D-F). Eighty-two virus Envs (the subset from Additional dataset S3 that were assayed for both neutralization and binding) were evaluated. The 3 Envs that had lost the PNG site at N332 were not included, as they were negative for all antibodies tested independently of V1 length. Only points from positive results were plotted: IC_{50} <50 μg/ml for neutralization in panels A-C, and AUC >1 for binding in panels D-F. DH270.1 gp120 Env binding is slightly negatively impacted by longer loops, but most Envs can be bound and the association with V1 length is essentially lost for DH270.5 and DH270.4. In contrast, DH270 lineage neutralization potency is profoundly impacted by V1 loop length. The ability to detectably neutralize autologous CH848 viruses increases as the lineage develops: N is the number of positive sample and ranges from N=45 (55%) for DH270.1 up to N=75 (91%) for DH270.4. These gains are mostly among viruses with long loops. The slope and relationship is most profound for DH270.4, because the viruses with short V1 loops are more potently neutralized (shifting points on the left hand side of the plot down) and the viruses with long V1 loops - which were not neutralized by earlier members of the DH270 lineage - are weakly neutralized.

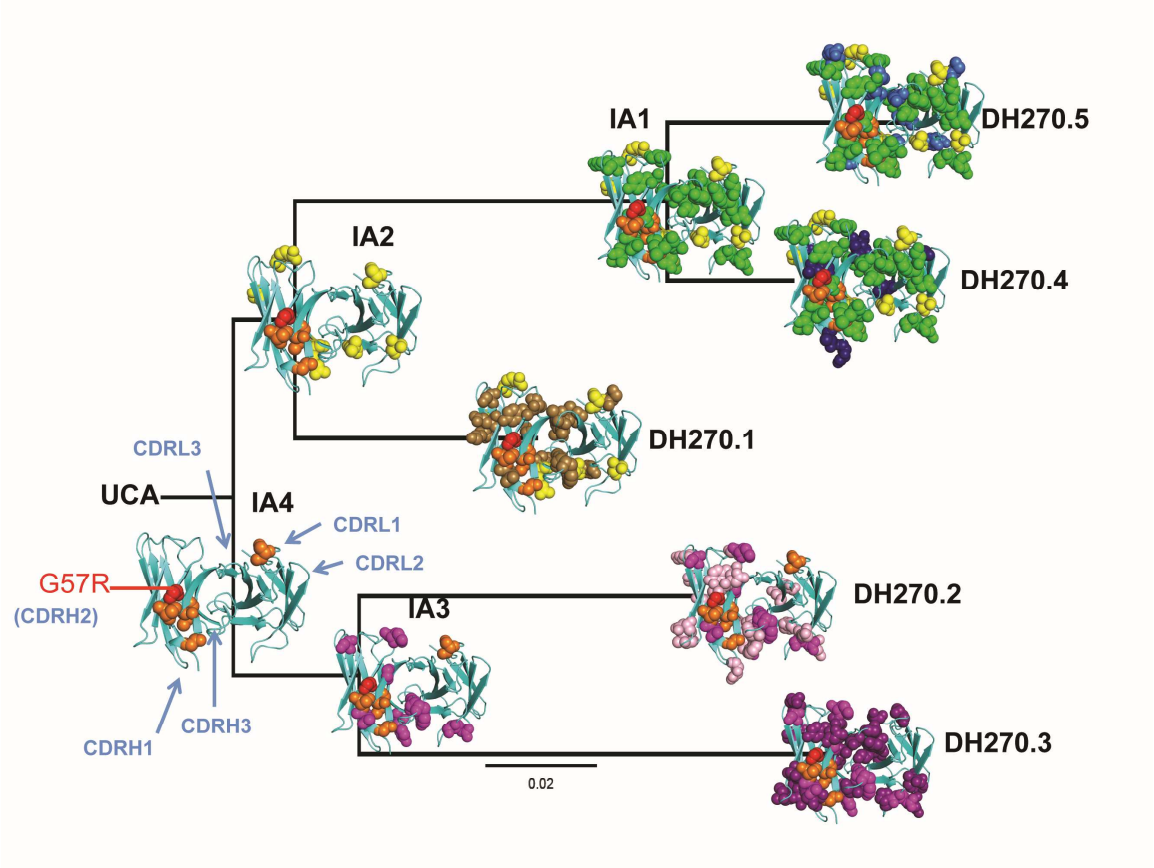


Figure 4.S14. Accumulation of mutation in DH270 lineage antibodies. Mutations are highlighted as spheres on the Fv region of each antibody, where the CDR regions, labeled on the backbone of the UCA, face outward. The G57R mutation is shown in red; the other mutations incurred between the UCA and IA4 are shown in orange. Mutations between intermediates are colored as follows: between IA2 and IA4, yellow; between IA1 and IA2, green; between IA3 and IA4, magenta. Mutations between the late intermediates and DH270.1, DH270.2, DH270.3, DH270.4, and DH270.5 are in brown, light purple, dark purple, blue, and dark blue, respectively.

A

```
UCA1_HC      QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNY
UCA3_HC      QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNY
UCA4_HC      QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMHWVRQAPGQGLEWMGWINPNSGGTNY
             ************************************************************

UCA1_HC      AQKFQGRVTMTRDTSISTAYMELSRLRSDDTAVYYCATGGWIGLYYDSSGYPNFDYWGQG
UCA3_HC      AQKFQGRVTMTRDTSISTAYMELSRLRSDDTAVYYCARGGWISLYYDSSGYPNFDYWGQG
UCA4_HC      AQKFQGRVTMTRDTSISTAYMELSRLRSDDTAVYYCARGGWIGLYYDSSGYPNFDYWGQG
             ********************************** ****.*****************

UCA1_HC      TLVTVS
UCA3_HC      TLVTVS
UCA4_HC      TLVTVS
             ******
```

B

```
UCA3_LC      QSALTQPASVSGSPGQSITISCTGTSSDVGSYNLVSWYQQHPGKAPKLMIYEVSKRPSGV
UCA1_LC      QSALTQPASVSGSPGQSITISCTGTSSDVGSYNLVSWYQQHPGKAPKLMIYEVSKRPSGV
             ************************************************************

UCA3_LC      SNRFSGSKSGNTASLTISGLQAEDEADYYCCSYAGSSTVIFGGGTKLTVL
UCA1_LC      SNRFSGSKSGNTASLTISGLQAEDEADYYCCSYAGSSIILFGGGTKLTVL
             ********************************** ::**********
```

C

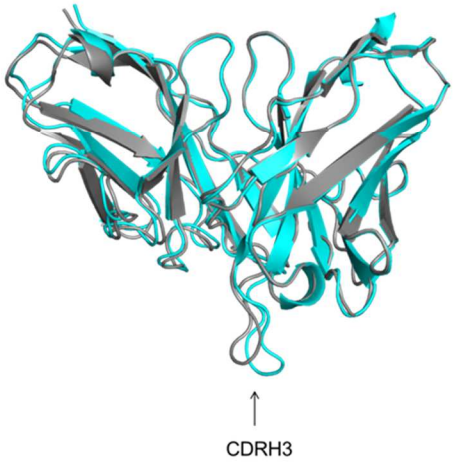


Figure 4.S15. UCA Differences. Sequence alignments of UCA3 and UCA1. (A) Heavy chains and (B) light chains, whose structures were obtained in this study, are aligned with UCA4, the germline antibody for the DH270 lineage (DH270.UCA). The UCA3 and UCA4 light chains are identical. Asterisks indicate positions in which the amino acids are the same. Colon ":", period "." And blanks " " correspond to strictly conserved, conserved and major differences, respectively. (C) Superposition of UCA3 (cyan) and

UCA1 (gray). Structural differences in CDR regions are indicated with an arrow.

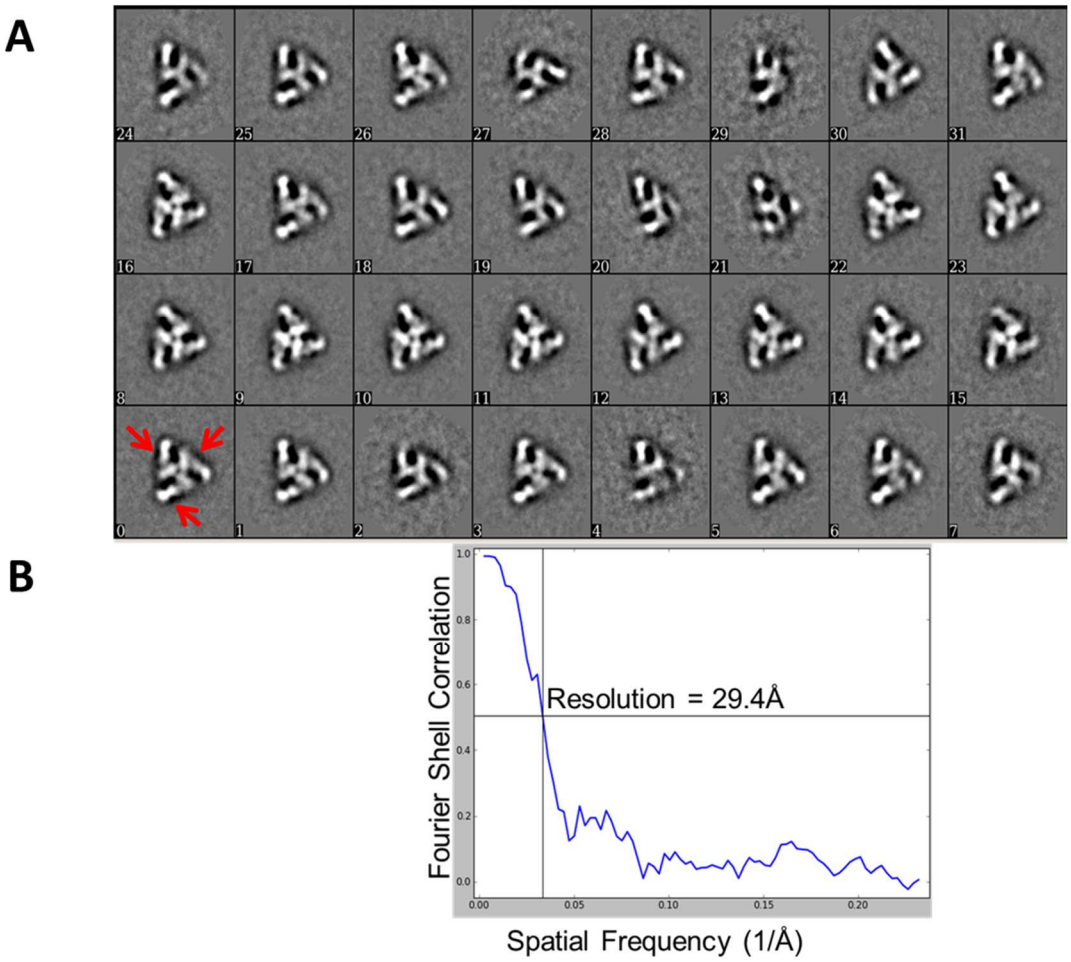


Figure 4.S16. Negative stain EM of DH270 Fab in complex with the 92BR SOSIP.664 trimer. (A) 2D class-averages of the complex. Fabs are indicated with a red arrow. (B) The Fourier shell correlation curve for the complex, along with the resolution determined using FSC = 0.5.

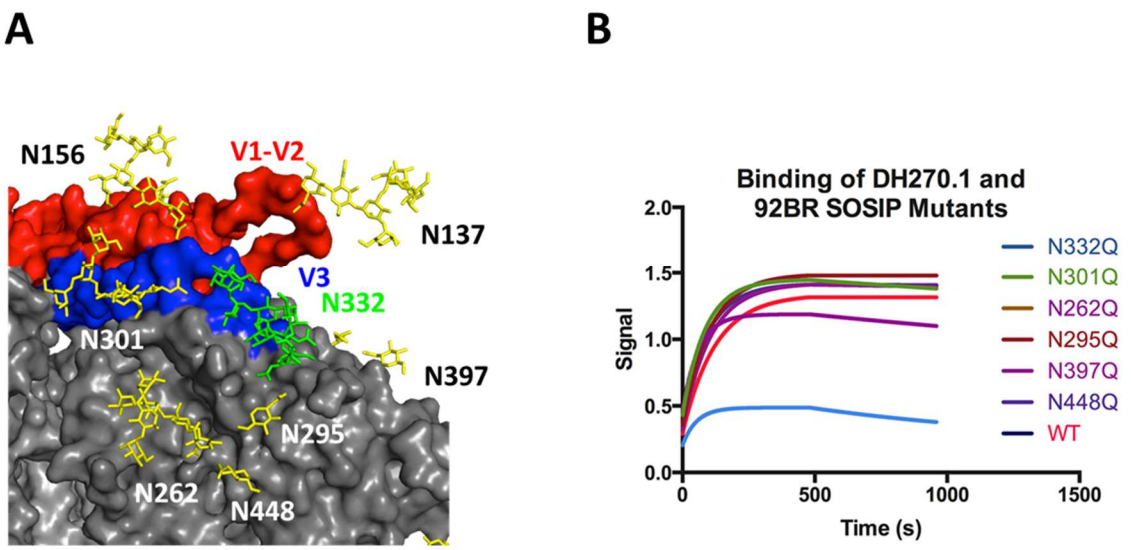


Figure 4.S17. Binding Kinetics. DH270.1 and 92BR SOSIP.664. (A) Glycans forming a "funnel" are shown on the surface of the trimer. V1-V2 and V3 loops are colored red and blue. (B) Association and dissociation curves, using biolayer interferometry, against different 92BR SOSIP.664 glycan mutants.

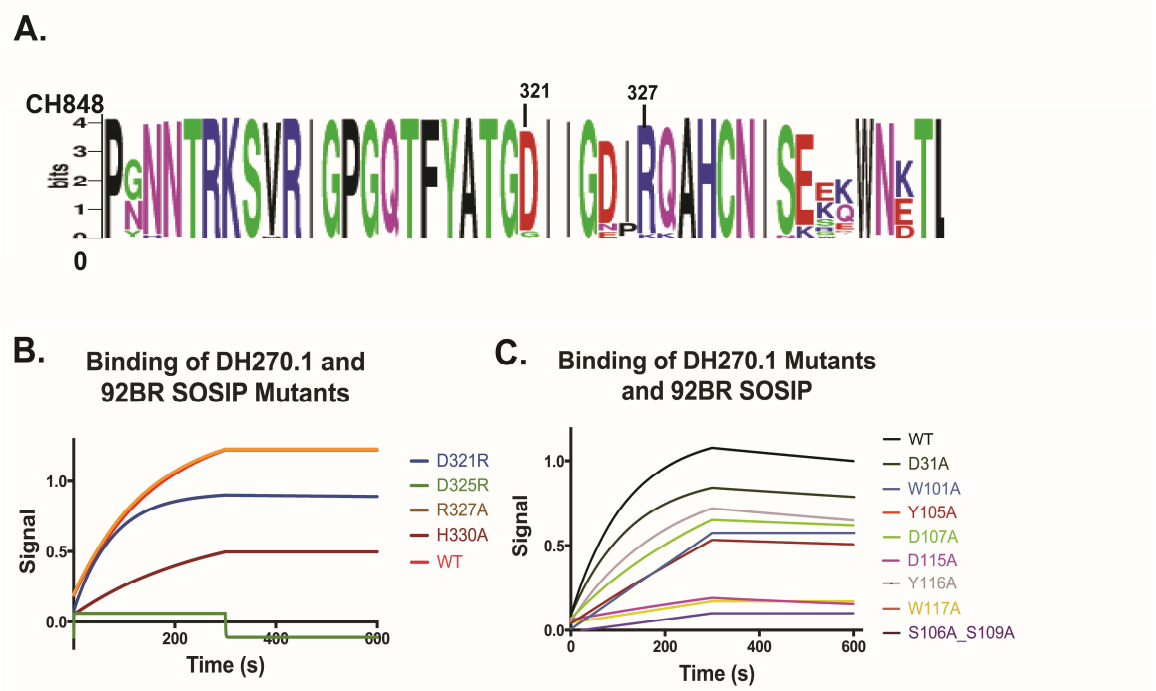
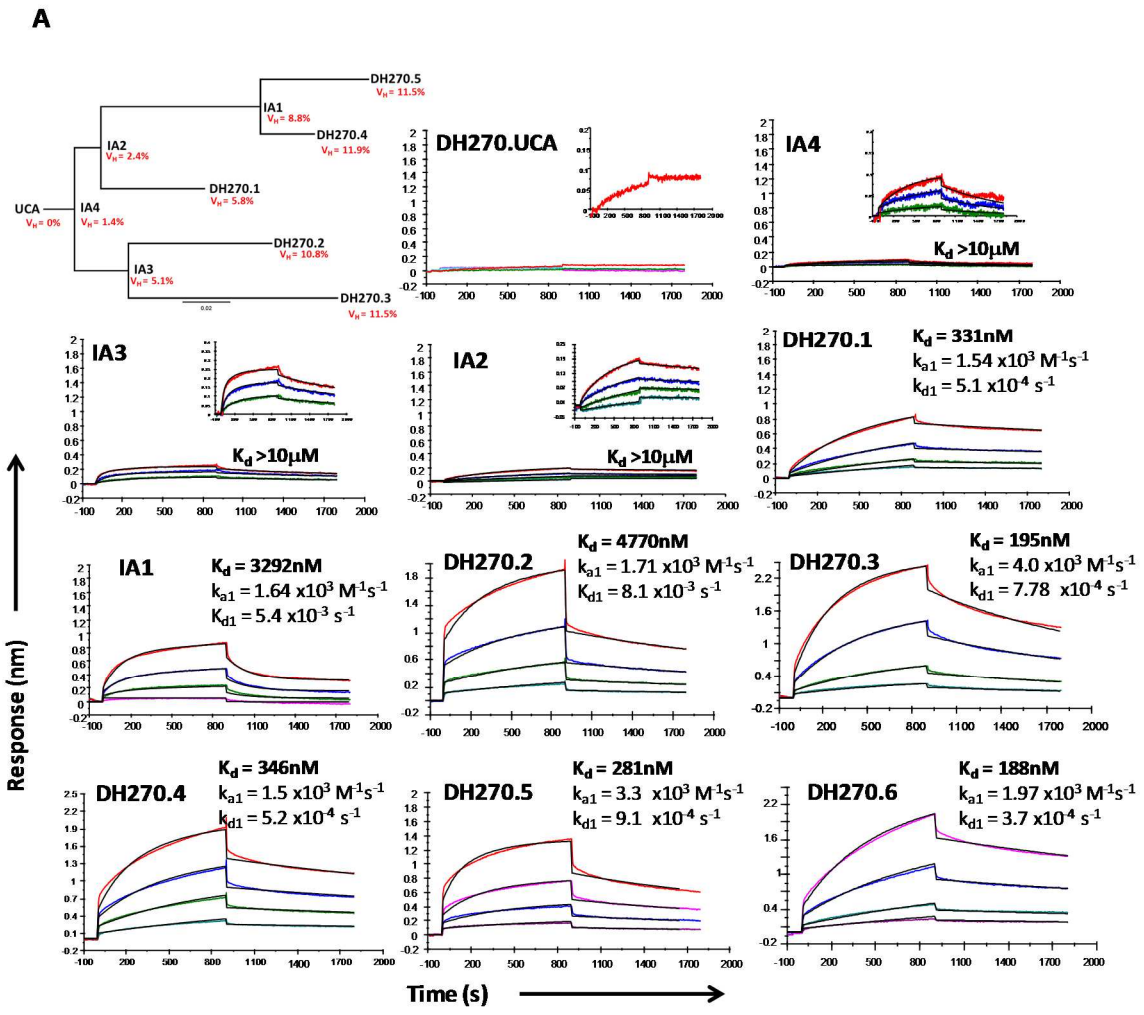


Figure 4.S18. Binding Kinetics: DH270.1 and 92BR SOSIP.664. (A) Sequence Logo of the V3 region of CH848 autologous viruses are shown. (B) Binding kinetics, using biolayer interferometry, against different 92BR SOSIP.664 V3 loop region mutants. (C) DH270.1 heavy chain mutants and 92BR SOSIP.664. Biolayer interferometry association and dissociation curves for the indicated Fab mutants for binding to 92BR SOSIP.664 (600nM curves are shown) Not shown are curves for DH270.1 heavy chain mutants K32A, R72A, D73A, S25D, S54D, S60D and double mutant S75/77A for which there was little or no reduction in affinity.


A
DH270.5
VH= 11.5%
IA1
VH= 8.8%
DH270.4
VH= 11.9%
IA2
VH= 2.4%
DH270.1
VH= 5.8%
UCA
VH= 0%
IA4
VH= 1.4%
DH270.2
VH= 10.8%
IA3
VH= 5.1%
DH270.3
VH= 11.5%
DH270.UCA
IA4
Kd >10μM
IA3
Kd >10μM
IA2
Kd >10μM
DH270.1
Kd = 331nM
ka1 = 1.54 x10³ M⁻¹s⁻¹
kd1 = 5.1 x10⁻⁴ s⁻¹
IA1
Kd = 3292nM
ka1 = 1.64 x10³ M⁻¹s⁻¹
Kd1 = 5.4 x10⁻³ s⁻¹
DH270.2
Kd = 4770nM
ka1 = 1.71 x10³ M⁻¹s⁻¹
Kd1 = 8.1 x10⁻³ s⁻¹
DH270.3
Kd = 195nM
ka1 = 4.0 x10³ M⁻¹s⁻¹
kd1 = 7.78 x10⁻⁴ s⁻¹
DH270.4
Kd = 346nM
ka1 = 1.5 x10³ M⁻¹s⁻¹
kd1 = 5.2 x10⁻⁴ s⁻¹
DH270.5
Kd = 281nM
ka1 = 3.3 x10³ M⁻¹s⁻¹
kd1 = 9.1 x10⁻⁴ s⁻¹
DH270.6
Kd = 188nM
ka1 = 1.97 x10³ M⁻¹s⁻¹
kd1 = 3.7 x10⁻⁴ s⁻¹
Response (nm)
Time (s)

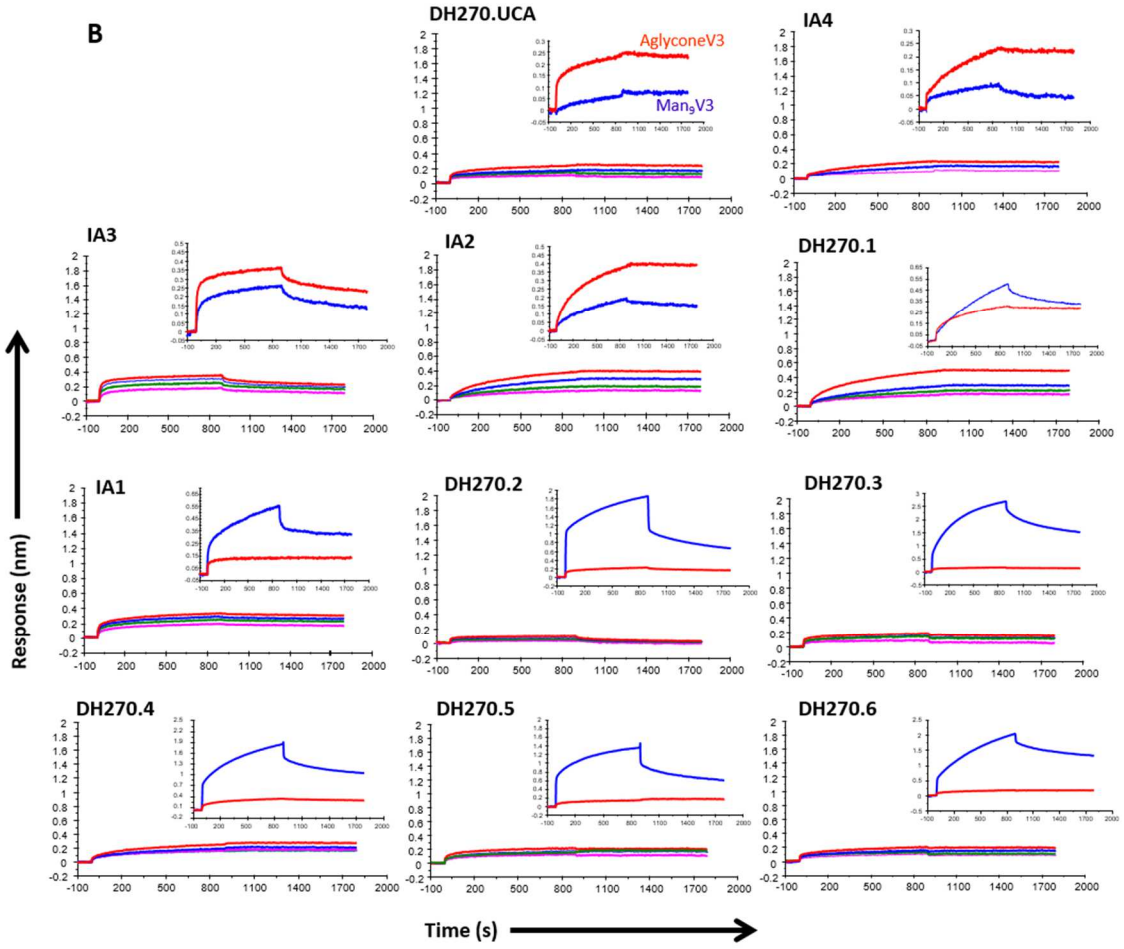
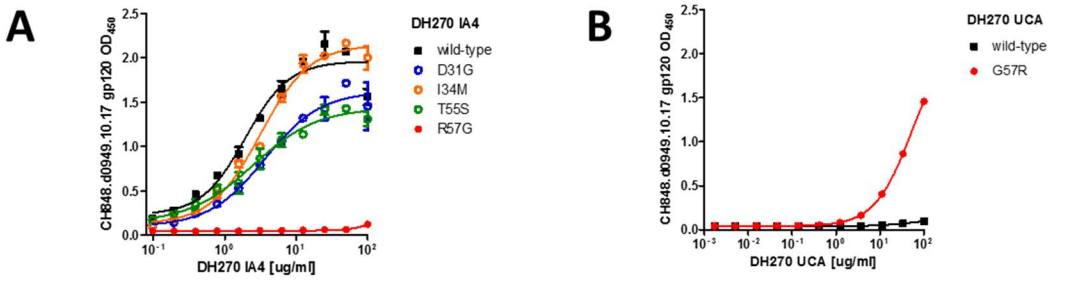


Figure 4.S19. Man$_9$-V3 glycopeptide binding of DH270 lineage antibodies. DH270 lineage tree (A, top left) is shown with V_H mutations of intermediates and mature antibodies. DH270.6 mAb, which clusters close to DH270.4 and DH270.5, is not shown in the phylogenetic tree. Binding of Man$_9$-V3 glycopeptide (*74*) and its aglycone form to DH270 lineage antibodies was measured by BLI assay using either biotinylated Man$_9$-V3 (A) or biotinylated aglycone V3 (B) as described in methods. DH270 lineage antibodies were each used at concentrations of 5, 10, 25, 50, 100, 150 µg/mL. Insets in (A) for UCA (150 µg/mL), IA4 (100, 50, 25 µg/mL), IA3 and IA2 (100, 50, 25, 10 µg/mL) show rescaled binding curves following subtraction of non-specific signal on a control antibody (Palivizumab). Rate (k_a, k_d) and dissociation constants (K_d) were measured for intermediate IA1 and mature mAbs with glycan-dependent binding to Man$_9$-V3. Kinetics analyses were performed by global curve fitting using bivalent avidity model and as described in methods ("Affinity measurements" section). Inset in (B) show overlay of binding of each mAbs to Man$_9$-V3 (blue) and aglycone V3 (red) at the highest concentration used in each of the dose titrations.



C

		gp120 Binding AUC									Neutralization IC50								
Boost	Subject.day.clone	UCA	IA4	IA3	IA2	I	IA1	4	5	6	UCA	IA4	IA3	IA2	I	IA1	4	5	V1 loop length
i.	CH848.0949.10.17	0	2.5	6.3	9.4	12.9	13.6	14.3	14.6	14.3	>50	0.64	0.2	0.14	0.18	<0.02	0.03	0.05	17
ii.	CH848.0836.10.31	0	0.0	4.9	9.4	13.2	13.5	13.2	14.5	14.2	>50	45.68	0.07	<0.02	<0.02	<0.02	<0.02	<0.02	17
iii.	CH848.0358.80.06	0	0.0	0.0	0.3	11.2	13.5	14.9	15.0	13.9	>50	>50	>50	>50	1.3	0.02	<0.02	<0.02	24
iii.	CH848.1432.5.41	0	0.0	0.0	0.8	11.7	12.7	12.7	14.6	14.0	>50	>50	>50	>50	2.6	0.33	0.26	0.21	30
iii.	CH848.0526.25.02	0	0.0	0.0	0.0	8.8	12.0	13.6	13.8	13.6	>50	>50	>50	>50	16.7	1.4	1.1	1.6	34

Figure 4.S20. Example of an immunization regimen derived from studies of virus-bnAb coevolution in CH848. (A) CH848 autologous gp120 Env clone d0949.10.17 binding depends on presence of Arg57 in DH270.IA4 and DH270.UCA. DH270.IA4 variants, reverted to UCA at positions of each of the 4 amino acids in V_H that had mutated in DH270.IA4 from their identities in DH270.UCA (colored circles), were tested by ELISA for binding to CH848.d0949.10.17 gp120, the only autologous Env bound by wild-type DH270.IA4 (black squares and Fig. 6A). Antibodies were serially diluted with a starting concentration of 100 μg/ml. Mean and standard deviation from duplicate observations are indicated for each datapoint and curve fitting (non-linear, 4-parameters) is shown for each dataset. Binding is quantified as background subtracted OD_{450} values. Reversion of R57G was necessary and sufficient to abrogate binding. **(B)** Introduction of G57R (red) in DH270.UCA (black) was sufficient to generate detectable binding with CH848.d0949.10.17 gp120 Env. (**C**) An immunization strategy composed of the following steps: first, prime with an immunogen that binds the UCA and the boost with immunogens with the following characteristics: i. engagement of DH270.IA4-like antibodies and selection for the G57R mutation; ii. Selection of antibodies that favor recognition of trimeric Env and expand the variation in the autologous signature residue to potentially expand recognition of diversity in population; iii. Exposing maturing antibodies to viruses with longer loops, even though these viruses are not bound or neutralized as well as viruses with shorter V1 loops, as this is the main constrain on antibody heterologous population neutralization breadth.

Table 4.S1. N332-dependent CH848 plasma neutralization.

	CH848 plasma neutralization, week 186 post-transmission (EC_{50} titer)		
HIV-1 strain	**Wild-type A**	**N332A mutant**	**Fold-difference**
Q23.17	962	150	6
Du156	170	<40	4
TRO.11	204	<40	5
Consensus C	4,261	1,569	3

Fold difference in CH848 plasma neutralization IC50 of selected wild-type and N332 mutant HIV-1 strains

Table 4.S2. NGS longitudinal sampling of V_HDJ_H rearrangements assigned to the DH270, DH272 and DH475 lineages from memory B cell mRNA.

A. DH270 Lineage

Week[a]	Sample A	Sample B	Overlap[d]	Total[e]
11	0	0	0	32,732
19	0	0	0	31,179
64	1	0	0	19,383
111	0	0	0	87,224
160	0	0	0	114,729
186	9,365	14,268	776	161,104
240[c]	3	251	1	171,012

B. DH272 Lineage

Week[a]	Sample A	Sample B	Overlap	Total
11	0	0	0	32,732
19	1,634	2,782	105	31,179
64	48	0	0	19,383
111	558	684	42	87,224
160	0	0	0	114,729
186	596	509	36	161,104
240[c]	0	7	0	171,012

C. DH475 Lineage

Week[b]	Sample A	Sample B	Overlap	Total
11	0	0	0	30,764
19	0	0	0	39,784
64	378	276	3	21,353
111	0	0	0	30,925
160	0	0	0	96,301
186	788	1590	50	37,648
240[c]	0	0	0	171,012

a Reads from V_H1 family-targeted Illumina
b Reads from V_H3 family-targeted Illumina
c Reads from V_H1 through V_H6 families-targeted Illumina
d Number of clonally-related V_HDJ_H sequences that replicated across samples A and B
e Total number of replicated sequences at each timepoint

Table 4.S3. CH848 plasma neutralization breadth over time.

	Week post-transmission, neutralization IC_{50} µg/ml							
HIV-1 strain	**51**	**100**	**111**	**113**	**136**	**160**	**186**	**205**
Du172.17	31	81	117	95	61	101	83	415
6535.3	<10	<10	<10	32	<10	115	460	1,363
Q23.17	<10	<10	<10	<10	64	<10	37	1,483
Du156.12	<10	<10	<10	<10	<10	25	280	682
SC422661.8	<10	<10	<10	<10	<10	38	92	207
Q842.d12	<10	<10	<10	<10	<10	50	74	54
QH0692.42	<10	<10	<10	<10	<10	<10	56	100
PVO.4	<10	<10	<10	<10	<10	<10	32	251
AC10.0.29	<10	<10	<10	<10	<10	<10	21	209
RHPA4259.7	<10	<10	<10	<10	<10	<10	59	250
Du422.1	<10	<10	<10	<10	<10	<10	152	341
ZM197M.PB7	<10	<10	<10	<10	<10	<10	43	105
Q259.d2.17	<10	<10	<10	<10	<10	<10	<10	112
ZM214M.PL15	<10	<10	<10	<10	<10	<10	<10	37
Q769.d22	<10	<10	<10	<10	<10	<10	<10	<10
CAP45.2.00.G3	<10	<10	<10	<10	<10	<10	<10	<10
SVA-MLV	<10	<10	<10	<10	<10	<10	<10	<10

Table 4.S4. Data collection and refinement statistics

Data Collection							
	UCA1	**UCA3**	**DH270.1**	**DH270.3**	**DH270.5**	**DH270.6**	**DH272**
Resolution (Å)	50 - 3.31 (3.37 - 3.31)	50 - 2.26 (2.30 - 2.26)	50 - 3.47 (3.53 - 3.47)	50 - 2.50 (2.54 - 2.50)	50 - 1.85 (1.88 - 1.85)	50 - 2.73 (2.78 - 2.73)	50 - 2.68 (2.73 - 2.68)
Space group	R 3 2	C 1 2 1	P 1 21 1	P 21 21 21	C 1 2 1	P 1 21 1	I 2 2 2
Unit cell a,b,c (Å)	209.1, 209.1, 83.5	98.5, 76.9, 137.1	83.8, 219.4, 103.1	63.2, 69.3, 115.2	135.1, 68.8, 59.5	67.1, 73.7, 112.7	60.9, 124.5, 146.2
Unit cell α, β, γ (°)	90 90 120	90,102, 90	90, 109, 90	90, 90, 90	90,113, 90	90, 107, 90	90, 90, 90
Total reflections	65247	142800	168772	57856	126544	57894	47496
Unique reflections	10410	46692	42952	17599	42441	27592	15439
Redundancy	6.3 (5.0)	3.1 (3.0)	3.9 (2.2)	3.3 (3.0)	3.0 (2.1)	2.1 (2.1)	1.7 (1.6)
Completeness (%)	97.6 (90.5)	99.1 (98.9)	94.1 (68.7)	97.1 (98.4)	96.2 (83.9)	97.2 (92.8)	97.5 (94.4)
$\langle I/\sigma_I \rangle$	7.0 (1.9)	7.3 (1.5)	6.7 (2.0)	6.8 (1.9)	5.6 (1.7)	6.00 (1.4)	9.3 (2.6)
R_{merge}	21.0 (88.3)	9.6 (87.8)	18.3 (32.0)	11.0 (66.2)	12.8 (56.0)	11.2 (64.2)	11.9 (63.8)
Refinement							
R_{work}/R_{free} (%)	26.4/27.6 (34.6/41.2)	23.6/26.5 (33.8/37.6)	22.5/25.3 (28.8/31.1)	21.2/24.8 (33.0/42.0)	17.6/21.4 (23.9/26.3)	24.3/26.5 (34.0/35.7)	22.0/26.7 (34.1/44.7)
No. atoms							
Protein	3215	6428	19474	3250	3303	6568	3376
Ligand		1					30
Water	0	195	0	114	353	0	41
R.M.S. deviations							
Bond lengths (Å)	0.004	0.005	0.006	0.005	0.007	0.004	0.005
Bond angles (°)	1.15	1.02	1.28	1.04	1.22	0.91	0.89
B-factors (Å^2)							
Protein	39.00	50.70	34.50	47.10	26.80	50.00	49.60
Ligand		30.00					69.20
Solvent		46.00		38.30	39.80		36.90

Statistics for the highest-resolution shell are shown in parentheses.

CHAPTER 5

HIV-1 THERAPY WITH MONOCLONAL ANTIBODY 3BNC117 ELICITS HOST IMMUNE RESPONSES AGAINST HIV-1

Reprinted from the publication:

Schoofs T, Klein F, Braunschweig M, **Kreider EF**, Feldmann A, Nogueira L, Oliveira T, Lorenzi JC, Parrish EH, Learn GH, West AP Jr, Bjorkman PJ, Schlesinger SJ, Seaman MS, Czartoski J, McElrath MJ, Pfeifer N, Hahn BH, Caskey M, Nussenzweig MC. HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1. Science. 2016 May 20;352(6288):997-1001. doi: 10.1126/science.aaf0972. Epub 2016 May 5. PubMed PMID: 27199429.

Contribution

For this study, I analyzed single genome sequencing data, correlated changes in neutralization to features of the evolving *env* quasispecies, and helped write/edit the manuscript. Along with Jerry Learn and Till Schoofs, I built Figures 3, 4, S6-S10, and Table S10

All figures are at the end of this chapter

Supplementary information available at

http://science.sciencemag.org/content/sci/suppl/2016/05/04/science.aaf0972.DC1/aaf0972-Schoofs-SM.pdf

Section 5.1 – Abstract

3BNC117 is a broad and potent anti-HIV-1 neutralizing antibody that targets the CD4 binding site on the viral envelope spike. When administered passively, this antibody can prevent infection in animal models and suppress viremia in HIV-1-infected individuals. Here we report that HIV-1 immunotherapy with a single injection of 3BNC117 impacts host antibody responses in viremic subjects and in individuals receiving anti-retroviral therapy. In comparison to untreated controls that showed little change in their neutralizing activity over a six-month period, 3BNC117 infusion significantly improved neutralizing responses to heterologous tier 2 viruses in nearly all study participants. We conclude that 3BNC117-mediated immunotherapy enhances host humoral immunity to HIV-1.

Section 5.2 – Main Text

Development of serum neutralization breadth during HIV-1 infection typically occurs after several years and exists on a continuum with ~50% of infected individuals developing some level of broad neutralization and a small fraction of individuals acquiring serum neutralizing activity of extraordinary breadth and potency (*1-4*). Antibody cloning experiments revealed that this activity is due to one or more potent broadly neutralizing antibodies (bNAbs) that target one or more epitopes on the viral spike protein, gp160 (*1, 5-10*).

bNAbs show exceptional breadth and potency *in vitro*, and can protect against or suppress active infection in humanized mice (*11-13*) and macaques (*14, 15*). Moreover, in a phase I clinical trial, a single injection of 3BNC117, a CD4-binding-site specific bNAb (*6*) was safe and effective in suppressing HIV-1 viremia by an average of 1.48 logs (*16*).

In addition to direct effects on target cells and pathogens, antibody-mediated immunotherapies have the potential to engage the host immune system and induce both innate and adaptive immune responses (*17*). In particular the Fc domains of antibodies interact with receptors on innate cells such as natural killer (NK) cells and phagocytes to enhance the clearance of viral particles and the killing of infected cells (Lu, 2016; Martin, 2016).

A single 3BNC117 infusion was administered to HIV-1-infected individuals at doses of 1, 3, 10, or 30 mg/kg (Fig. 1A, Table S1A) (*16*). To determine whether 3BNC117 therapy is associated with changes in viral sensitivity and serologic responses to autologous viruses, we cultured HIV-1 from peripheral blood mononuclear cells (PBMCs) of 9 viremic individuals before (d0) and 4 weeks (wks) after 3BNC117 infusion (*16*). On d0, all but one of the cultured viruses were sensitive to 3BNC117 with IC_{50} values ranging from 0.09 - 8.8 μg/ml (Fig. 1B and (*16*)). At wk 4, we found increased resistance to 3BNC117 in most individuals indicating selection for viral escape variants (Fig. 1B and (*16*)).

When the same viral isolates were tested for sensitivity to the matched individual's immunoglobulins (IgG) obtained before (d0) and 24 wks after 3BNC117 infusion (Fig. 1A), we found increased neutralizing activity in the wk 24 IgG against both d0 and wk 4 autologous viruses (p=0.0078, Fig. 1C, Table S2). Thus, while 3BNC117 infusion selected for 3BNC117-resistant HIV-1 variants, neutralizing antibody responses continued to develop against autologous viruses (*18*).

To test for changes in heterologous neutralizing activity following 3BNC117 treatment, we assayed patients' d0 and wk 24 IgG against a panel of tier 1 (n=1) and tier 2 (n=12) HIV-1 pseudoviruses that included globally circulating HIV-1 strains (*19*) (Fig. 2, Table

S1, S3, Table S4). Neutralizing activity was compared between the two time points by measuring the area under the neutralization curve for subjects' isolated IgG against each virus (AUC) (Table S4B). 15 subjects that received 3BNC117 were not on anti-retroviral therapy (ART) and had starting viral loads from 640 - 53,470 copies/ml (Table S1A). Paired control IgGs were obtained from 36 viremic individuals, who did not receive 3BNC117 and had starting viral loads ranging from 150 – 303,200 copies/ml (Fig. 2, Table S1B).

During a 6-month observation period, control individuals' neutralizing activity showed no consistent improvement in either breadth or potency (Fig. 2A and B, Fig. S1A, S2, Table S4, S5) (*4, 20*). In contrast, all but one of the 15 viremic individuals infused with 3BNC117 showed increased breadth and/or potency against the pseudovirus panel at wk 24 ($p=7.1 \times 10^{-7}$, Fig. 2A, S1B, S2, Table S4, S5, S6). The absolute change in neutralizing activity varied between viruses and individuals, ranging from small effects to dramatic increases (Fig. 2C, Table S1, S4, S5, S6). Significant differences were also evident between treated and control groups regardless whether sera from all individuals were considered in aggregate, or examined against individual viruses ($p=1.9 \times 10^{-9}$, Fig. 2B, D).

In addition to viremic subjects, we examined 12 individuals that received 3BNC117 while on ART, with no detectable or low-level viremia (<20 - 100 copies/ml). In

comparison to viremic subjects, the increase in heterologous neutralizing activity was less pronounced in ART-treated individuals (p=0.037, Fig. 2A, B, and D, S1B, S2, Table S4, S5).

The observed improvement in neutralizing activity could not be explained by confounding factors such as differences in initial viral load and $CD4^+$ T cell levels (Fig. S3, Table S1, S7). Moreover, we found no correlation between d0 neutralizing activity and neutralization improvement (Fig. S4). A comparison of the pattern of neutralization increase with 3BNC117's neutralization profile ruled out that remaining antibody is responsible for the effect (Fig. S5, Table S8). We conclude that 3BNC117 enhances host immunity to heterologous tier 2 HIV-1 viruses irrespective of initial neutralization breadth and potency.

To examine the effects of 3BNC117 immunotherapy on the plasma viral population of treated individuals, we performed single genome sequencing (SGS) of over 1,000 plasma-derived gp160 *env* genes (gp160) before (d0) and 4 (6), 12, or 24 weeks (wks) after infusion (Fig. 3A, B and S6, Table S9). With the exception of two individuals who were sexual partners, all other volunteers had epidemiologically unrelated infections (Fig. 3A). On d0, *env* sequences from subjects 2A1, 2A3, and 2C4 comprised multiple lineages, which was reflected in a multimodal distribution of pairwise diversity measurements from these individuals (Fig. 3B, S6). Analysis of *env* sequences from

subsequent time points revealed significant shifts in both nucleotide (6 out of 9 individuals, Fig. 3B) and amino acid sequence diversity (7 out of 9 individuals, Fig. S6). Consistent with the observation that *env* diversity is associated with neutralization breadth (*21-23*), there was a strong correlation between the initial level of neutralizing activity and the initial diversity of the circulating viral swarm ($R^2 = 0.92$, Fig. 3C).

We next evaluated viral sequence evolution in each of the 3BNC117-treated subjects over time. Shifts in the viral quasispecies were evident regardless of initial 3BNC117 neutralization sensitivity and bNAb dose (Fig. 4, S7). However, the nature of these shifts differed depending on the subject (Fig. 4, S7-S9). For example, in subject 2A1, 15/27 d0 sequences fell into a single clade marked "group A" (Fig. 4A, S8). Four weeks following 3BNC117 infusion, group A viruses contracted (2/25 sequences) and group C viruses expanded (16/25). At wk 24, the viral quasispecies was primarily comprised of group B and D viruses (Fig. 4A, S8). This pattern of "clade shifting" was also seen in subjects 2A3 and 2C4 (Fig. S7). Subjects with lower initial *env* diversities, such as 2E1, did not harbor distinct viral sublineages at d0 (Fig. 3A, B), but continued to accrue mutations some of which became fixed during the 24-week follow up (*e.g.* changes in V1/V2 in 2E1, Fig. S9).

To assess viral sequence changes following 3BNC117 infusion, we generated longitudinal logo plots depicting 3BNC117 contact residues (*24, 25*) for each subject

(Fig. 4B, S7, S10). While viruses from all nine subjects exhibited mutations within 3BNC117 contact residues relative to the d0 consensus sequence, their number and position varied considerably as exemplified by subjects 2A1 and 2E1 (Fig. 4B, Fig. S7, S10). Using LASSIE (Longitudinal Antigenic Sequences and Sites from Intrahost Evolution)(*26*), we scanned the entire *env* protein sequence for sites selected within the 24 wk time frame (selection cutoff ≥80% relative to d0 consensus) (Table S10). While selected sites were identified in all subjects, no consistent mutational pattern was observed (Table S10). These data suggest that 3BNC117 immunotherapy is associated with shifts in circulating quasispecies and a number of different *env* mutations, some of which persist even after the infused antibody levels drop below detection.

To better understand the virus host-interactions that led to the development of enhanced heterologous neutralizing breadth, we performed neutralization assays on 63 pseudoviruses expressing the gp160s found in the circulation on d0, wk 4, 12 and 24 from 5 individuals (Fig. 4, S7, Table S11). The pseudoviruses were tested for sensitivity to the corresponding individual's IgG obtained on d0 and wk 24. In all cases, we were able to identify d0 or wk 4 viruses that exhibited greater neutralization sensitivity to wk 24 IgG compared to d0 IgG (Fig. 4, Fig. S7, Table S11). For example, all 2A1 and 2E1 viruses were 3BNC117 sensitive and exhibited a wk 24/d0 fold change of 1.7 and 4.8 in IgG IC_{50} respectively (Fig. 4). On the other hand, all 2C4 viruses were 3BNC117-resistant (mean IC_{50}: >20 μg/ml), yet they were 5.8-fold more sensitive to wk 24 IgG

versus d0 IgG (Fig. S7). In conclusion, viremic individuals receiving 3BNC117 produced antibodies to autologous viruses that were both sensitive and resistant to 3BNC117.

While exceptional broadly neutralizing antibodies to HIV-1 develop only sporadically in a fraction of infected individuals, most HIV-1 infected individuals develop some level of neutralization breadth (*1-4*). Here we show that 3BNC117 immunotherapy accelerates this process. This boost in heterologous breadth occurs irrespective of demographic, virologic, or dosage factors and was associated with both transient and lasting changes to the viral quasi-species. Of note, neutralization improvements observed were modest in most individuals, potentially owing to the transient nature of therapy with a single antibody as well as the short timeframe of observation.

Although the effect of 3BNC117 on neutralizing responses to heterologous HIV-1 viruses may seem surprising, anti-HIV-1 antibodies have been associated with enhanced immunity in infants born to HIV-1-infected mothers that have circulating anti-HIV-1 antibodies and macaques treated with monoclonal antibodies or neutralizing serum (*27-29*).

How passively administered antibodies to HIV-1 accelerate the emergence of bNAbs is not completely understood. One possibility is that 3BNC117 infusion selected for viral

variants with altered antigenic properties, which in turn stimulated new B cell lineages. (*21-23, 30-32*). A second possibility is that immune complexes formed by 3BNC117 and circulating viruses act as potent immunogens, a phenomenon that is believed to be responsible for the enhanced $CD8^+$ T cell immunity to tumor antigens in individuals receiving monoclonal antibody based immunotherapy (*33, 34*).

Irrespective of the mechanism(s), the enhanced antibody response found in individuals receiving 3BNC117 therapy indicates that immunotherapy boosts host immunity to HIV-1. Moreover, the finding that antibody responses to heterologous tier 2 viruses develop in nearly all 3BNC117-treated individuals suggests that host genetics or a specific viral envelope sequence do not limit the development of neutralizing antibodies to HIV-1.

Section 5.3 – References:

1. F. Klein *et al.*, Antibodies in HIV-1 vaccine development and therapy. *Science* **341**, 1199-1204 (2013).

2. A. P. West, Jr. *et al.*, Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* **156**, 633-648 (2014).

3. I. Mikell *et al.*, Characteristics of the earliest cross-neutralizing antibody response to HIV-1. *PLoS pathogens* **7**, e1001251 (2011).

4. P. Hraber *et al.*, Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* **28**, 163-169 (2014).

5. X. Wu *et al.*, Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**, 856-861 (2010).

6. J. F. Scheid *et al.*, Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. *Science*, (2011).

7. L. M. Walker *et al.*, Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466-470 (2011).

8. F. Klein *et al.*, Broad neutralization by a combination of antibodies recognizing the CD4 binding site and a new conformational epitope on the HIV-1 envelope protein. *J Exp Med* **209**, 1469-1479 (2012).

9. M. Bonsignori *et al.*, Two Distinct Broadly Neutralizing Antibody Specificities of Different Clonal Lineages in a Single HIV-1-infected Donor: Implications for Vaccine Design. *Journal of virology*, (2012).

10. J. F. Scheid *et al.*, Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636-640 (2009).

11. F. Klein *et al.*, HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* **492**, 118-122 (2012).

12. J. A. Horwitz *et al.*, HIV-1 suppression and durable control by combining single broadly neutralizing antibodies and antiretroviral drugs in humanized mice. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16538-16543 (2013).

13. J. Pietzsch *et al.*, A mouse model for HIV-1 entry. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15859-15864 (2012).

14. D. H. Barouch *et al.*, Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. *Nature* **503**, 224-228 (2013).

15. M. Shingai *et al.*, Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. *Nature* **503**, 277-280 (2013).

16. M. Caskey *et al.*, Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* **522**, 487-491 (2015).

17. S. Bournazos, J. V. Ravetch, Fcgamma receptor pathways during active and passive immunization. *Immunol Rev* **268**, 88-103 (2015).

18. X. Wei *et al.*, Antibody neutralization and escape by HIV-1. *Nature* **422**, 307-312 (2003).

19. A. deCamp *et al.*, Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. *Journal of virology* **88**, 2489-2507 (2014).

20. S. G. Deeks *et al.*, Neutralizing antibody responses against autologous and heterologous viruses in acute versus chronic human immunodeficiency virus (HIV) infection: evidence for a constraint on the ability of HIV to completely evade neutralizing antibody responses. *Journal of virology* **80**, 6155-6164 (2006).

21. H. X. Liao *et al.*, Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* **496**, 469-476 (2013).

22. N. A. Doria-Rose *et al.*, Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* **509**, 55-62 (2014).

23. P. L. Moore, C. Williamson, L. Morris, Virological features associated with the development of broadly neutralizing antibodies to HIV-1. *Trends Microbiol* **23**, 204-211 (2015).

24. T. Zhou *et al.*, Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science* **329**, 811-817 (2010).

25. F. Klein *et al.*, Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* **153**, 126-138 (2013).

26. P. Hraber *et al.*, Longitudinal Antigenic Sequences and Sites from Intra-Host Evolution (LASSIE) Identifies Immune-Selected HIV Variants. *Viruses* **7**, 5443-5475 (2015).

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

28. N. L. Haigwood *et al.*, Passive immunotherapy in simian immunodeficiency virus-infected macaques accelerates the development of neutralizing antibodies. *Journal of virology* **78**, 5983-5995 (2004).

29. C. T. Ng *et al.*, Passive neutralizing antibody controls SHIV viremia and enhances B cell responses in infant macaques. *Nat Med* **16**, 1117-1119 (2010).

30. P. L. Moore *et al.*, Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. *Nat Med* **18**, 1688-1692 (2012).

31. F. Gao *et al.*, Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell* **158**, 481-491 (2014).

32. J. N. Bhiman *et al.*, Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nat Med* **21**, 1332-1336 (2015).

33. T. T. Wang *et al.*, Anti-HA Glycoforms Drive B Cell Affinity Selection and Determine Influenza Vaccine Efficacy. *Cell* **162**, 160-169 (2015).

34. S. Bournazos, D. J. DiLillo, J. V. Ravetch, The role of Fc-FcgammaR interactions in IgG-mediated microbial neutralization. *J Exp Med* **212**, 1361-1369 (2015).

35. S. J. Ratcliffe, J. Shults, GEEQBOX: A MATLAB toolbox for generalized estimating equations and quasi-least squares. *J Stat Softw* **25**, 1-14 (2008).

36. E. E. Giorgi, T. Bhattacharya, A note on two-sample tests for comparing intra-individual genetic sequence diversity between populations. *Biometrics* **68**, 1323-1326; author reply 1326 (2012).

37. A. B. van 't Wout, H. Schuitemaker, N. A. Kootstra, Isolation and propagation of HIV-1 on peripheral blood mononuclear cells. *Nature protocols* **3**, 363-370 (2008).

38. M. S. Seaman *et al.*, Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *Journal of virology* **84**, 1439-1452 (2009).

39. M. Li *et al.*, Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *Journal of virology* **79**, 10108-10125 (2005).

40. A. P. West, Jr. *et al.*, Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 10598-10603 (2013).

41. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519-1522 (2014).

42. M. A. Larkin *et al.*, Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).

43. M. Kearse *et al.*, Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649 (2012).

44. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772 (2012).

45. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307-321 (2010).

46. W. Deng *et al.*, DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques* **48**, 405-408 (2010).

47. D. C. Nickle *et al.*, HIV-specific probabilistic models of protein evolution. *PloS one* **2**, e503 (2007).

48. P. B. Gilbert, A. J. Rossini, R. Shankarappa, Two-sample tests for comparing intra-individual genetic sequence diversity between populations. *Biometrics* **61**, 106-117 (2005).

49. J. L. Kirchherr *et al.*, High throughput functional analysis of HIV-1 env genes without cloning. *J Virol Methods* **143**, 104-111 (2007).

Section 5.4 – Acknowledgments:

We thank all study participants who devoted time to our research. We thank the Rockefeller University Hospital Clinical Research Support Office and nursing staff for patient care and recruitment and members of the Nussenzweig lab for helpful discussions. We thank M. Schechter and C. Baro for technical assistance, P. Fast and H. Park for Clinical monitoring, E. Gotschlich and B. Coller for input on study design and P. Hraber for helping with LASSIE analyses. The data reported in this study are tabulated in the main paper and in the supplementary material. Envelope sequencing data can be downloaded from GenBank (Accession no. KX027737-KX028736). This work was supported in part by the Bill and Melinda Gates Foundation Collaboration for AIDS Vaccine Discovery (CAVD) Grants OPP1032144 (M.S.S.), OPP1092074 and OPP1124068 (M.C.N), the Robertson Foundation to M.C.N., the NIH Centers for HIV/AIDS Vaccine Immunology and Immunogen Discovery (CHAVI-ID) 1UM1 AI100663-01 (M.C.N) and 1UM1 AI00645 (B.H.H.), the University of Pennsylvania Center for AIDS Research (CFAR) Single Genome Amplification Service Center P30 AI045008 (B.H.H.), NIH grants UM1 AI068618 (MJM), UM1 AI069481 (MJM), F30 AI112426 (E.F.K), HIVRAD P01 AI100148 (P.J.B.). T.S. is supported by a German Research Foundation postdoctoral fellowship (SCHO 1612/1-1). F.K. is supported by the Heisenberg-Program of the DFG (KL 2389/2-1), the European Research Council (ERC-StG639961), and the German Center for Infection Research (DZIF), partner site Bonn-Cologne, Cologne, Germany. M.B. was supported by the German National Academic

Foundation. J.C.L. is supported by an award from CNPq "Ciencia sem Fronteiras" Brazil (248676/2013-0). M.C.N. is a Howard Hughes Medical Investigator.

Section 5.5 - Figures

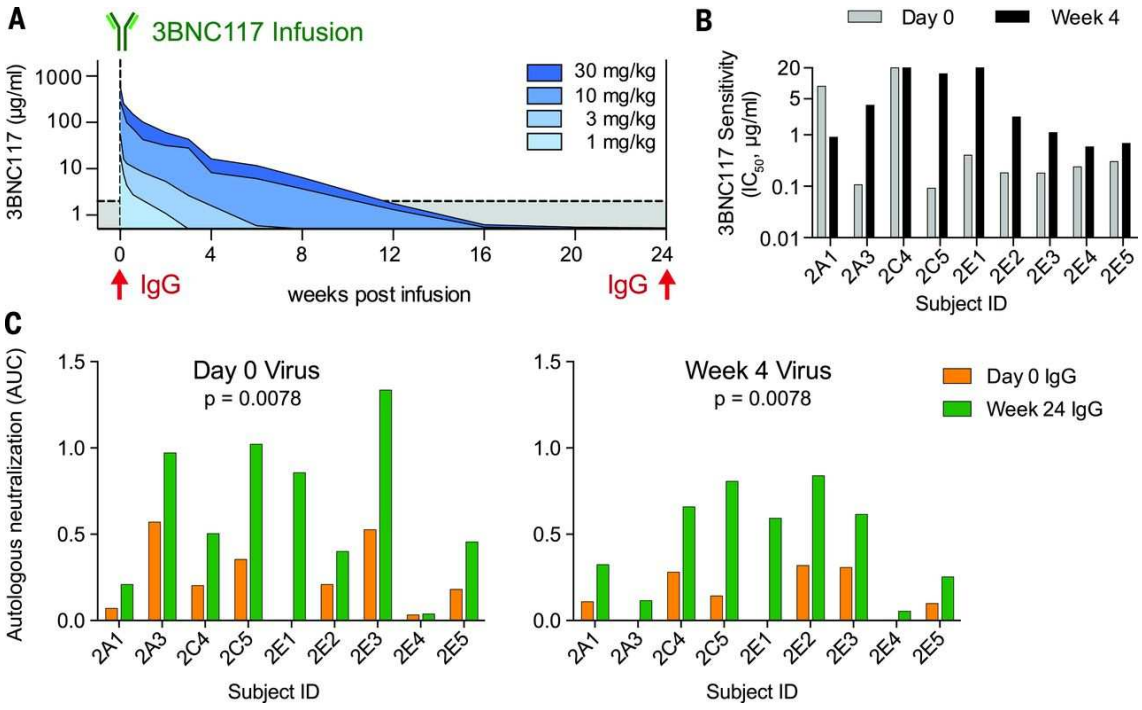


Figure 5-1 Virus sensitivity to 3BNC117 and autologous antibody responses

A. Graph displays kinetics of 3BNC117 antibody decay in HIV-1-infected individuals as determined by a validated anti-idiotype ELISA (*16*). Each sample measured in duplicates. Graphs show the mean antibody concentration of all patients infused for each of the indicated dose groups (n=3 (1 mg/kg), 3 (3 mg/kg), 8 (10 mg/kg), 13 (30 mg/kg)). Red arrows indicate the timepoints of IgG purification. **B.** Autologous virus sensitivity to 3BNC117 before (day 0, grey) and 4 wks (black) after 3BNC117 infusion. Y-axis shows IC_{50}s for 3BNC117 on viral culture supernatants from PBMCs determined by TZM.bl assay. Neutralization assays performed in duplicates. **C.** Graph shows the AUC of the neutralization curves of purified IgGs obtained from sera on day 0 (orange) or wk 24 (green) against day 0 (left panel) or wk 4 (right panel) autologous viruses. Neutralization assays performed in duplicates. p-values determined by Wilcoxon signed-rank test.

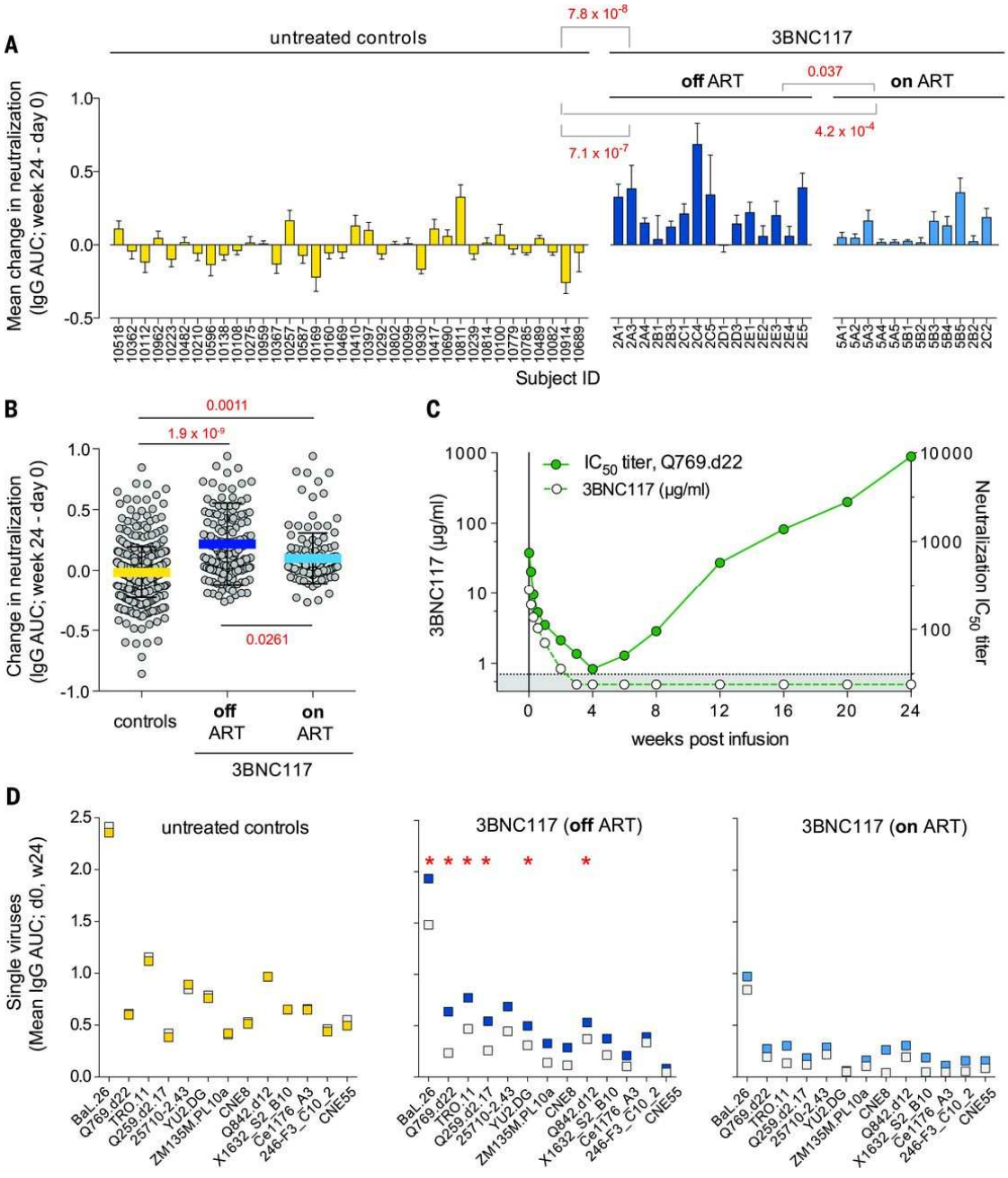


Figure 5-2 Heterologous antibody responses

A. Graph shows the difference in overall AUC (mean AUC change) per individual in TZM.bl assays against 13 heterologous viruses (see **2D**) for day 0 vs. wk 24 IgG obtained from 36 untreated viremic controls, 15 viremic individuals infused with 3BNC117, and 12 ART-treated individuals receiving 3BNC117 infusion (*16*). Neutralization assays performed in duplicates. p-values determined by unpaired Wilcoxon rank-sum test. **B.** Graph shows the aggregated differences in AUC between d0 and wk 24 IgG assayed by

TZM.bl for all viruses and all individuals. Each dot represents a single AUC difference for a single virus from one individual displayed in A. p-values determined using generalized estimating equations (*35*). **C**. Graph shows 3BNC117 antibody levels and TZM.bl neutralization titer against tier 2 strain Q769.d22 in subject 2A3 **D.** Mean AUCs of IgGs of all individuals at d0 (grey) and wk 24 (color of respective group) for each HIV-1 pseudovirus tested. Changes in neutralization of viremic control individuals without 3BNC117 infusion are shown in yellow (left). Change in neutralization of 3BNC117-treated individuals shown in dark (off ART, middle) and light blue (on ART, right)**.** p-values determined using unpaired Wilcoxon test (rank-sum test). Red stars indicate significant p-values after Bonferroni-correction (threshold $p < 0.0038$).

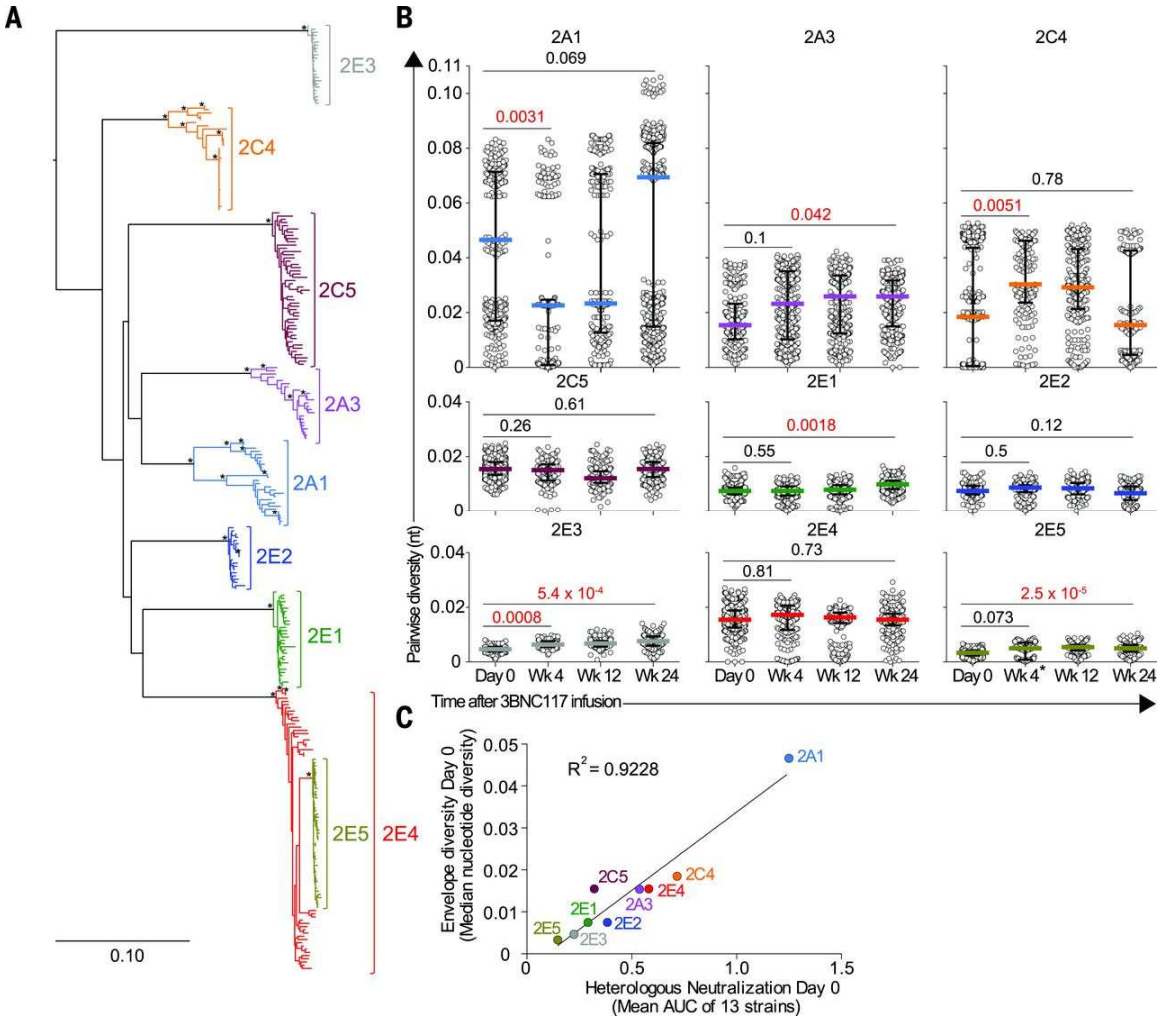


Figure 5-3 HIV-1 quasispecies diversity before and after 3BNC117 infusion

A. Maximum likelihood phylogenetic tree of single genome-derived *env* sequences from d0 plasma, before therapy with 3BNC117 (Table S9). Asterisks indicate bootstrap values of 100%. Individual viral sequences are color coded as indicated. **B.** Scatter plots depicting pairwise nucleotide sequence diversity of plasma *env* sequences on d0, and wk 4 (2E5, wk 6), 12 and 24 after infusion. Each dot represents the pairwise genetic difference between two sequences at a given timepoint. Colored bars indicate median diversity, while black bars indicates the interquartile range. P-values were determined using a two-sample U-statistic based Z-test(*36*). **C.** Graph showing the relationship between d0 mean heterologous neutralizing AUC against a panel of tier 1 (n=1) and tier 2 (n=12) viruses (abscissa) and the median pairwise nucleotide diversity (ordinate).

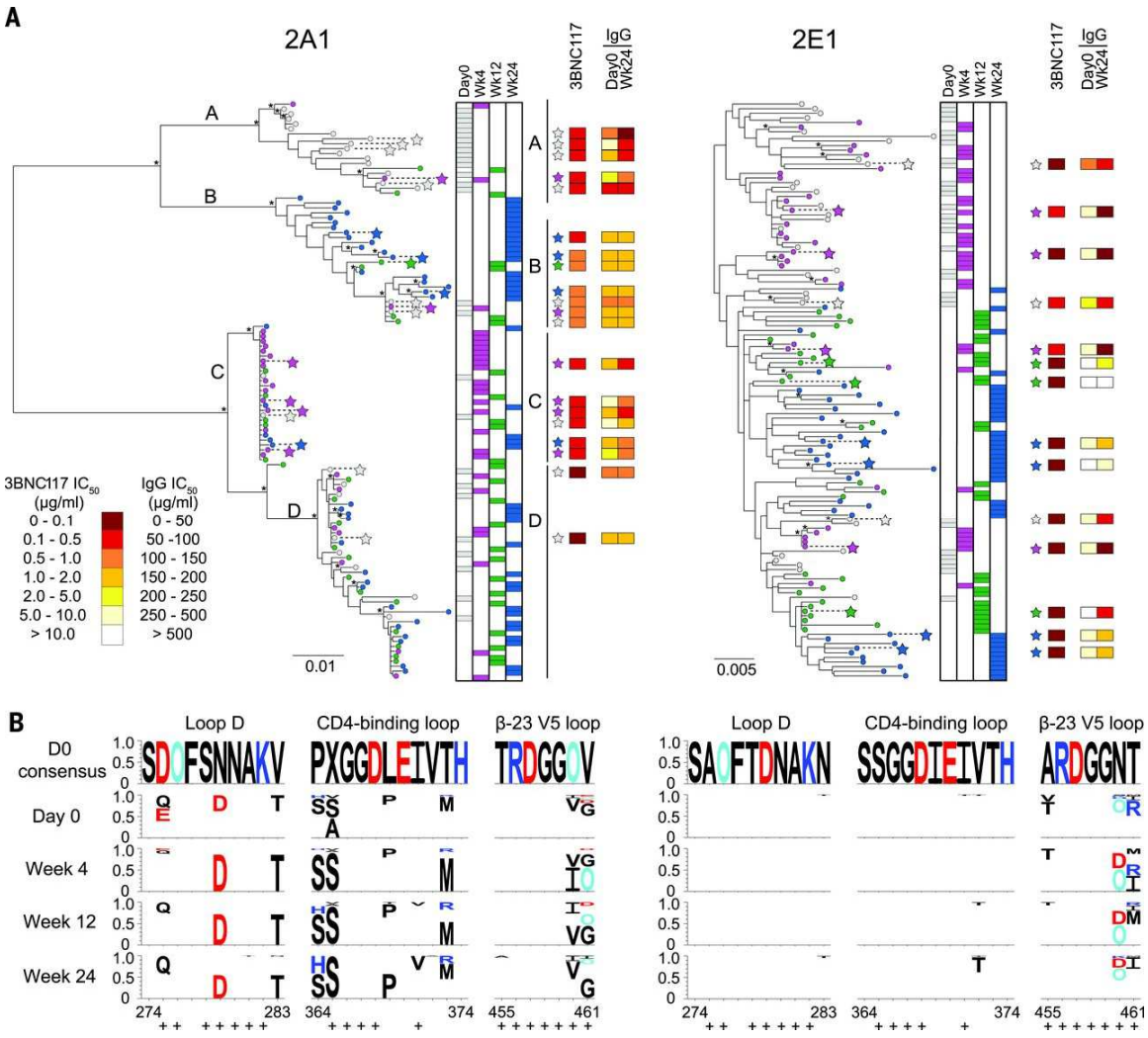


Figure 5-4 Antibody responses to the evolving viral quasispecies

A. Maximum-likelihood phylogenetic trees of single genome-derived *env* gene sequences from subjects 2A1 and 2E1 sampled on d0 and wk 4, 12, and 24 after 3BNC117 infusion (left). Clades with bootstrap support ≥ 70% are indicated by a black asterisk and are arbitrarily named groups A-D in case of subject 2A1. Bar graphs (middle) indicate the timepoints from which sequences in the tree are derived. Heat maps (right) show the 3BNC117 IC_{50}, d0 IgG IC_{50} and wk 24 IgG IC_{50} values against autologous pseudoviruses using *env*s as indicated by colored stars. Neutralization assays performed in duplicates. **B.** Sequence logo plots illustrating longitudinal amino acid changes in and around known 3BNC117-contact residues (*24, 25*) in subject 2A1 and 2E1. White boxes indicate that sequence matches to the d0 consensus, grey boxes indicate gaps in alignment. Colors

indicate basic (dark blue) and acidic (red) residues and a turquoise “O” is used instead of “N” to indicate a potential N-glycosylation site. Logo plots were generated using LASSIE (*26*). + indicate 3BNC117 contact residues confirmed by two crystal structures (*24, 25*)

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Edward F. Kreider

Departments of Medicine and Microbiology, Perelman School of Medicine at the University of Pennsylvania

Section 6.1 – HIV-1 5' leader sequence evolution during infection

Cytotoxic T lymphocyte responses against 5' leader peptides

We recently demonstrated that the SIVmac766 5' leader sequence, despite its alternative designation as an "untranslated" region, expresses a T cell antigen (KA9). This finding stemmed from the observation of viral escape mutations within the 5' leaders of SIVmac strains sequenced longitudinally during infection. To test if similar patterns of sequence evolution and peptide expression occurred during HIV-1 infection, I studied a cohort of acutely infected human subjects who were followed for 1-4 years of infection. As described in Chapter 2, I used longitudinal single genome sequencing, *ex vivo* immunologic assays, and a mass spectrometry-based approach to identify six different 5' leader peptides that were expressed during HIV-1 infection either *in vivo* or *in vitro*. These peptides were encoded by reading frames +1, +2, and +3 and overlapped with either the R element or the Ψ packaging signal. As discussed in Chapter 2, these findings suggest that HIV-1 translation initiation, which has been a controversial topic within the field, is mediated through cap-dependent ribosomal scanning (1-5). Moreover, 5' leader encoded peptides can be classified as viral Defective Ribosomal Products, or DRiPs. DRiPs are a class of T cell epitopes that are rapidly presented in major histocompatibility (MHC) complexes following translation and may play an important role in early immune surveillance of invading pathogens (6-9).

Due to some constraints of our initial analysis, future studies will be needed to characterize the extent of 5' leader DRiP expression during HIV-1 infection. A major limitation of our study was the method we employed to discover 5' leader DRiPs. By testing for virus escape mutations in longitudinal sequence datasets, we only detected DRiPs that were bound to the human leukocyte antigen (HLA) class I alleles expressed by the infected individual. To address this problem, we could implement two complementary approaches; we could 1) screen for 5' leader sequence evolution in a large, diverse acute infection cohort and 2) apply the mass spectrometry approach introduced in Chapter 2 in cell lines and donor cells that express diverse MHCs. The U.S. Army has recently conducted a trial called RV 217 in Africa and Thailand in which study volunteers who were at high risk for HIV-1 infection were screened twice weekly for seroconversion. If a study volunteer became infected, s/he was sampled twice weekly for four weeks and every three months thereafter (10). With over 112 subjects longitudinally sampled from acute infection, this cohort would be ideal for screening a large number of HIV-1 infected humans for 5' leader responses. Our current 5' leader PCR protocol requires two amplification reactions to obtain the entire 5' leader sequence. To make this process more high throughput, we could implement a 5' Rapid Amplification of cDNA Ends (RACE) protocol to amplify the entire 5' leader sequence in a single PCR reaction (11). 5' RACE and amplicon sequencing of samples taken at an acute time point and ~six months post-infection in RV217 subjects could then be used to test for sequence changes suggestive of CTL escape. This approach would not only assist in the identification of new subjects who mount 5' leader T cell responses, it would also give an accurate

estimate of the frequency of these responses during typical HIV-1 infection. While this approach will increase the number of viruses and subjects tested, it may still be limited by the human leukocyte antigen (HLA) genotype expressed by the study subjects. Thus, in parallel, we could extend the application of mass spectrometry-based approaches for peptide identification to include many donors or cell lines that express diverse HLA genotypes (12). Using multiple donors would potentially allow for a more complete characterization of protein expression and subsequent MHC-complex presentation of 5' leader peptides expressed by a particular virus (12). The combination of these two methods could give insight into the extent of 5' leader DRiP expression and the diversity of 5' leader ORFs among globally circulating HIV-1 strains.

Because 5' leader peptides represent a novel set of DRiPs, we hypothesized that they will be rapidly cell surface expressed in MHC complexes (8,9,13). One approach to test the kinetics of 5' leader peptide expression could be through implementation of the rhesus cytomegalovirus (RhCMV) 68.1 vaccine platform. This vaccine vector is capable of eliciting T cell responses against ~7 MHC-E and MHC-class II epitopes per 100 amino acids of vaccine insert in an MHC class-I independent manner (14-16). T cells are very sensitive to peptide-MHC complexes (requiring fewer than 10 on the cell surface for activation), so RhCMV-vaccine-primed PBMCs could serve as a sensitive reagent to screen for peptide-MHC surface expression of their target antigen (17). In preliminary experiments, we generated three RhCMV 68.1 vaccine vectors – one encoding each forward reading frame of the SIVmac766 5' leader – and vaccinated macaques with one or all three vectors. PBMCs from each vaccinated macaque specifically reacted with

SIV-infected CD4+ T cells, demonstrating that 5' leader peptides are expressed from all three forward reading frames and that these primed PBMCs could be used to monitor for cell-surface expression of 5' leader DRiP-MHC complexes (Figure 6.1). Moving forward, we can use these vaccine-primed PBMCs to generate 5' leader peptide-specific T cell lines that could be used in cytotoxicity and viral suppression assays (18-20). For comparison to canonical epitopes, we could run these same assays using PBMCs from macaques vaccinated using the RhCMV/gag, RhCMV/rtn (rev/tat/nef), or RhCMV/env (21). Finally, we will be using these 5'leader-RhCMV vectors in a vaccine efficacy trial to test whether macaques vaccinated against 5' leader DRiPs can clear newly acquired SIV infection (16,22).

If 5' leader DRiP vaccination proves effective in the rhesus-SIV model system, the next step will be identifying conserved vaccine targets in globally circulating HIV-1 strains. While the HIV-1 5' leader RNA sequence and structure are heavily conserved, the open reading frames (ORFs) are not (Fischer, W., Kreider, E., Shaw, G., Korber, B., unpublished). Insertions and deletions are common within the 5' leader and sometimes clade-specific. For instance, ~61% of Subtype A viruses encode a single nucleotide deletion in the Tat binding bulge of TAR (23). Further, in Chapter 2 I reported that 5' leader CTL escape mutations included start site mutations, nonsense mutations, and insertions/deletions. Despite this heterogeneity, we have identified 12 conserved regions in the HIV-1 5' leader that could serve as vaccine targets (Fischer, W., Korber, B., unpublished). These 12 regions, which are comprised of ~140 amino acids, could be used

to generate a 5' leader vaccine insert that, if immunogenic and expressed by HIV-1 strains, may elicit cross-subtype immune responses.

While the data I have presented thus far demonstrate that the HIV-1 5' leader expresses peptides, they do not address an important question: do 5' leader peptides perform a function? T cell recognition of an influenza alternative reading frame was previously used to discover a novel influenza protein, PB1-F2. *In vitro* and *in vivo* studies of PB1-F2 revealed that this protein compromises mitochondrial function, signals apoptosis, and can modify pathogenesis of both the primary viral infection and secondary bacterial pneumonia (24-26). An interesting future study could be a comparison of viral dynamics, T cell responses, and pathology in Mamu-B*29+ and Mamu-B*29- rhesus macaques infected with either the wildtype SIVmac766 clone or the 5'L-KA9 start site mutant (KA9 is a Mamu-B*29-restricted epitope). While I am unable to predict the outcome of such an experiment, even a negative result would be informative since it would indicate that KA9 expression is dispensable for viral replication and pathogenesis.

Overextension mutation and expression of 5' leader DRiPs

Longitudinal analysis of HIV-1 5' leaders has revealed multiple patterns of sequence evolution that did not fall in cognate CTL epitopes. One region that consistently exhibited mutation among different subjects was a 4-nt motif called the primer overextension sequence (POS) (27). In Chapter 3, I used longitudinal single genome

sequencing, an *in vitro* extension assay, and *ex vivo* immunologic assays to provide evidence for two mechanisms that underlie this sequence variation. The first was reverse transcriptase (RT) overextension, a phenomenon in which RT replicates 22 nucleotides (nts) of tRNA sequence instead of 18, incorporating an additional TTGA downstream of the primer binding site (PBS) (27,28). Embedded in these four nts is a one-off AUG (**UUG**A) that lies upstream of a 5' leader open reading frame (ORF) that I described in Chapter 2. A second pattern of POS mutation was the mutation of this one-off AUG during immune escape from a 5' leader-directed T cell response. Collectively, these two findings demonstrated that processes like RT overextension can restore one-off AUGs in the 5' leader that are deleted during virus immune escape.

While the data presented in Chapter 3 support the conclusion that RT overextension incorporates TTGA into the POS, the mechanism by which this mutation occurs remains unclear. A recent study of the simian immunodeficiency virus clone SIVmac239 has sparked a debate surrounding this topic (29-31). SIVmac239 differs from the host tRNA(Lys,3) at one position within the PBS and this site is rapidly mutated to match the tRNA upon viral passage *in vitro* and *in vivo* (30). Fennessey and colleagues argued that this base pair mismatch within the PBS conferred a fitness cost to the virus and invoked DNA mismatch repair mechanisms to explain the rapid rate with which this mutation is reverted (30,32). Berkhout and colleagues, however, contested this interpretation, arguing that there is no fitness impact of a single base pair mismatch within the PBS and that the rapid reversion is likely due to the preferential incorporation of the tRNA(Lys,3) into assembling virions (31). Further, they argued that cell division

likely occurs too soon after proviral integration for mismatch repair to take place (31,33). However, mismatch repair of a similar mutation in MLV infection has been detected (32). To resolve this conflict, one would need to test frequency of PBS reversion during infection of CD4+ T cells that lacked mismatch repair mechanisms. While one could employ transformed cell lines or CRISPR technology to generate these mismatch repair-deficient CD4+ T cells, there is a hereditary illness called Lynch Syndrome that is associated with mutations in mismatch repair genes (34). Thus, one could compare the relative rates of PBS (or POS) reversion in CD4+ T cells from either healthy donors or donors with Lynch syndrome to determine the contribution of mismatch repair to this phenomenon.

No matter how these mutations are incorporated, the process of one-off AUG incorporation during overextension has implications for 5' leader peptide vaccine development. As presented in Chapter 2, viral escape from 5' leader CTL responses differs from that seen in the canonical proteome; we have observed mutations in one-off AUGs, nonsense mutations, and even frameshift mutations. This observation raises an obvious question – why would the 5' leader encode any one-off AUGs? One possible explanation is that the 5' leader may require some one-off AUGs based on sequence constraints of the underlying functional RNA elements (35,36). Another possibility is that certain processes, like overextension, may restore these one-off AUGs as soon as they are deleted. This second possibility is particularly relevant to the development 5' leader DRiP-based vaccine interventions – if the virus cannot eliminate a one-off AUG, then an intervention targeting the downstream peptide may be effective. Future studies will need

to investigate whether other processes maintain or incorporate one-off AUGs in the 5' leader.

In summary, the results presented in Section 6.1 provide insight into HIV-1 5' leader peptide expression and immune recognition. Using HIV-1 sequence evolution as an indicator of underlying biology, I was able to identify peptide expression and immune recognition of five 5' leader peptides and understand why certain translation initiation codons for these peptides are maintained in the 5' leader. These findings have implications for future interventions that could target this region of the genome and deepen our understanding of retrovirus biology.

Section 6.2 – Env evolution during bnAb development

Env evolution in natural infection

The recent identification of over a hundred broadly neutralizing antibodies (bnAbs) that can potently inhibit diverse HIV-1 strains has led to general interest in their use in the prevention of HIV-1 infection (37-39). Eliciting these antibodies with a vaccine, however, has been challenging (40). One approach to understanding bnAb development has been the study of virus-antibody co-evolution in naturally infected human subjects who develop these responses (37,41-46). In Chapter 4, I presented one such study of a human subject called CH0848. By combining longitudinal plasma viral single genome sequencing, monoclonal antibody isolation, binding ELISAs, and neutralization assays, we were able to map the virus and antibody evolutionary pathways during CH0848 infection, which ultimately resulted in the development of the potent V3-glycan bnAb, DH270.

A major caveat to our interpretation of these data is that this report is a case study of only one human subject. Moreover, we only characterized three CH0848 neutralizing antibody lineages (DH270/DH272/DH475) out of a diverse, polyclonal response. This shortcoming is exemplified by the discrepancy between neutralization breadth of monoclonal DH270.6 and polyclonal CH0848 plasma – on the same 16 virus neutralization panel, CH0848 polyclonal plasma neutralized ~88% of viruses whereas the

monoclonal bnAb DH270.6 only neutralized 69% (Haynes, B., unpublished). Thus, understanding the complete development of heterologous neutralization breadth in CH0848 will require future studies.

Despite this caveat, we were able to identify at least three major evolutionary events along the path to DH270 development that may inform vaccine immunogen design: a ten-amino acid (aa) V1 deletion in Env one year post-infection, a mutation within an AID coldspot during DH270 affinity maturation, and diversification of V1 length and V3-glycan bnAb contact residues and signatures. We proposed a four-stage vaccine regimen in Chapter 4 that could theoretically recapitulate these events and will be tested in non-human primate trials in the near future.

Beyond the identification of candidate vaccine immunogens, analysis of CH0848 Env evolution revealed that the plasma virus population split into two persistent viral lineages starting week 186 post-infection (also the first time point that DH270 was detected). This finding, which was briefly covered in Chapter 4, is shown again in Figure 6.2. Analysis of the DH270 bnAb contact residues and neutralization phenotype of these major (X, cross, and triangle) and minor (O) persistent lineages revealed differences. Minor lineage Envs encoded a known V3-glycan bnAb resistance mutation at the base of V3 that shifted a potential N-linked glycosylation (PNG) site from HXB2 position 332 to 334 and were resistant to DH270 neutralization. Major lineage Envs, however, retained the PNG at N332 and exhibited a spectrum of DH270 binding and neutralization phenotypes. Surprisingly, a subset of major lineage Envs (indicated with an X in Figure 6.2) remained neutralization sensitive to contemporaneous DH270 mAbs. I hypothesize

that these DH270 neutralization-sensitive Envs were important for late stages of DH270 development – without sustained antigenic stimulation, DH270 may have been arrested at an earlier stage of maturation and never developed into a cross-reactive bnAb.

Since the initial description of this pattern of lineage divergence in CH0848, we have evaluated Env evolution in other subjects who do and do not develop broadly cross-reactive plasma neutralization. We can detect multiple persistent viral lineages in all subjects who develop bnAbs (N=4), but cannot in subjects who do not develop bnAbs (N=2, Kreider, E., Learn, G., Li, Y., Hahn, B., unpublished). Thus, while these data are preliminary, it appears that a viral correlate of neutralization breadth development may be the establishment and persistence of multiple, genetically distinct viral lineages. Future studies will be needed to increase the number of subjects evaluated for this quasispecies structure and to test whether these distinct viral lineages have different phenotypic properties as was observed in CH0848.

Finally, comparison of virus evolution among human subjects, as I have done above, is confounded by many factors including genetic differences between transmitted founder viruses, the number of transmitted founder virus, viral load setpoint, frequency of sampling, and duration of follow up (41-47). To better understand the viral correlates of bnAb development, a more controlled, reproducible experimental system for virus-antibody co-evolution studies is needed. A recent advancement in simian/human immunodeficiency virus (SHIV) design may be useful in controlling for some of these differences. SHIVs are chimeric viruses that encode an HIV-1 *env* within a SIV backbone and can cause persistent infection in rhesus macaques that recapitulates many of the

pathogenic features of HIV-1 infection (48,49). Li and colleagues have recently developed a method that has permitted the rapid cloning of SHIVs encoding primary, unadapted HIV-1 *envs* (48), including transmitted founder *envs* like those from CH0848 and CH0505 (Li, H., Shaw, G., unpublished). Interestingly, preliminary studies have shown that early virus and antibody evolution over the first year of SHIV.CH0505.TF infection in rhesus macaques recapitulates events that occurred within the human subject CH0505 (Li, H., Williams, W., Haynes, B., Shaw, G., unpublished). Thus, while this experimental model is still being developed, it will likely be an invaluable tool for understanding virus-antibody coevolution in the future.

Virus evolution during bnAb immunotherapy

In the last two years, multiple bnAbs have entered phase I human clinical trials for both safety and efficacy testing (50-53). A recent trial using the CD4 binding site (CD4bs) bnAb 3BNC117 found that viremic subjects experienced a 0.8-2.5 log drop in viral load upon bnAb infusion that was sustained for ~1 month (50). Rebounding viruses from these treated individuals demonstrated increased 3BNC117 neutralization resistance after 1-6 months as compared to pre-treatment (50). In order to understand the potential effects of bnAb therapy on viral populations and/or neutralizing antibody responses, we performed a follow-up longitudinal study of these 3BNC117-treated subjects, which was presented in Chapter 5. Subjects were followed for 6 months after infusion and assessed

for both virus evolution and changes in plasma neutralization. Using a combination of longitudinal single genome sequencing and neutralization assays, we demonstrated that 3BNC117-treated subjects exhibited an increase in heterologous neutralization breadth as compared to untreated controls. Concomitantly, we observed measurable shifts within the virus population using phylogenetic and pairwise sequence comparisons. Interestingly, both increase in heterologous neutralization and virus population shifts were observed regardless of initial 3BNC117 sensitivity.

While 3BNC117 infusion uniformly decreased viral loads, selected for increase bnAb resistance, and boosted host immune responses, the associated changes within the viral quasispecies were not consistent across subjects. As described in Chapter 5, certain subjects exhibited a pattern of "clade shifting," whereas others demonstrated mutational selection within the 3BNC117 Env contact residues. In hindsight, this pattern may have been expected based on a rhesus macaque trial in which SHIV-infected macaques were treated with either 3BNC117 or a V3-glycan bnAb 10-1074 (54). When the rebound viruses were sequenced in this macaque trial and assessed for changes in bnAb contact residues, no mutations were observed in rebound viruses from the 3BNC117-treated macaques. Viruses from 10-1074-treated macaques, however, all harbored deletions of the potential N-linked glycosylation (PNG) site at HXB2 position 332, a known 10-1074 resistance mutation (54,55). 10-1074 has since been infused into viremic HIV-infected human subjects in a Phase I human clinical trial. Similar to sequence evolution in the macaques, nearly 100% of rebound viruses during this trial exhibited the deletion of the PNG site at N332 (Schoofs, T., Kreider, E., Hahn, B., Nussenzweig, M., unpublished).

This difference in virus escape patterns in treatment with a CD4 binding site antibody and a V3-glycan dependent antibody may be due to the nature of the targeted epitope. CD4 binding site antibodies engage the virus via the entry receptor contact residues – if these residues were substantially mutated, the virus may no longer be able to enter target cells (56). Shifts and deletions in the PNG site at HXB2 position 332, however, are commonly seen during autologous neutralizing antibody escape (41,57,58). Futher, because mutations that delete the PNG at N332 often abrogate 10-1074 binding, it will be interesting to see if subjects treated with 10-1074 also exhibit a boost in heterologous neutralization breadth.

Finally, while this study had an interesting and somewhat unexpected outcome, it will require follow-up. Out of the 36 untreated and 15 treated viremic subjects within this trial, we focused our virus sequence analysis on nine individuals who demonstrated a range of 1) viral load setpoints, 2) years of HIV-infection, 3) initial 3BNC117 neutralization sensitivities, and 4) infusion doses (50). Controlling for all of these variables while accounting for virus evolution would be impossible and precludes our ability to make generalizable conclusions. Further, we did not have virus sequencing on matched, untreated controls. Thus, while we interpreted that 3BNC117 infusion was followed by shifts in the virus population, we cannot conclude whether these shifts are larger than that seen in untreated controls.

In Chapters 4 and 5, I evaluated virus evolution during the development of heterologous neutralization breadth in two settings – natural infection and bnAb immunotherapy. Although these two settings are quite different, common themes emerged from our analysis. First, heterologous breadth correlated with the presence of cooperating neutralizing antibody lineages (42). In CH0848, these lineages were DH272 and DH475. In the 3BNC117 trial, the cooperating antibody was the infused antibody, 3BNC117 itself. A second common theme was the evolution of viral diversity during bnAb development. As discussed previously, we have observed multiple persistent viral lineages in longitudinally followed subjects who develop bnAbs. Interestingly, the subjects from the 3BNC117 trial with the broadest initial heterologous neutralization (2A1, 2A3, and 2C4) all harbored viral quasispeices with complex clades structures. Future studies will be needed to increase the number of subjects evaluated for this phenomenon and understand what role antigenic diversity plays in bnAb development.

In conclusion, longitudinal single genome sequencing of plasma HIV-1 can be used to discover novel aspects of virus biology and host-pathogen interactions. Future studies should not only continue to build on these findings, hopefully revealing new aspects of virus and host biology, but also translate these discoveries into novel HIV vaccines and therapeutics.

Section 6.3 – References

1. Bolinger C, Boris-Lawrie K. Mechanisms employed by retroviruses to exploit host factors for translational control of a complicated proteome. Retrovirology. 2009 ed. 2009;6:8.

2. Berkhout B, Arts K, Abbink TE. Ribosomal scanning on the 5'-untranslated region of the human immunodeficiency virus RNA genome. Nucleic Acids Res. 2011 ed. 2011 Jul;39(12):5232–44.

3. Buck CB, Shen X, Egan MA, Pierson TC, Walker CM, Siliciano RF. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. Journal of Virology. 2000 ed. 2001 Jan;75(1):181–91.

4. Brasey A, Lopez-Lastra M, Ohlmann T, Beerens N, Berkhout B, Darlix JL, et al. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. Journal of Virology. 2003rd ed. 2003 Apr;77(7):3939–49.

5. Plank T-DM, Whitehurst JT, Kieft JS. Cell type specificity and structural determinants of IRES activity from the 5' leaders of different HIV-1 transcripts. Nucleic Acids Res. 2013 May 9;41(13):6698–714.

6. Qian S-B, Princiotta MF, Bennink JR, Yewdell JW. Characterization of rapidly degraded polypeptides in mammalian cells reveals a novel layer of nascent protein quality control. J Biol Chem. American Society for Biochemistry and Molecular

Biology; 2006 Jan 6;281(1):392–400.

7. Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW, Bennink JR. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature. 2000 Apr 13;404(6779):770–4.

8. Reits EA, Vos JC, Gromme M, Neefjes J. The major substrates for TAP in vivo are derived from newly synthesized proteins. Nature. 2000 ed. 2000 Apr 13;404(6779):774–8.

9. Antón LC, Yewdell JW. Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors. J Leukoc Biol. 2014 ed. 2014 Apr;95(4):551–62.

10. Robb ML, Eller LA, Kibuuka H, Rono K, Maganga L, Nitayaphan S, et al. Prospective Study of Acute HIV-1 Infection in Adults in East Africa and Thailand. N Engl J Med. Massachusetts Medical Society; 2016 Jun 2;374(22):2120–30.

11. Frohman MA. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. Meth Enzymol. 1993;218:340–56.

12. Ternette N, Block PD, Sanchez-Bernabeu A, Borthwick N, Pappalardo E, Abdul-Jawad S, et al. Early Kinetics of the HLA Class I-Associated Peptidome of MVA.HIVconsv-Infected Cells. Journal of Virology. 2015 ed. 2015 Jun;89(11):5760–71.

13. Schubert U, Antón LC, Gibbs J, Norbury CC, Yewdell JW, Bennink JR. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature. 2000 ed. 2000 Apr 13;404(6779):770–4.

14. Hansen SG, Sacha JB, Hughes CM, Ford JC, Burwitz BJ, Scholz I, et al. Cytomegalovirus vectors violate CD8+ T cell epitope recognition paradigms. Science. 2013 May 24;340(6135):1237874.

15. Hansen SG, Vieville C, Whizin N, Coyne-Johnson L, Siess DC, Drummond DD, et al. Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. Nature Medicine. 2009 Mar;15(3):293–9.

16. Hansen SG, Piatak MJ, Ventura AB, Hughes CM, Gilbride RM, Ford JC, et al. Immune clearance of highly pathogenic SIV infection. Nature. 2013 ed. 2013 Oct 3;502(7469):100–4.

17. George AJT, Stark J, Chan C. Understanding specificity and sensitivity of T-cell recognition. Trends Immunol. 2005 Dec;26(12):653–9.

18. Goonetilleke N, Moore S, Dally L, Winstone N, Cebere I, Mahmoud A, et al. Induction of multifunctional human immunodeficiency virus type 1 (HIV-1)-specific T cells capable of proliferation in healthy subjects by using a prime-boost regimen of DNA- and modified vaccinia virus Ankara-vectored vaccines expressing HIV-1 Gag coupled to CD8+ T-cell epitopes. Journal of Virology.

American Society for Microbiology; 2006 May;80(10):4717–28.

19. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, et al. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nature Medicine. 1997 Feb;3(2):205–11.

20. Hooijberg E, Ruizendaal JJ, Snijders PJ, Kueter EW, Walboomers JM, Spits H. Immortalization of human CD8+ T cell clones by ectopic expression of telomerase reverse transcriptase. J Immunol. 2000 Oct 15;165(8):4239–45.

21. Hansen SG, Sacha JB, Hughes CM, Ford JC, Burwitz BJ, Scholz I, et al. Cytomegalovirus vectors violate CD8+ T cell epitope recognition paradigms. Science. 2013 ed. 2013 May 24;340(6135):1237874.

22. Hansen SG, Ford JC, Lewis MS, Ventura AB, Hughes CM, Coyne-Johnson L, et al. Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. Nature. 2011 May 26;473(7348):523–7.

23. Kuiken C, Foley B, Leitner T, Apetrei C, Hahn B, Mizrachi I, et al. HIV Sequence Compendium. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR 12-24653; 2012.

24. McAuley JL, Hornung F, Boyd KL, Smith AM, McKeon R, Bennink J, et al. Expression of the 1918 influenza A virus PB1-F2 enhances the pathogenesis of viral and secondary bacterial pneumonia. Cell Host and Microbe. 2007 Oct

11;2(4):240–9.

25. Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, et al. A novel influenza A virus mitochondrial protein that induces cell death. Nature Medicine. Nature Publishing Group; 2001 Dec;7(12):1306–12.

26. Gibbs JS, Malide D, Hornung F, Bennink JR, Yewdell JW. The influenza A virus PB1-F2 protein targets the inner mitochondrial membrane via a predicted basic amphipathic helix that disrupts mitochondrial function. Journal of Virology. American Society for Microbiology (ASM); 2003 Jul;77(13):7214–24.

27. Muthuswami R, Chen J, Burnett BP, Thimmig RL, Janjic N, McHenry CS. The HIV plus-strand transfer reaction: determination of replication-competent intermediates and identification of a novel lentiviral element, the primer over-extension sequence. Journal of Molecular Biology. 2002nd ed. 2002 Jan 18;315(3):311–23.

28. Auxilien S, Keith G, Le Grice SF, Darlix JL. Role of post-transcriptional modifications of primer tRNALys,3 in the fidelity and efficacy of plus strand DNA transfer during HIV-1 reverse transcription. J Biol Chem. 1999 ed. 1999 Feb 12;274(7):4412–20.

29. Alexander L, Denekamp L, Czajak S, Desrosiers RC. Suboptimal Nucleotides in the Infectious, Pathogenic Simian Immunodeficiency Virus Clone SIVmac239. Journal of Virology. 2001 Apr 15;75(8):4019.

30. Fennessey CM, Reid C, Lipkey L, Newman L, Oswald K, Piatak M, et al. Generation and characterization of a SIVmac239 clone corrected at four suboptimal nucleotides. Retrovirology. 2015 ed. 2015 Jun 16;12(1):49.

31. Berkhout B, Das AT. On the primer binding site mutation that appears and disappears during HIV and SIV replication. Retrovirology. 2015 ed. 2015;12:75.

32. Tang LY, Zhang J. The cellular mismatch repair system is able to repair mismatches within MLV retroviral double-stranded DNA at a low frequency. Nucleic Acids Res. Oxford University Press; 2000 Jun 15;28(12):2302–6.

33. Berwin B, Barklis E. Retrovirus-mediated insertion of expressed and non-expressed genes at identical chromosomal locations. Nucleic Acids Res. Oxford University Press; 1993 May 25;21(10):2399–407.

34. Li SKH, Martin A. Mismatch Repair and Colon Cancer: Mechanisms and Therapies Explored. Trends Mol Med. 2016 Apr;22(4):274–89.

35. van Bel N, Ghabri A, Das AT, Berkhout B. The HIV-1 leader RNA is exquisitely sensitive to structural changes. Virology. 2015 Sep;483:236–52.

36. van Bel N, Das AT, Berkhout B. In Vivo SELEX of Single-Stranded Domains in the HIV-1 Leader RNA. Journal of Virology. 2014 Jan 31;88(4):1870–80.

37. Haynes BF, Shaw GM, Korber B, Kelsoe G, Sodroski J, Hahn BH, et al. HIV-Host Interactions: Implications for Vaccine Design. Cell Host and Microbe. 2016 Mar

9;19(3):292–303.

38. Eroshkin AM, LeBlanc A, Weekes D, Post K, Li Z, Rajput A, et al. bNAber: database of broadly neutralizing HIV antibodies. Nucleic Acids Res. Oxford University Press; 2014 Jan;42(Database issue):D1133–9.

39. Kwong PD, Mascola JR. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. Immunity. 2012 Sep 21;37(3):412–25.

40. Kwong PD, Mascola JR, Nabel GJ. Rational design of vaccines to elicit broadly neutralizing antibodies to HIV-1. Cold Spring Harb Perspect Med. Cold Spring Harbor Laboratory Press; 2011 Sep;1(1):a007278–8.

41. Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature. 2013 Apr 25;496(7446):469–76.

42. Gao F, Bonsignori M, Liao H-X, Kumar A, Xia S-M, Lu X, et al. Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. Cell. 2014 Jul 31;158(3):481–91.

43. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, et al. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature. Nature Research; 2014 May 1;509(7498):55–62.

44. Bonsignori M, Hwang KK, Chen X, Tsao CY, Morris L, Gray E, et al. Analysis of

a Clonal Lineage of HIV-1 Envelope V2/V3 Conformational Epitope-Specific Broadly Neutralizing Antibodies and Their Inferred Unmutated Common Ancestors. Journal of Virology. 2011 Sep 9;85(19):9998–10009.

45. MacLeod DT, Choi NM, Briney B, Garces F, Ver LS, Landais E, et al. Early Antibody Lineage Diversification and Independent Limb Maturation Lead to Broad HIV-1 Neutralization Targeting the Env High-Mannose Patch. Immunity. 2016 May 17;44(5):1215–26.

46. Bhiman JN, Anthony C, Doria-Rose NA, Karimanzira O, Schramm CA, Khoza T, et al. Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. Nature Medicine. 2015 Nov;21(11):1332–6.

47. Kong R, Xu K, Zhou T, Acharya P, Lemmin T, Liu K, et al. Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. Science. American Association for the Advancement of Science; 2016 May 13;352(6287):828–33.

48. Li H, Wang S, Kong R, Ding W, Lee F-H, Parker Z, et al. Envelope residue 375 substitutions in simian-human immunodeficiency viruses enhance CD4 binding and replication in rhesus macaques. Proc Natl Acad Sci USA. National Acad Sciences; 2016 Jun 14;113(24):E3413–22.

49. Shingai M, Donau OK, Schmidt SD, Gautam R, Plishka RJ, Buckler-White A, et al. Most rhesus macaques infected with the CCR5-tropic SHIV(AD8) generate cross-reactive antibodies that neutralize multiple HIV-1 strains. Proc Natl Acad

Sci USA. National Acad Sciences; 2012 Nov 27;109(48):19769–74.

50. Caskey M, Klein F, Lorenzi JCC, Seaman MS, West AP, Buckley N, et al. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. Nature. Nature Research; 2015 Jun 25;522(7557):487–91.

51. Schoofs T, Klein F, Braunschweig M, Kreider EF, Feldmann A, Nogueira L, et al. HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1. Science. American Association for the Advancement of Science; 2016 May 20;352(6288):997–1001.

52. Lynch RM, Boritz E, Coates EE, DeZure A, Madden P, Costner P, et al. Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. Sci Transl Med. American Association for the Advancement of Science; 2015 Dec 23;7(319):319ra206–6.

53. Scheid JF, Horwitz JA, Bar-On Y, Kreider EF, Lu C-L, Lorenzi JCC, et al. HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. Nature. Nature Research; 2016 Jul 28;535(7613):556–60.

54. Shingai M, Nishimura Y, Klein F, Mouquet H, Donau OK, Plishka R, et al. Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia. Nature. Nature Research; 2013 Nov 14;503(7475):277–80.

55. Mouquet H, Scharf L, Euler Z, Liu Y, Eden C, Scheid JF, et al. Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. Proc Natl Acad

Sci USA. National Acad Sciences; 2012 Nov 20;109(47):E3268–77.

56. Lynch RM, Wong P, Tran L, O'Dell S, Nason MC, Li Y, et al. HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. Doms RW, editor. Journal of Virology. American Society for Microbiology; 2015 Apr;89(8):4201–13.

57. Moore PL, Gray ES, Wibmer CK, Bhiman JN, Nonyane M, Sheward DJ, et al. Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. Nature Medicine. Nature Publishing Group; 2012 Nov;18(11):1688–92.

58. Moody MA, Gao F, Gurley TC, Amos JD, Kumar A, Hora B, et al. Strain-Specific V3 and CD4 Binding Site Autologous HIV-1 Neutralizing Antibodies Select Neutralization-Resistant Viruses. Cell Host and Microbe. 2015 Sep 9;18(3):354–62.

Section 6.4 – Figures

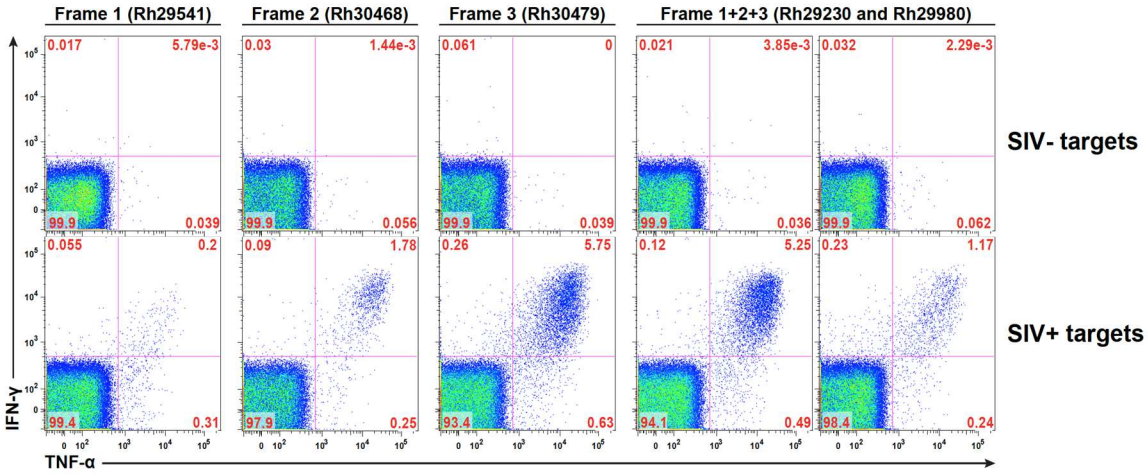


Figure 6-1 RhCMV68.1-5'L vaccination of five macaques

Reading frame +1, +2, and +3 of the SIVmac766 5' leader were each incorporated into a separate RhCMV68.1 vaccine vector. Macaques were immunized against RF+1, RF+2, RF+3, or all three. PBMCs from vaccinated macaques were then incubated with autologous T cells (taken prior to vaccination) that were either infected or not infected with SIVmac766. T cell recognition of infected cells was assayed with interferon-γ and TNF-α intracellular cytokine staining. SIV-specific T cell responses were elicited in macaques immunized against RF+1, RF+2, RF+3, or all three.

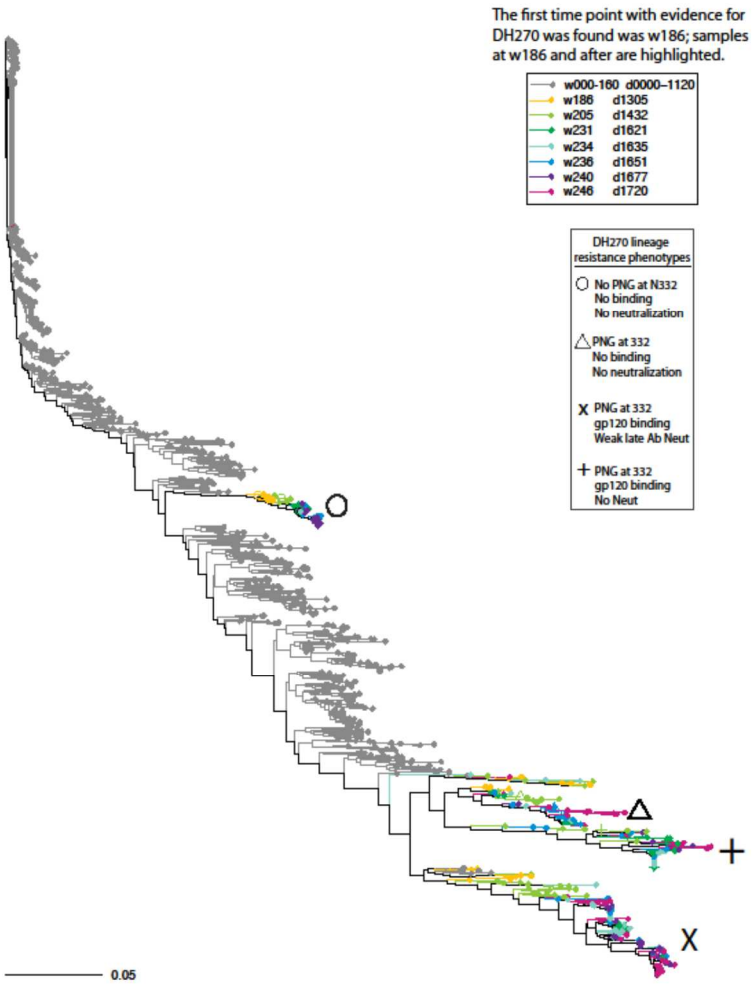


Figure 6-2 Multiple viral lineages persisted during bnAb development

1,223 Env protein sequences translated from single genome sequences were used to generate a Maximum Likelihood phylogenetic tree rooted on the transmitted founder sequence. Sequences sampled prior to the development of Tier 2 heterologous breadth (week 186) are shaded in grey and sequences from after week 186 are highlighted. By week 186, the quasispecies resolved into at least two distinct lineages with different DH270 binding and neutralization profiles. Envs indicated with a circle simultaneously lack the potential N-linked glycosylation (PNG) site (NXS/T) at HXB2 position 301 and shift the PNG site at 332 to 334. Loss of critical V3 glycan bnAb-binding amino acids conferred complete resistance to DH270 lineage Abs, and by week 246 this virus was undetected. Binding and neutralization assays revealed that certain variants escape DH270 lineage pressure by abrogating antibody binding (circle and triangle). Others

maintain binding while disrupting neutralization (cross) or maintain both binding and neutralization (X).