

**Antisense-Derived HIV-1 Cryptic Epitopes Are Not Major Drivers of Viral Evolution
during the Acute Phase of Infection**

Binghao J. Peng^a, Jonathan M. Carlson^b, Michael K. P. Liu^{c*}, Feng Gao^d, Nilu Goonetilleke^{c*},
Andrew J. McMichael^c, Persephone Borrow^c, Jill Gilmour^e, Sonya L. Heath^a, Eric Hunter^f, Anju
Bansal^a, Paul A. Goepfert^{a#}

^aDepartments of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

^bMicrosoft Research, Redmond, Washington, USA

^cNuffield Department of Medicine, University of Oxford, Oxford, UK

^dDepartment of Medicine, Duke Human Vaccine Institute, Duke University Medical Center,
Durham, North Carolina, USA

^eIAVI Human Immunology Laboratory, Imperial College London, London, UK

^fEmory Vaccine Center at Yerkes National Primate Research Center and Department of
Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, USA.

Running Title: Cryptic epitopes and viral evolution

#Address correspondence to Paul A. Goepfert, pgoepfert@uabmc.edu.

*Present address: Michael K. P. Liu, Center for Immunology and Vaccinology, Imperial College
London, London, United Kingdom; Nilu Goonetilleke, Dept. of Microbiology and Immunology,
University of North Carolina at Chapel Hill, North Carolina, USA.

Word counts: Abstract: 186, Main text: 3705

ABSTRACT

While prior studies have demonstrated that CD8 T cell responses to cryptic epitopes (CE) are readily detectable during HIV-1 infection, their ability to drive escape mutations following acute infection is unknown. We predicted 66 CE in a Zambian acute infection cohort based on escape mutations occurring within or near the putatively predicted HLA-I restricted epitope. The CE were evaluated for CD8 T cell responses in patients with chronic and acute HIV infection. Of the 66 predicted CE, 10 were recognized in 8/32 and 4/11 patients with chronic, and acute infection respectively. The immunogenic CE were all derived from a single antisense reading frame within *pol*. However, when these CE were tested using longitudinal study samples, CE specific T cell responses were detected but did not consistently select for viral escapes. Thus, while we demonstrated that CE are immunogenic in acute infection, the immune responses to CE are not major drivers of viral escape in the initial stages of HIV infection. This latter finding may be due to either the subdominant nature of CE-specific responses, the low antigen sensitivity, and magnitude of CE responses during acute infections.

IMPORTANCE

Although prior studies demonstrated that cryptic epitopes of HIV-1 induce CD8 T cell responses, evidence supporting that targeting these epitopes to drive HIV escape mutations have been substantially limited and none have addressed this question following acute infection. In this comprehensive study, we utilized longitudinal viral sequencing data obtained from three separate acute infection cohorts to predict potential cryptic epitopes based on HLA-I associated viral escape. Our data shows that cryptic epitopes are immunogenic during acute infection and many of these responses are elicited towards translation products of HIV-1 antisense reading frames.

47 However, despite cryptic epitope targeting, our study did not find any evidence of early CD8
48 mediated immune escape. Nevertheless, improving cryptic epitope specific CD8 T cell responses
49 may still be beneficial in both preventative and therapeutic HIV-1 vaccines.

50

INTRODUCTION

CD8 T cells play a vital role in controlling intracellular pathogens by recognizing epitopes presented on infected cells. In human immunodeficiency virus type 1 (HIV-1) infections, CD8 T cells have been shown to be essential in control of viremia and disease progression, and part of this evidence comes from the fact that they drive HIV escape mutations that may impact overall viral fitness (1). However, due to the mutable nature of HIV, achieving a broader range of epitope recognition by targeting beyond the scope of the nine conventional proteins of the virus may be necessary for augmenting the efficacy for both preventative and therapeutic vaccines (2, 3). Several studies have described a lesser known form of epitopes within several viruses including HIV and SIV, termed cryptic epitopes (CE)(4-24). These epitopes are predominantly derived from the translation of two sense and three antisense alternative open reading frames. Translation products of any of these five alternate reading frames (ARF) were found to be recognized by CD8 T cells during the acute and chronic stages of infection (6, 21, 25-29). While CE usually induce T cell responses of lower magnitude when compared to responses induced by their main reading frame counterparts, past studies have demonstrated that immune responses to these ARF epitopes are observed in many infected individuals or are immunoprevalent during the course of the disease. (25, 27). Furthermore, HIV was demonstrated to escape CE-specific CD8 T cell responses in chronic infection, demonstrating that they exert significant selective pressure on viral replication *in vivo*. However, the contribution made by such responses in driving viral evolution during acute and early HIV-1 infection has not been adequately addressed.

Here, we utilized a variety of methods to predict CE that potentially induced CD8 T cell responses during the acute phase of infection. By validating the epitope predictions from one of

the acute cohorts through immunogenicity screening assays, we identified ten previously undescribed CE, to be immunogenic in both acute and chronic infection. However, no reliable CE responses can be detected when using similar methods of epitope prediction in two additional acute cohorts. From this study, we conclude that CE are indeed present during early stages of infection and can be targeted by T cells. Nevertheless, due to the overall subdominant T cell response seen to these epitopes, they appear to contribute minimally in driving viral adaptation.

RESULTS

Cryptic epitope predictions and T cell responses

To analyze the role of CE targeting during acute infections, we adopted an approach similar to previous studies, utilizing viral sequence data obtained from a serodiscordant cohort (30, 31). We first identified sites of HLA-I associated polymorphisms within or near previously predicted CE that may have evolved to escape CD8 T cell responses during the first two years of infection in 78 acutely infected Zambian patients. Due to the limited Zambian sample availability, and to facilitate analysis in subtype B HIV infected patients, we narrowed the list of all predicted epitopes down to 66 putative CE for testing (Table 1). The CE were first tested as seven distinct peptide pools (6-10 peptides/pool) for recognition by PBMCs from chronically infected HIV-1 subtype B patients in an IFN- γ ELISpot assay. Thirty-two patient samples were used. 78% (25/32) patients were not on active ART, with the median viral load of the cohort being 3,380 RNA copies/ml and median absolute CD4 count of 646 (Table 2). We observed T cell responses to the CE peptide pools in 25% of the subjects tested (8/32, Table 2, shaded in gray and Figure 1A). In contrast, CE pools did not elicit T cell responses in any of the 16 HIV seronegative donors ($p < 0.0001$) evaluated. Next, we deconvoluted the positive pool responses to the single

peptide level. Single peptide responses were successfully deconvoluted from six out of the eight CE pool responders, which allowed us to identify ten novel immunogenic CE (Table 3) with no significant homology to HIV-1 proteins encoded by the main reading frame (Table 4).

Interestingly, all immunogenic CE were derived from the antisense frame 3 (RF3) of *pol*. The magnitude of the responses to these CE in the six patients tested varied, ranging from 63 to 487 SFU/10⁶ PBMC (Figure 1B, median magnitude of 133 SFU/10⁶ PBMC).

Polyfunctionality of T cells responding to CE

Since T cell responses to cryptic epitopes WL9 and FY11 were frequently observed in the ELISpot assays, ICS assays were employed to determine CD4/CD8 restriction of CE-reactive T cell responses. The CE specific CD8 T cells produced IFN- γ and upregulated perforin expression in an *ex vivo* intracellular cytokine staining assay, confirming that CE could elicit antiviral responses (Figure 2A). To demonstrate the ability of these cells to kill virally infected targets, we generated CD8 T cell lines using two rounds of peptide pulsed monocytes as antigen presenting cells and tested their effector functions in an ICS assay as described in methods. CD8 T cells specific for WL9 and FY11 were able to expand after the two-week culture and were highly specific for CE based on effector molecule expression after re-stimulation (Figure 2B). Furthermore, these CE specific cell lines were capable of killing HIV(NL4.3) infected HLA-I matched target cells *in vitro*, demonstrating their effectiveness at eliminating HIV infected target cells (Figure 2C, p-value <0.001).

CD8 T cell responses to *pol* CE in acute infection

We next analyzed samples from the Zambian cohort to determine whether T cell responses to the 10 *pol* CE epitopes could drive viral escape during acute and early infection. With limited quantities of PBMC samples remaining from the cohort, we were only able to identify two

120 Zambian patients with PBMCs available who showed a sequence-based mutation in or around
121 any of the CE. These patients each had a synonymous mutation in *pol*, which generated a fixed
122 amino acid change near the CE-KL10 epitope. For ZA1, the mutation did not occur until 24
123 months post estimated date of infection, and for ZA2 the mutation occurred very early,
124 detectable by month 3 after transmission of the virus (Figure 3A). No T cell responses were
125 detected by IFN- γ ELISPOT assay in ZA1 at either 590 or 841dpi. The magnitude of the IFN- γ
126 response for ZA2 at 280 days post infection was below our criteria for a positive response, but
127 the IFN- γ level was subsequently increased to 150 SFU/10⁶ PBMC at 389 days post infection
128 (Figure 3B). To examine whether CE specific T cell responses were indeed stronger in earlier
129 stages of infection, we examined 11 subtype B individuals with PBMC samples collected within
130 the first three months of infection (Table 2). When tested, 4/11 patients showed T-cell responses
131 to at least one of the novel CE identified, with the highest magnitude of response being detected
132 at 443 SFU/10⁶ PBMC (Figure 3C). CE Antigen sensitivity testing was further performed using
133 acute patient sample (Figure 4A and B). Although at higher EC50 levels, the CE tested show
134 comparable avidity to some main reading frame epitopes tested in the past (32, 33). To
135 strengthen our finding, we examined whether the emergence of synonymous mutations in early
136 HIV infections resulted from CD8 T cell targeting of cryptic epitopes in two additional acute
137 cohorts (Rwandan/Ugandan and CHAVI001, Supplemental Table 1). T cell responses to putative
138 autologous virus CE were tested using IFN- γ ELISpot assays as described previously (*ex-vivo*
139 and cultured). No CE-specific T cell responses were detected during acute/early infection in the
140 16 individuals studied (data not shown), suggesting that although CE can be targeted by CD8 T
141 cells, it may not be common for these responses to drive viral escapes during early stages
142 infection.

DISCUSSION

Beginning with the discovery of antisense reading frame protein specific antibodies in the serum of HIV-1 infected individuals nearly two decades ago (22), alternative reading frame transcription and translation have remained an exciting and active facet of viral immunology research (8, 9, 19, 23, 34). Several groups have since then reported immunogenicity of peptides derived from the ARF of HIV-1 that can trigger CD8 T cell activation and effector functions. Although some studies have attempted to unravel the biological role of these alternate reading frame derived proteins and their derivative epitopes, the exact role of cryptic epitopes and their relationship with host immunity remains an enigma. The potential ability for HIV to produce up to six different amino acid combinations per single genomic coding region remains a tantalizing idea for broadening T cell responses in vaccine designs. Although many have described the ability of CE to trigger CD8 T cell responses in both acute and chronic stages of HIV-1 and SIV infection (20, 25, 27, 35, 36), our findings here do not indicate that CE-specific CD8 T cells select for HIV escape mutations. This conclusion is based on longitudinal data performed on 16 patients following acute infection whereby we did not demonstrate evolutionary pressure from cryptic epitope-specific CD8 T cells. This finding is in sharp contrast to what was observed for CD8 T cells targeting epitopes in the main reading frame in the CHAVI001 cohort (42). 11 participants of this cohort demonstrated an average of 4 CD8 T cell responses (range: 2-11) to the main reading frame epitopes that were temporally associated with escape. In a similar fashion, the two Rwandan individuals had 12 and 13 total T cell responses to the main reading frame (48). When all immune data were combined, a median of 3 main reading frame CD8 T cell responses were detected which were associated with viral escape in these cohorts. These numbers are significantly different than what was seen for cryptic epitopes tested, where we

showed only one possible detectable T cell responses in the 176 cryptic epitopic regions tested (Fisher's exact test: $p < 0.0001$).

The Zambian sequence data allowed us to identify ten novel cryptic epitopes, each being derived from the antisense reverse frame 3 in the *pol*-coding region of HIV. Perhaps the most intriguing and perplexing observation in our study is the fact that half of the novel immunogenic CE do not appear to be a translation product of an ORF initiated from the typical AUG^{met} start codon. In addition, *in silico* analysis of HIV-1 consensus sequence shows an enrichment of CUG^{leu} initiated ORF within reverse frame 3 of the virus (data not shown). With *pol* being the largest gene of the virus, this observation could help explain why all our immunogenic CE were derived from the reverse frame 3 of *pol*. Nevertheless, this is a similar phenomenon to what was previously observed by Berger et al. (27), where an immunogenic CE of HIV from the alternative forward frame was also found to be initiated by a leucine (e.g., CUG^{leu}) start codon. As in this prior study, we too show killing of virus-infected cells by a CD8 T cell line specific to a cryptic epitope translated from a leucine initiated ORF (Figure 2C). Furthermore, the studies of Shastri et al. in recent years have provided evidence in both cell lines and animal models, of how CUG^{leu} start codons are frequently used by cells to generate epitopes that can trigger T cell activation (37-40). Moreover, using *in silico* analysis, we have found an enrichment of CUG ORF within the antisense RF3 of HIV-1 consensus sequence (data not shown), providing further anecdotal explanation to our data. In conjunction with our observation of multiple CE being translated from a CUG initiated ORF, we hypothesize that the production of HIV CE relies heavily on the utilization of alternative start codons. It has been shown that the translation efficiency of CUG initiated ORFs appears to depend heavily on other molecular mechanisms and environmental factors can impact the protein translation process. In particular, inflammation and

cytokine milieu seem to have a dramatic effect on the translation efficiency of CUG^{leu} initiated ORF (41). When pro-inflammatory cytokines such as type I and type II interferons and tumor necrosis factor α were added exogenously to cell lines, CUG^{leu} ORF-encoded peptides were, in turn, more efficiently produced. We speculate that CE production from CUG^{leu} initiated ORFs may therefore be modulated by the levels of inflammatory cytokines produced during different stages of infection.

We have previously shown that the magnitude of CD8 T cell responses to CE are often lower during acute than in chronic HIV-1 infection (25). Unfortunately, with the limited availability of acute Zambian samples, we were unable to directly test the magnitude of T cell responses to the predicted CE immediately after infection. Nevertheless, in the longitudinal Zambian samples obtained, we could capture archival T cell responses to CE-KL10 definitively in one subject (ZA2) but not in the other (ZA1). The lone T cell responses to CE-KL10 was detected at time-points that were well into the chronic stage of infection. The kinetics of selection of T cell escape mutations during acute and early HIV-1 infection is strongly dependent on both the immunodominance of the epitope-specific T cell response, and the fitness costs associated with acquisition of escape-conferring mutations in and around the epitope sequence (42-46). Although acquisition of escape mutations in epitopes derived from alternative open reading frames may potentially incur lower average costs to viral fitness than for epitopes derived from the main reading frames, the fact that responses to CE are typically of very low magnitude during acute infection, likely explains why virus escape is slower and not detected until later in infection.

Another potential explanation for the lack of escape can be the result of lower antigen avidity noticed for some of the CE in acute samples (Figure 4). In line with all observations, an SIV-macaque study published by Harris et al. provided evidence for strong CD8 T cell responses to

CE mainly when the infecting SIV strain has had an accumulation of mutations in epitopes derived from the main reading frame (47). This would indicate that HIV does evolve to escape CE-T cell responses but does so more during the chronic stages of infection, where CE are generated more frequently due to the accumulation of both inflammatory stimuli and viral escape has occurred from responses to immunodominant epitopes in the main reading frame (41, 47). Together, our data support the notion that cryptic epitopes are translated by HIV during acute stages of infection, but due to their subdominant nature during natural infections, selection pressure on these epitopes is likely minimal. However, with vaccination strategies that specifically target cryptic epitopes, frequencies of these responses would likely increase, augmenting the immune pressure targeted at these ARF regions of the virus. Future studies are needed to determine if vaccine boosted HIV-CE immune responses can enhance the overall quality of response directed against the virus.

MATERIALS AND METHODS

Patient Cohorts: Autologous HIV-1 sequence and HLA-I allelic data from three separate acute infection cohorts were used in this study for cryptic epitope prediction. These included data from a small cohort of two Rwandans and two Ugandans, Zambian (N=78) patients enrolled in the International AIDS Vaccine Initiative (IAVI) protocol C study, and patients enrolled in the CHAVI001 study (N=12). Putative CE predicted based on information from the Zambian cohort (see below) were tested in 32 patients with chronic infection and 11 with acute infection enrolled at the 1917 HIV-1 clinic at the University of Alabama at Birmingham (UAB). Healthy donor PBMC from the Alabama Vaccine Research Clinic were used as HIV-1 seronegative controls. IRB approval was obtained, and all participants consented to this study. Demographic details of the 1917 clinic cohort can be found in Table 2.

CE prediction and peptide synthesis: For the patients in the Rwandan/Ugandan and CHAVI001 cohorts, HIV-1 single genomic amplification (SGA) and sequencing were performed at serial post-infection timepoints as previously described (42, 48) (CHAVI001 cohort sequences can be accessed via Los Alamos HIV sequence database, <http://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>, PubMed ID 23221345), and viral sequences were evaluated for the emergence of synonymous mutations (with respect to the main reading frame). Peptides were designed that overlapped the region containing each synonymous mutation, and putative optimal epitopes in this region were also predicted using either Microsoft EpiPred (<http://boson.research.microsoft.com/bio/epipred.aspx>), or LANL ELF software (http://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer.html). Following acute HIV infection of the patients in the Zambian cohort, bulk plasma viral sequences were previously analyzed at the time of infection and every 3-6 months after that for two years (30, 31). We used prior predicted CE using HLA-I associated polymorphism in the ARF of HIV in subtype C chronically infected patients (25). We then applied these CE predictions to the longitudinal sequence data obtained from 78 patients in a linked transmission cohort in Zambia to determine if synonymous mutations were fixed in or within 3 amino acids of these epitopes during the first two years of follow up. A fixed mutation was defined as being present for at least the penultimate and the final sequencing time points. A list of likely previously predicted epitopes (25) containing the fixed mutation were generated based on HLA-I of the recipient and the sequence flanking the fixed mutation. Additional criteria were adopted in selecting CE including predicted CE that had a posterior probability greater than 50%, greater than 80% peptide homology to subtype B consensus sequence at the amino acid level, and the mutation observed in least three or more individuals. Based on these criteria, 66 epitopes were

selected for testing in the Zambian cohort. All peptides were synthesized in a 96 well peptide array format (New England Peptides (Gardner, MA) or as previously described (42)). The peptides were reconstituted in DMSO and stored at -80°C. The peptides were tested in pools (6-10 peptides/pool), and responses were further mapped to the single peptide level.

IFN- γ ELISpot assay: ELISpot assays were performed as previously described by our group and others (25, 26, 28, 42, 44, 48-50). In brief, nitrocellulose 96-well plates were first coated with anti-IFN- γ monoclonal antibody. PBMCs, rested for 2-6 hrs after thawing, were then incubated in the plates (100,000~200,000 cells/well) in at least duplicates with the appropriate antigen for 20-24 hrs at 37°C in 5% CO₂. The final concentrations of CE used were 2 μ M (peptide pools) and 10 μ M (single peptide). PBMCs were washed off (PBS and Tween-20) after overnight culture, and biotinylated anti-IFN- γ antibody (Mabtech) was added to the plates for 2hr at room temperature. Another round of washing was performed, followed by the addition of streptavidin-conjugated alkaline phosphatase (Mabtech) for 45 minutes. Finally, NBT/BCIP (Southern Biotech) was added to develop spots as readout. ELISpot reader (C.T.L.) counted individual spot forming units (SFU), and the counts were then normalized to SFU/10⁶ PBMCs. Criteria for positive ELISpot responses were as described previously (25, 26, 42, 48, 50).

Intracellular cytokine staining (ICS) assay: Cryopreserved PBMCs were stained in an ICS assay following stimulation with appropriate CE peptides at 10 μ M final concentration. In brief, co-stimulatory monoclonal antibodies anti-CD28 and anti-CD49d, each at 2 μ g/ml was added to each tube containing 10⁶ PBMCs. Cells were then stimulated with the appropriate peptide, monensin, and brefeldin A for 12 hrs at 37°C. The cells were then washed and stained with Live/Dead cell dye (Invitrogen) for 30 min, washed again, and stained for surface markers for 30 min. Surface labeling of cells was done using anti-CD3 (Pacific Blue), anti-CD8 (V500), anti-

282 CD4 (Alexa Fluor 780) conjugated antibodies. Anti-CD14 and anti-CD19 labeled with PerCp-
283 Cy5.5 were used as dump channels. Cells were washed with PBS before treating with
284 Cytotfix/Cytoperm reagent (BD) for 20 min at RT. Following permeabilization, cells were
285 labeled with intracellular antibodies for 30 min at room temperature Intracellular staining was
286 performed using anti-IFN γ (Alexa Fluor 700), anti-IL-2 (APC), anti-TNF α (PE-Cy7), and anti-
287 Perforin (PE). Stained cells were fixed with 2% paraformaldehyde and acquired on a BD LSRII
288 flow cytometer. Data collected were analyzed using FlowJo software version 9.9.7 (Tree Star,
289 Inc.).

290 **CD8 T cell line generation:** Epitope-specific CD8⁺ T cell lines were generated from patient
291 PBMC initially identified with a positive response to CE based on the IFN- γ ELISpot assay. As
292 previously described by our group, cell lines were *in vitro* expanded over the course of 14 days
293 using two rounds of irradiated monocytes pulsed with peptide as antigen presenting cells (26,
294 44). In brief, at the end of 2 weeks, CD8 T cell lines were re-stimulated with cognate antigen in
295 the presence of co-stimulatory antibodies and intracellular transport inhibitors (monensin,
296 brefeldin A) for 12 hrs and analyzed via flow cytometry for effector function, as described
297 above.

298 ***In-vitro* viral killing assay (iVKA):** To obtain target cells needed for viral infection, PBMC
299 from HIV seronegative donors were first depleted of CD8⁺ T cells using magnetic bead
300 separation (CD8 Dynabeads, Invitrogen). CD8⁻ PBMC were then activated with PHA (5ug/ml)
301 and IL-2 (50U/ml) for two days before infection with HIV NL4.3 virus at MOI of 0.5. Two days
302 post infection, 1 x 10⁵ cells of the CE-FY11 specific CD8⁺ T cell line (HLA-A*0201/A*0201,
303 B*1503/B*3501, C*0210/0401) were then added to infected HLA-matched (Match target 1:
304 HLA-A*0201/A*0201, B*5501/B*5801, C*0303/C*0701, Match target 2: HLA-

305 A*0201/A*0201, B*4402/B4402, C*0501/0501), and completely HLA-mismatched targets
306 (HLA-A*0301/A*3101, B*0702/B*4001, C*0304/C*0702) at 0:1, 1:1, at 3:1 effector/target
307 (E/T) ratios and co-cultured in duplicate for 48 hrs at 37°C in 5% CO₂. Cells were then stained
308 with Live/dead stain and surface markers using anti-CD3 (Pacific Blue), anti-CD8 (V500), anti-
309 CD4 (Alexa Fluor 780), anti-CD14 (Percpcy5.5) conjugated antibodies. Following PBS wash,
310 cells were first permeabilized with Perm A reagent (Invitrogen), then stained intracellularly with
311 Gag p24-PE antibody (Beckman Coulter) diluted in Perm B buffer (Invitrogen) according to the
312 manufacturer's instructions. Gates for p24 expression were set based on uninfected targets.
313 Percentage of killing for each E/T ratio was determined by the following formula:
314 $\% \text{ killing} = 1 - \left(\frac{\% P24+effector}{\% P24 \text{ without effector}} \right) \times 100$. For purpose of calculating statistical significance,
315 values obtained for the two HLA matched targets were combined (Figure 2C).
316 **Antigen sensitivity:** Serial 10-fold dilutions of peptides were used in IFN-γ ELISPOT assays to
317 stimulate functional responses. CE antigen sensitivity was measured as the peptide concentration
318 eliciting 50% of the maximal IFN-γ response, or EC50, which was calculated and graphed using
319 GraphPad Prism software (version 7.0). Additional evaluations were based on response
320 magnitude (SFU/10⁶ PBMCs).
321 **Statistics:** Fisher's tests, paired test, and nonparametric Mann-Whitney U test were used to
322 determine statistical significance in this study. An area-under-the-curve based paired *t*-test was
323 used to determine significance between p24 expression of match and mismatch targets in iVKA.
324 GraphPad Prism (ver. 7.0) was used to perform these tests, and significance is determined as *p*
325 value <0.05.

326

327 **ACKNOWLEDGEMENTS**

328 We acknowledge the International AIDS Vaccine Initiative (IAVI), and the U.S. Agency for
329 International Development (USAID) for the support and funding of the IAVI cohort. We are
330 grateful to Karen Conrod for her contribution to screening for T cell responses to cryptic
331 peptides in the acutely-infected subtype A patients. This work was supported by the RO1 A1
332 084772 (PAG), RO1 A1 112566 (PAG), R56 A1 098551 (PAG), R01 A1 064060 (PAG and
333 EH), the Center for HIV/AIDS Vaccine Immunology grant A1 067854 (AJMcM) and the Grand
334 Challenges in Global Health Program of the Bill and Melinda Gates Foundation grant 37874
335 (PB). Parts of the work were performed in the UAB CFAR BSL-3 facilities and by the UAB
336 CFAR Flow Cytometry Core/Joint UAB Flow Cytometry Core, which are funded by
337 NIH/NIAID P30 AI027767 and by NIH 5P30 AR048311.

338 BJP, PB, NG, AB, EH, PAG, JG, and AJM conceived the study and designed the experiments.
339 JMC, PB, MKPL, NG, and FG produced the peptide predictions. BJP and MKPL performed the
340 experiments. BJP, AB, PAG, PB, MKPL, and NG analyzed the data. BJP, AB, and PAG wrote
341 the manuscript.

342 All authors declare no financial conflict of interest for this study.

343

344 REFERENCES

- 345 1. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, Carlson JM, Derdeyn CA, Tang J,
346 Kaslow RA, Bansal A, Yusim K, Heckerman D, Mulenga J, Allen S, Goulder PJ, Hunter E. 2008.
347 Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in
348 linked recipients. *J Exp Med* 205:1009-17.
- 349 2. Liu J, O'Brien KL, Lynch DM, Simmons NL, La Porte A, Riggs AM, Abbink P, Coffey RT, Grandpre
350 LE, Seaman MS, Landucci G, Forthal DN, Montefiori DC, Carville A, Mansfield KG, Havenga MJ,
351 Pau MG, Goudsmit J, Barouch DH. 2009. Immune control of an SIV challenge by a T-cell-based
352 vaccine in rhesus monkeys. *Nature* 457:87-91.
- 353 3. Wilson NA, Keele BF, Reed JS, Piaskowski SM, MacNair CE, Bett AJ, Liang X, Wang F, Thoryk E,
354 Heidecker GJ, Citron MP, Huang L, Lin J, Vitelli S, Ahn CD, Kaizu M, Maness NJ, Reynolds MR,
355 Friedrich TC, Loffredo JT, Rakasz EG, Erickson S, Allison DB, Piatak M, Jr., Lifson JD, Shiver JW,
356 Casimiro DR, Shaw GM, Hahn BH, Watkins DI. 2009. Vaccine-induced cellular responses control
357 simian immunodeficiency virus replication after heterologous challenge. *J Virol* 83:6508-21.
- 358 4. Bukrinsky MI, Etkin AF. 1990. Plus strand of the HIV provirus DNA is expressed at early stages of
359 infection. *AIDS Res Hum Retroviruses* 6:425-6.
- 360 5. Bullock TN, Patterson AE, Franlin LL, Notidis E, Eisenlohr LC. 1997. Initiation codon scanthrough
361 versus termination codon readthrough demonstrates strong potential for major
362 histocompatibility complex class I-restricted cryptic epitope expression. *J Exp Med* 186:1051-8.
- 363 6. Cardinaud S, Moris A, Fevrier M, Rohrlich PS, Weiss L, Langlade-Demoyen P, Lemonnier FA,
364 Schwartz O, Habel A. 2004. Identification of cryptic MHC I-restricted epitopes encoded by HIV-1
365 alternative reading frames. *J Exp Med* 199:1053-63.
- 366 7. Cardinaud S, Starck SR, Chandra P, Shastri N. 2010. The synthesis of truncated polypeptides for
367 immune surveillance and viral evasion. *PLoS One* 5:e8692.
- 368 8. Dolan BP, Bennink JR, Yewdell JW. 2011. Translating DRiPs: progress in understanding viral and
369 cellular sources of MHC class I peptide ligands. *Cell Mol Life Sci* 68:1481-9.
- 370 9. Dolan BP, Li L, Veltri CA, Ireland CM, Bennink JR, Yewdell JW. 2011. Distinct pathways generate
371 peptides from defective ribosomal products for CD8+ T cell immunosurveillance. *J Immunol*
372 186:2065-72.
- 373 10. Garrison KE, Champiat S, York VA, Agrawal AT, Kallas EG, Martin JN, Hecht FM, Deeks SG, Nixon
374 DF. 2009. Transcriptional errors in human immunodeficiency virus type 1 generate targets for T-
375 cell responses. *Clin Vaccine Immunol* 16:1369-71.
- 376 11. Gaur A, Green WR. 2005. Role of a cytotoxic-T-lymphocyte epitope-defined, alternative gag
377 open reading frame in the pathogenesis of a murine retrovirus-induced immunodeficiency
378 syndrome. *J Virol* 79:4308-15.
- 379 12. Ho O, Green WR. 2006. Cytolytic CD8+ T cells directed against a cryptic epitope derived from a
380 retroviral alternative reading frame confer disease protection. *J Immunol* 176:2470-5.
- 381 13. Ho O, Green WR. 2006. Alternative translational products and cryptic T cell epitopes: expecting
382 the unexpected. *J Immunol* 177:8283-9.
- 383 14. Landry S, Halin M, Lefort S, Audet B, Vaquero C, Mesnard JM, Barbeau B. 2007. Detection,
384 characterization and regulation of antisense transcripts in HIV-1. *Retrovirology* 4:71.
- 385 15. Larocque E, Halin M, Landry S, Marriott SJ, Switzer WM, Barbeau B. 2011. Human T-cell
386 lymphotropic virus type 3 (HTLV-3)- and HTLV-4-derived antisense transcripts encode proteins
387 with similar Tax-inhibiting functions but distinct subcellular localization. *J Virol* 85:12673-85.
- 388 16. Laverdure S, Gross A, Arpin-Andre C, Clerc I, Beaumelle B, Barbeau B, Mesnard JM. 2012. HIV-1
389 antisense transcription is preferentially activated in primary monocyte-derived cells. *J Virol*
390 86:13785-9.

- 391 17. Lefebvre G, Desfarges S, Uyttebroeck F, Munoz M, Beerenwinkel N, Rougemont J, Telenti A,
392 Ciuffi A. 2011. Analysis of HIV-1 expression level and sense of transcription by high-throughput
393 sequencing of the infected cell. *J Virol* 85:6205-11.
- 394 18. Li C, Goudy K, Hirsch M, Asokan A, Fan Y, Alexander J, Sun J, Monahan P, Seiber D, Sidney J, Sette
395 A, Tisch R, Frelinger J, Samulski RJ. 2009. Cellular immune response to cryptic epitopes during
396 therapeutic gene transfer. *Proc Natl Acad Sci U S A* 106:10770-4.
- 397 19. Malarkannan S, Horng T, Shih PP, Schwab S, Shastri N. 1999. Presentation of out-of-frame
398 peptide/MHC class I complexes by a novel translation initiation mechanism. *Immunity* 10:681-
399 90.
- 400 20. Maness NJ, Valentine LE, May GE, Reed J, Piaskowski SM, Soma T, Furlott J, Rakasz EG, Friedrich
401 TC, Price DA, Gostick E, Hughes AL, Sidney J, Sette A, Wilson NA, Watkins DI. 2007. AIDS virus
402 specific CD8+ T lymphocytes against an immunodominant cryptic epitope select for viral escape.
403 *J Exp Med* 204:2505-12.
- 404 21. Maness NJ, Walsh AD, Piaskowski SM, Furlott J, Kolar HL, Bean AT, Wilson NA, Watkins DI. 2010.
405 CD8+ T cell recognition of cryptic epitopes is a ubiquitous feature of AIDS virus infection. *J Virol*
406 84:11569-74.
- 407 22. Miller RH. 1988. Human immunodeficiency virus may encode a novel protein on the genomic
408 DNA plus strand. *Science* 239:1420-2.
- 409 23. Saulquin X, Scotet E, Trautmann L, Peyrat MA, Halary F, Bonneville M, Houssaint E. 2002. +1
410 Frameshifting as a novel mechanism to generate a cryptic cytotoxic T lymphocyte epitope
411 derived from human interleukin 10. *J Exp Med* 195:353-8.
- 412 24. Zook MB, Howard MT, Sinnathamby G, Atkins JF, Eisenlohr LC. 2006. Epitopes derived by
413 incidental translational frameshifting give rise to a protective CTL response. *J Immunol*
414 176:6928-34.
- 415 25. Bansal A, Carlson J, Yan J, Akinsiku OT, Schaefer M, Sabbaj S, Bet A, Levy DN, Heath S, Tang J,
416 Kaslow RA, Walker BD, Ndung'u T, Goulder PJ, Heckerman D, Hunter E, Goepfert PA. 2010. CD8 T
417 cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense
418 transcription. *J Exp Med* 207:51-9.
- 419 26. Bansal A, Mann T, Sterrett S, Peng BJ, Bet A, Carlson JM, Goepfert PA. 2015. Enhanced
420 Recognition of HIV-1 Cryptic Epitopes Restricted by HLA Class I Alleles Associated With a
421 Favorable Clinical Outcome. *J Acquir Immune Defic Syndr* 70:1-8.
- 422 27. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, Henry LM, Rosato PC, Piechocka-
423 Trocha A, Brockman MA, Harrigan PR, Heckerman D, Kaufmann DE, Brander C. 2010. Viral
424 adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting
425 epitopes in HIV frameshift sequences. *J Exp Med* 207:61-75.
- 426 28. Bet A, Sterrett S, Sato A, Bansal A, Goepfert PA. 2014. Characterization of T-cell responses to
427 cryptic epitopes in recipients of a noncodon-optimized HIV-1 vaccine. *J Acquir Immune Defic*
428 *Syndr* 65:142-50.
- 429 29. Champiat S, Raposo RA, Maness NJ, Lehman JL, Purtell SE, Hasenkrug AM, Miller JC, Dean H, Koff
430 WC, Hong MA, Martin JN, Deeks SG, Spotts GE, Pilcher CD, Hecht FM, Kallas EG, Garrison KE,
431 Nixon DF. 2012. Influence of HAART on alternative reading frame immune responses over the
432 course of HIV-1 infection. *PLoS One* 7:e39311.
- 433 30. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, Deymier MJ, Ende ZS,
434 Klatt NR, DeZiel CE, Lin TH, Peng J, Seese AM, Shapiro R, Frater J, Ndung'u T, Tang J, Goepfert P,
435 Gilmour J, Price MA, Kilembe W, Heckerman D, Goulder PJ, Allen TM, Allen S, Hunter E. 2014.
436 HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science*
437 345:1254031.

31. Monaco DC, Dilerenia DA, Fiore-Gartland A, Yu T, Prince JL, Dennis KK, Qin K, Schaefer M, Claiborne DT, Kilembe W, Tang J, Price MA, Farmer P, Gilmour J, Bansal A, Allen S, Goepfert P, Hunter E. 2016. Balance between transmitted HLA preadapted and nonassociated polymorphisms is a major determinant of HIV-1 disease progression. *J Exp Med* 213:2049-63.
32. Carlson JM, Du VY, Pfeifer N, Bansal A, Tan VY, Power K, Brumme CJ, Kreimer A, DeZiel CE, Fusi N, Schaefer M, Brockman MA, Gilmour J, Price MA, Kilembe W, Haubrich R, John M, Mallal S, Shapiro R, Frater J, Harrigan PR, Ndung'u T, Allen S, Heckerman D, Sidney J, Allen TM, Goulder PJ, Brumme ZL, Hunter E, Goepfert PA. 2016. Impact of pre-adapted HIV transmission. *Nat Med* 22:606-13.
33. Du VY, Bansal A, Carlson J, Salazar-Gonzalez JF, Salazar MG, Ladell K, Gras S, Josephs TM, Heath SL, Price DA, Rossjohn J, Hunter E, Goepfert PA. 2016. HIV-1-Specific CD8 T Cells Exhibit Limited Cross-Reactivity during Acute Infection. *J Immunol* 196:3276-86.
34. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW. 2008. The antisense transcriptomes of human cells. *Science* 322:1855-7.
35. Cardinaud S, Consiglieri G, Bouziat R, Urrutia A, Graff-Dubois S, Fourati S, Malet I, Guernon J, Guihot A, Katlama C, Autran B, van Endert P, Lemonnier FA, Appay V, Schwartz O, Kloetzel PM, Moris A. 2011. CTL escape mediated by proteasomal destruction of an HIV-1 cryptic epitope. *PLoS Pathog* 7:e1002049.
36. Walsh AD, Bimber BN, Das A, Piaskowski SM, Rakasz EG, Bean AT, Mudd PA, Ericson AJ, Wilson NA, Hughes AL, O'Connor DH, Maness NJ. 2013. Acute phase CD8+ T lymphocytes against alternate reading frame epitopes select for rapid viral escape during SIV infection. *PLoS One* 8:e61383.
37. Schwab SR, Shugart JA, Horng T, Malarkannan S, Shastri N. 2004. Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* 2:e366.
38. Starck SR, Jiang V, Pavon-Eternod M, Prasad S, McCarthy B, Pan T, Shastri N. 2012. Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* 336:1719-23.
39. Starck SR, Ow Y, Jiang V, Tokuyama M, Rivera M, Qi X, Roberts RW, Shastri N. 2008. A distinct translation initiation mechanism generates cryptic peptides for immune surveillance. *PLoS One* 3:e3460.
40. Starck SR, Shastri N. 2011. Non-conventional sources of peptides presented by MHC class I. *Cell Mol Life Sci* 68:1471-9.
41. Prasad S, Starck SR, Shastri N. 2016. Presentation of Cryptic Peptides by MHC Class I Is Enhanced by Inflammatory Stimuli. *J Immunol* 197:2981-2991.
42. Liu MK, Hawkins N, Ritchie AJ, Ganusov VV, Whale V, Brackenridge S, Li H, Pavlicek JW, Cai F, Rose-Abrahams M, Treurnicht F, Hraber P, Riou C, Gray C, Ferrari G, Tanner R, Ping LH, Anderson JA, Swanstrom R, B CC, Cohen M, Karim SS, Haynes B, Borrow P, Perelson AS, Shaw GM, Hahn BH, Williamson C, Korber BT, Gao F, Self S, McMichael A, Goonetilleke N. 2013. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. *J Clin Invest* 123:380-93.
43. Dong T, Stewart-Jones G, Chen N, Easterbrook P, Xu X, Papagno L, Appay V, Weekes M, Conlon C, Spina C, Little S, Screatton G, van der Merwe A, Richman DD, McMichael AJ, Jones EY, Rowland-Jones SL. 2004. HIV-specific cytotoxic T cells from long-term survivors select a unique T cell receptor. *J Exp Med* 200:1547-57.
44. Akinkunle OT, Bansal A, Sabbaj S, Heath SL, Goepfert PA. 2011. Interleukin-2 production by polyfunctional HIV-1-specific CD8 T cells is associated with enhanced viral suppression. *J Acquir Immune Defic Syndr* 58:132-40.

- 484 45. Liu D, Zuo T, Hora B, Song H, Kong W, Yu X, Goonetilleke N, Bhattacharya T, Perelson AS, Haynes
485 BF, McMichael AJ, Gao F. 2014. Preexisting compensatory amino acids compromise fitness costs
486 of a HIV-1 T cell escape mutation. *Retrovirology* 11:101.
- 487 46. Song H, Hora B, Bhattacharya T, Goonetilleke N, Liu MK, Wiehe K, Li H, Iyer SS, McMichael AJ,
488 Perelson AS, Gao F. 2014. Reversion and T cell escape mutations compensate the fitness loss of
489 a CD8+ T cell escape mutant in their cognate transmitted/founder virus. *PLoS One* 9:e102734.
- 490 47. Harris M, Burns CM, Becker EA, Braasch AT, Gostick E, Johnson RC, Broman KW, Price DA,
491 Friedrich TC, O'Connor SL. 2013. Acute-phase CD8 T cell responses that select for escape
492 variants are needed to control live attenuated simian immunodeficiency virus. *J Virol* 87:9353-
493 64.
- 494 48. Yue L, Pfafferott KJ, Baalwa J, Conrod K, Dong CC, Chui C, Rong R, Claiborne DT, Prince JL, Tang J,
495 Ribeiro RM, Cormier E, Hahn BH, Perelson AS, Shaw GM, Karita E, Gilmour J, Goepfert P,
496 Derdeyn CA, Allen SA, Borrow P, Hunter E. 2015. Transmitted virus fitness and host T cell
497 responses collectively define divergent infection outcomes in two HIV-1 recipients. *PLoS Pathog*
498 11:e1004565.
- 499 49. Bansal A, Yue L, Conway J, Yusim K, Tang J, Kappes J, Kaslow RA, Wilson CM, Goepfert PA. 2007.
500 Immunological control of chronic HIV-1 infection: HLA-mediated immune function and viral
501 evolution in adolescents. *AIDS* 21:2387-97.
- 502 50. Bet A, Maze EA, Bansal A, Sterrett S, Gross A, Graff-Dubois S, Samri A, Guihot A, Katlama C,
503 Theodorou I, Mesnard JM, Moris A, Goepfert PA, Cardinaud S. 2015. The HIV-1 antisense protein
504 (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* 12:15.

FIGURE LEGENDS

Figure 1. T cell responses to predicted cryptic epitope (CE) in chronically HIV infected

individuals. (A) PBMC obtained from thirty-two chronically HIV infected individuals (HIV+) were evaluated for responses to 10 predicted CE using an IFN- γ ELISpot assay. Samples obtained from HIV seronegative (HIV-) donors served as a control group. PBMC were stimulated overnight with CE peptide pools (6-10 peptides/pool) and the number of cytokine producing cells was enumerated and is shown as spot forming units (SFU) per million PBMC. Each dot represents the average magnitude of response for one individual to a single peptide pool. A positive response is defined as ≥ 55 SFU/ 10^6 PBMC (dotted line), and three times greater magnitude than unstimulated control (media wells). Mann-Whitney U test was used to determine significance between the two groups. **(B)** Deconvoluted responses to single peptides in six chronic patients are shown. Controls and criteria for positive responses were identical to those of (A).

Figure 2. Ex-vivo and in-vitro functional responsiveness of T cells responding to cryptic

epitopes. (A) Representative flow cytometry data shown for an individual who tested positive for CE-WL9 in an IFN- γ ELISpot assay. PBMCs were incubated with peptide for 12 hrs before antibody staining for flow cytometry-based analysis. Cell population shown in panel were pre-gated on live CD3⁺ lymphocyte population, and excluded cells expressing CD14, CD19, and CD4. **(B)** WL9 and FY11 specific cells were expanded in vitro by culturing for two weeks with peptide pulsed monocyte stimulation (see methods). Cell populations were pre-gated in the same format as (A), with the panel depicting secretion of IFN γ , TNF α , and perforin in the top, middle, and bottom respectively. **(C)** A CD8⁺ T cell line specific for CE-FY11 was generated using peptide pulsed monocyte stimulation and co-cultured with HLA A*02 matched and mismatched,

NL4.3 infected CD4⁺ T-cell targets (see methods). Percentage of killing was determined by the following equation: $100 - (\% \text{ p24}^+ \text{ staining without any effector} / \% \text{ p24}^+ \text{ at the different E/T ratios})$.

Figure 3. Analysis of cryptic epitope targeting in acutely HIV-1 infected individuals. (A)

Amino acid (AA) sequence of CE-KL10 is shown for two Zambian subtype C infected patients (ZA1 and ZA2) who were predicted to respond to this epitope. Longitudinal changes in AA at the polymorphic site within each CE are shown for each of these individuals. **(B)** IFN- γ ELISpot assays were performed on PBMC cryopreserved at the indicated number of days after the estimated day of infection (dpi) from ZA1 and ZA2, using 10 μ M of CE-KL10 as the antigen. Error bars represent mean magnitude of responses in triplicate wells. A positive response was defined as ≥ 55 SFU/10⁶ PBMC (dotted line), and three times greater magnitude than unstimulated control (media alone). **(C)** Eleven subtype B acutely infected individuals were tested for T cell responses to 10 CE, and 4/11 had one or more positive T cell responses. The magnitude of the response to each of the 10 peptides in these four individuals are shown in SFC/10⁶ PBMC. All samples used here were cryopreserved within three months of the estimated date of infection.

Figure 4. Antigen sensitivity of immunogenic cryptic epitopes. (A) Serial 10-fold dilution of

cryptic peptides were used stimulate USAHI-3 PBMC in an ELISpot assay. Mean with standard deviation of IFN- γ (x-axis) at different concentrations of each peptide (y-axis) is displayed. Dotted line represents limit of positive response. **(B)** Data in panel (A) normalized to show

percent maximal IFN- γ response induced at each peptide concentration. Dotted line shows the half-maximal effective concentration (EC50).

Table 1. Cryptic epitopes predicted from HLA-I associated polymorphisms

ARF ^a	Gene	Predicted HLA-I	Peptide Sequence	PP ^b	Esc/Rev ^c	Conserv ^d
FF2	<i>gag</i>	A74, A3301	PTRGGQGISFR	0.69	3/3	0.81
FF3	<i>gag</i>	A0205, B0801	EMWKRTPNERL	0.64	3/8	0.87
FF3	<i>gag</i>	B39, B4201	FPQTLWQRPL	0.79	6/6	0.97
FF3	<i>gag</i>	A74	NLAFFPQGEAR	0.51	3/3	0.88
RF1	<i>gag</i>	A340201	EPVYIVSK	0.68	1/4	0.94
RF1	<i>gag</i>	A3001	FPRGPAIFGY	0.87	2/4	0.85
RF1	<i>gag</i>	A0201	GLMYHLPL	0.94	0/7	0.97
RF1	<i>gag</i>	C0802	HPCYLFLKV	0.86	1/4	0.80
RF1	<i>gag</i>	B5101	EPVYIVSKGF	0.56	0/4	0.93
RF2	<i>gag</i>	B57	HAFKVLGDMAW	0.66	2/2	0.84
RF2	<i>gag</i>	B58	HAFKVLGDMDW	0.78	4/2	0.78
RF2	<i>gag</i>	C0804	HAVISSNVA	0.77	3/1	0.90
RF2	<i>gag</i>	B4201	LPQLKHLTL	0.56	3/4	0.82
RF2	<i>gag</i>	B1402	WRKFPGPL	0.66	5/2	0.88
RF2	<i>gag</i>	C0401, C17	YCFPPGLNL	0.63	4/6	0.78
RF3	<i>gag</i>	A0202	FLLCHFPLV	0.83	1/7	0.97
RF3	<i>gag</i>	B5703	LLLLPMHSKF	0.80	0/3	0.87
RF3	<i>gag</i>	A2402	SFPLALTFF	0.66	1/6	0.82
RF3	<i>gag</i>	A4301	WLLDVPPLY	0.96	0/3	0.84
RF2	<i>nef</i>	A2402	GYLTPGPGVF	0.59	2/2	0.89
FF2	<i>pol</i>	B5702	IAGLSMIYRSW	0.70	2/1	0.81
FF2	<i>pol</i>	B5301	IPLPNYGTSW	0.91	0/5	0.88
FF2	<i>pol</i>	B5703	ITAIGEQW	0.88	0/6	0.84
FF2	<i>pol</i>	A33	KWPSHKGR	0.73	7/2	0.95
FF2	<i>pol</i>	A0301, A340201, A66	QIRRLNYK	0.68	4/3	0.84
FF2	<i>pol</i>	A2402	QYWMWGMHI	0.99	0/4	0.85
FF2	<i>pol</i>	A0202, A2402, A6801, B1402, B3001	WMWGMHIFQL	0.82	2/17	0.79
FF2	<i>pol</i>	A3201	YIMTHQKTW	0.61	1/12	0.76
FF3	<i>pol</i>	A33	GLPVMGTSTQR	0.66	0/3	0.82
RF1	<i>pol</i>	B1801	CPFVRMEF	0.76	1/5	0.85
RF1	<i>pol</i>	B15	FITASYYFCY	0.62	2/5	0.88
RF1	<i>pol</i>	A0201	FLFPCWVWYSL	0.80	0/4	0.81
RF1	<i>pol</i>	B15	FLHCLCLFY	0.53	5/2	0.94
RF1	<i>pol</i>	A0301, A68	FLMFFVWCGK	0.74	1/5	0.92
RF1	<i>pol</i>	B4201	FPSLRKHIL	0.80	1/3	0.78
RF1	<i>pol</i>	B0801, C18	FYKYLIL	0.63	1/2	0.85
RF1	<i>pol</i>	A2902	GLDFCHATLEY	0.56	1/7	0.85
RF1	<i>pol</i>	C0501	HIQYCHFF	0.83	0/6	0.83
RF1	<i>pol</i>	A2902	LYFCSMLLY	0.86	1/3	0.86
RF1	<i>pol</i>	B1801	NPWCFFVYTRY	0.65	0/7	0.84
RF1	<i>pol</i>	A4301	QIIHILIDDY	0.95	1/2	0.76
RF1	<i>pol</i>	A4301	QLLYFCSMPLPY	0.70	0/3	0.79
RF1	<i>pol</i>	A29	WTTVYLSMY	0.85	1/2	0.87
RF1	<i>pol</i>	C18	YCHLYCCLKVL	0.87	2/3	0.80
RF1	<i>pol</i>	B18	YPQVDPFLFY	0.93	0/10	0.76
RF1	<i>pol</i>	A24	YYYAFHGYF	0.94	1/2	0.84
RF2	<i>pol</i>	A2902	AMGSFSSWY	0.73	0/6	0.76
RF2	<i>pol</i>	B5301	APAMGSFSSW	0.73	1/5	0.78
RF2	<i>pol</i>	B3501	KPSSKGTEKY	0.58	5/1	0.79
RF2	<i>pol</i>	A30, A3002	KSDPTYKSSY	0.73	2/6	0.90
RF2	<i>pol</i>	A0201	SIACNSVFFV	0.54	1/3	0.89
RF2	<i>pol</i>	B3501	YPAFPILVSL	0.52	5/1	0.89
RF3	<i>pol</i>	C18	DYSLLVHVWL	0.68	2/3	0.87
RF3	<i>pol</i>	A0201	ELPKSLEFFYV	0.70	0/4	0.81
RF3	<i>pol</i>	A26	FLLLYHLLLY	0.54	0/4	0.94
RF3	<i>pol</i>	B1801, B4201	FPWLFFAL	0.66	1/5	0.79
RF3	<i>pol</i>	A0101, A29, A2902, A4301	GTILLGEGY	0.63	6/14	0.81
RF3	<i>pol</i>	C0501	HNILHVLVL	0.82	1/3	0.84
RF3	<i>pol</i>	A02, A0201, B15:16	KLLLMLLFFL	0.81	5/8	0.86
RF3	<i>pol</i>	B35	LPRSLEFFY	0.59	0/4	0.86
RF3	<i>pol</i>	B42	RPFPLLIVL	0.71	0/4	0.77
RF3	<i>pol</i>	A0202	SLIFSFLTLL	0.65	0/6	0.76
RF3	<i>pol</i>	A3201, A6801, A6802, B58	STMSPMFLFGW	0.80	5/15	0.76
RF3	<i>pol</i>	A02	TMFPMFLFGWV	0.81	1/4	0.79
RF3	<i>pol</i>	A0202	WLFFALPLL	0.74	0/5	0.92
RF3	<i>pol</i>	B5702	YQPTQESRW	0.52	0/5	0.95

^aFF= Forward reading Frame and RF= Reverse reading Frame^bPP=posterior probability^cesc/rev = escapes or reversions^dconservation= score based on average percentage of occurrence for listed amino acids within HIV subtype C infected cohort

Table 2. Demographics and clinical parameters of chronic and acute clade B HIV infected individuals

^a ID	Gender	Age	^b CD4	^c VL	^d Risk Factor	^e ART
CHI-1	M	54	705	47	IVDU	No
CHI-2	F	31	606	4684	Heterosexual	No
CHI-3	M	51	468	6900	Heterosexual	No
CHI-4	F	45	829	132	heterosexual	No
CHI-5	F	38	772	12700	heterosexual	No
CHI-6	F	46	650	33543	Heterosexual	No
CHI-7	F	63	827	47	heterosexual	No
CHI-8	F	42	717	47	heterosexual	No
CHI-9	M	62	593	5857	Heterosexual	No
CHI-10	M	52	474	36538	MSM	No
CHI-11	M	44	848	561	MSM	No
CHI-12	M	26	454	435	IVDU	No
CHI-13	M	52	1155	47	MSM	No
CHI-14	M	35	522	3860	MSM	No
CHI-15	F	47	667	132	Heterosexual	No
CHI-16	M	34	681	2359	MSM	No
CHI-17	M	43	686	277	Heterosexual	No
CHI-18	M	52	711	112	MSM	No
CHI-19	F	27	641	1260	Heterosexual	No
CHI-20	M	42	632	305	Heterosexual	No
CHI-21	F	22	656	47	Heterosexual	No
CHI-22	F	37	731	324	Heterosexual	No
CHI-23	F	51	550	14100	Heterosexual	No
CHI-24	M	52	567	8424	Heterosexual	No
CHI-25	M	31	289	5258	MSM	No
CHI-26	F	46	357	119	Heterosexual	Yes
CHI-27	F	46	1562	47	Heterosexual	Yes
CHI-28	F	52	445	49	Heterosexual	Yes
CHI-29	F	53	360	187	Heterosexual	Yes
CHI-30	M	42	892	229	MSM	Yes
CHI-31	F	24	619	13371	IVDU	Yes
CHI-32	M	43	240	5934	IVDU	Yes
USAHI-1	M	34	503	2090000	IVDU	No
USAHI-2	F	31	619	492		No
USAHI-3	M	19				Yes
USAHI-4	M	28		<20		Yes
USAHI-5	M	33				Yes
USAHI-6	M	28	868	1083	MSM	Yes
USAHI-7	M	33	534	461		Yes
USAHI-8	M	30	995	64	Heterosexual	Yes
USAHI-9	M	29	1429	683		Yes
USAHI-10	M	49	1129	152	MSM	Yes
USAHI-11	M	50	541	1580	MSM	Yes

Patients with grayed-out boxes tested positive for at least one CE used in the study

^aCHI=chronic infection and USAHI=acute infection

^bAbsolute CD4 count per mm³ blood

^cCopies of HIV RNA detected per mm³ of blood

^dRoute of virus acquisition. MSM-men who have sex with men and IVDU- intravenous drug user

^eActive antiretroviral therapy status

Table 3. Novel immunogenic cryptic epitopes (CE) identified during acute and chronic HIV-1 infection

Name	AA Sequence	HLA-I ^a	Frequency ^b	
			Acute	Chronic
FY11	FLLILYHLLLY	A*26	18%	6%
FL8	FPWLFFAL	B*1801, B*4201	0%	3%
GY10	GTIILLGEGY	A*0101, A*29, A*4301	18%	3%
HL9	HNILHVLVL	Cw*0501	0%	3%
KL10	KLLLMLLFFL	A*02	18%	6%
RL10	RPFPFLIVL	B*42	18%	3%
SW11	STMSPMFLFGW	A*3201, A*6801, A*6802, B*58	9%	3%
TV11	TMFPMFLFGWV	A*02	9%	9%
WL9	WLFFALPLL	A*0202	0%	13%
YW10	YQTPTQESRW	B*5702	9%	3%

All ten epitopes described here are translated from the antisense, reverse frame 3 of the *pol* coding region.

^aEpiPred predicted HLA-I restriction

^bFrequency of US acute (n=11), and chronic (n=32) patient cohort responding to each CE.

Table 4. Amino acid sequence homology of cryptic epitopes compared to HXB2 derived sequences

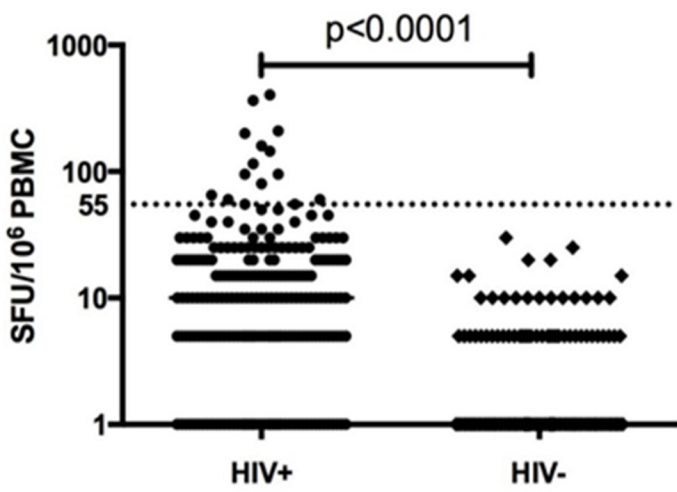
CE sequence	HXB2 sequence^a	HXB2 protein region^b	% homology^c
FLLILYHLLLY	LDLWIYHTQG	Nef	36
FPWLFFAL	FLWMGYEL	Pol	38
GTIILLGEGY	VTVLDVGDAY	Pol	30
HNILHVLVL	HNWATHACV	Env	36
KLLLMLLFFL	RILQQLLFH	Vpr	40
RFPFLLIVL	QPIPIVAIVA	Vpu	40
STMSPMFLFGW	QAISPRTLNA	Gag	36
TMFPMFLFGWV	VRYPL-TFGW	Nef	36
WLFFALPLL	WEFVNTPL	Pol	44
YQTPTQESRW	FKLPIQKETW	Pol/RT	30

^aClosest matching HXB2 AA sequence based on LANL HIV sequence locator

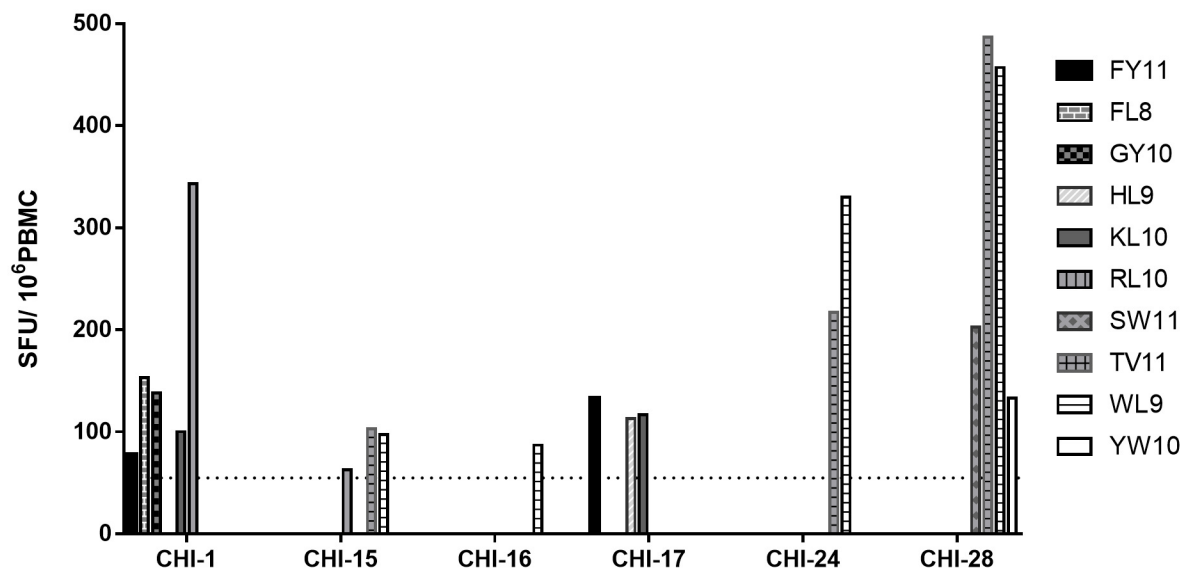
^bHIV-1 protein where the HXB2 sequence is located

^c% identity between CE and HXB2 sequences

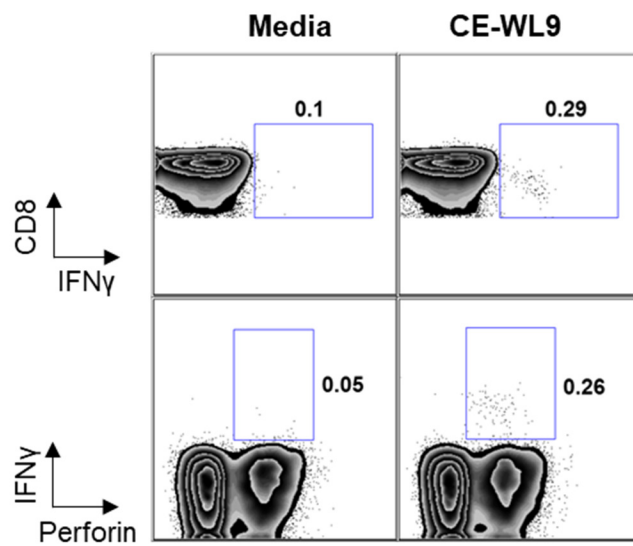
(A)



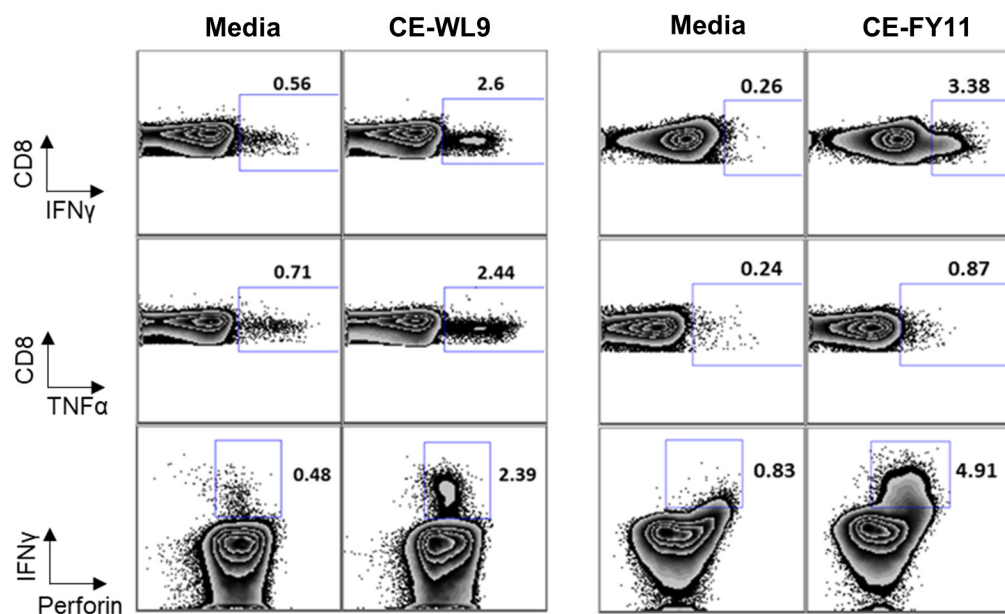
(B)



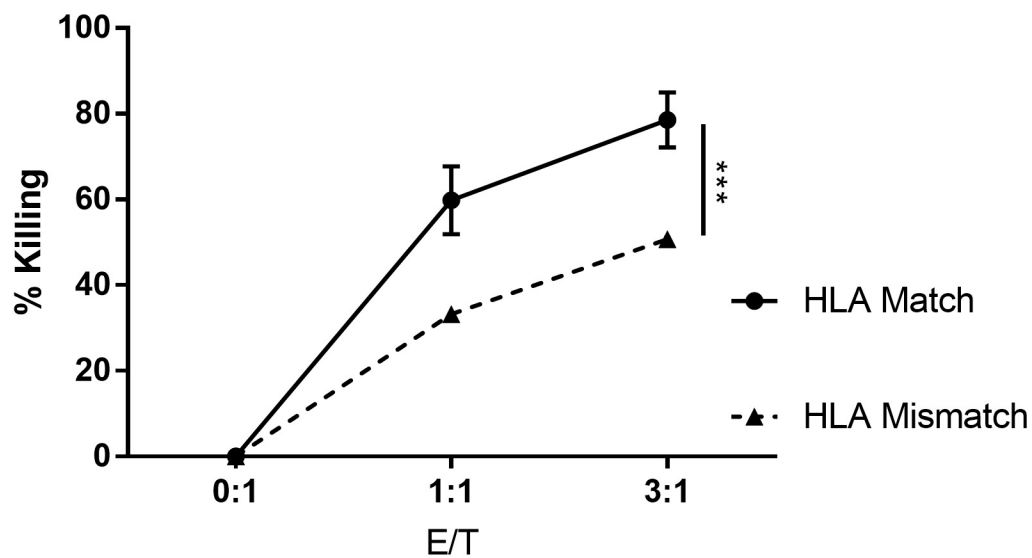
(A)



(B)



(C)



(A)

Patient ZA1		Patient ZA2	
Time	Polymorph sequence	Time	Polymorph sequence
donor	F	donor	L
3 months	F	3 months	P
6 months	F	6 months	P
9 months	F	9 months	-
12 months	-	12 months	P
15 months	F	15 months	P
18 months	F	18 months	P
21months	F	21months	P
24 months	YF	24 months	P

pol - 195

Epitope Region

180

FLRK**KFLLMLLFSL**QLM

Y-----

pol - 195

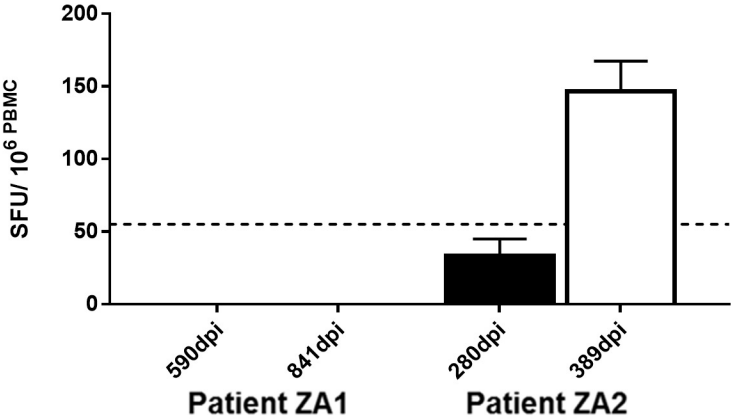
Epitope Region

180

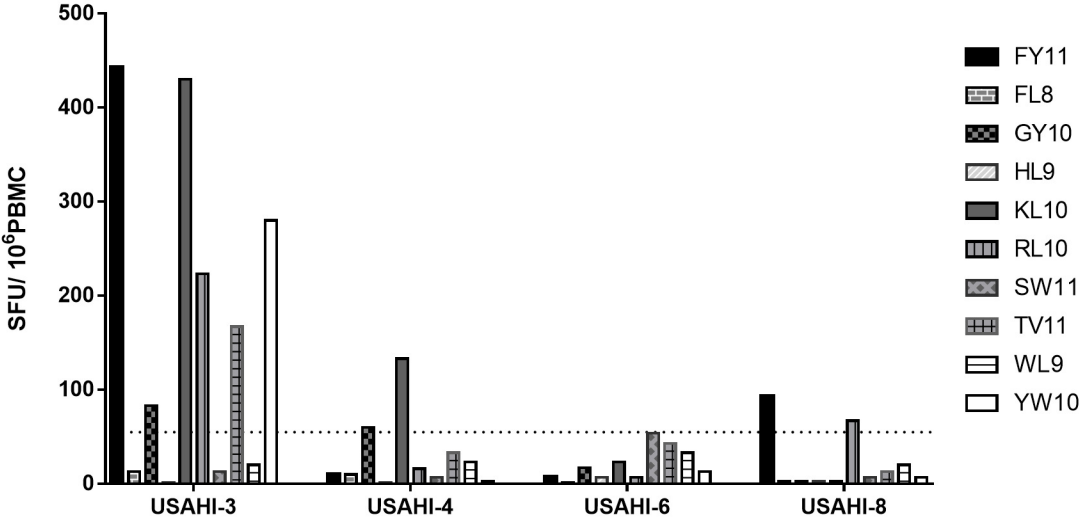
SLH**KLLLMLLFFLL**LM

-P-----

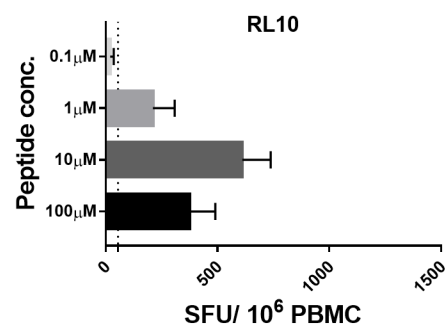
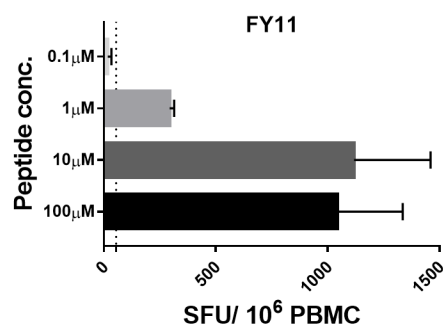
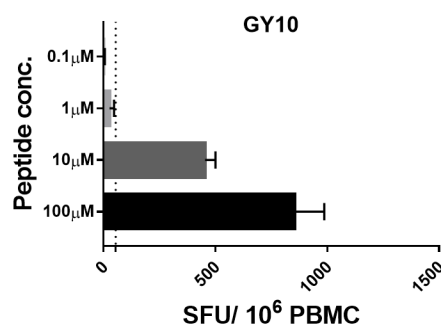
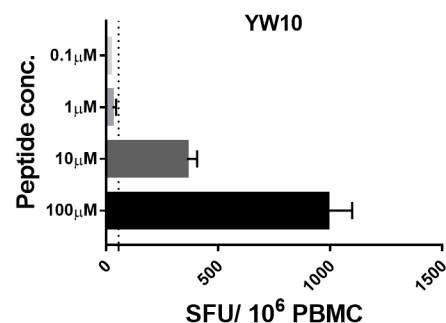
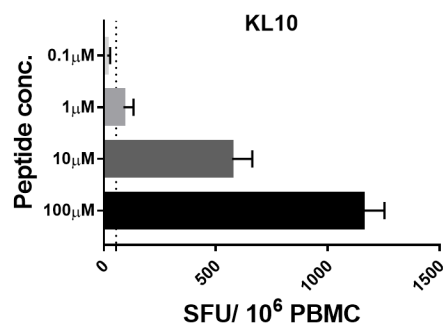
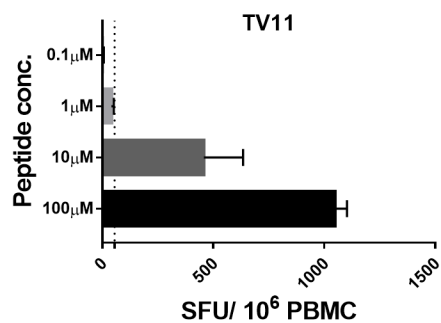
(B)



(C)



(A)



(B)

