Proteome-wide analysis of HIV-specific naive and memory CD4+ T cells in unexposed blood donors

Suzanne L. Campion,[1] Tess M. Brodie,[2] William Fischer,[3] Bette T. Korber,[3] Astrea Rossetti,[2] Nilu Goonetilleke,[4,5] Andrew J. McMichael,[1] and Federica Sallusto[2]

[1]Nuffield Department of Medicine Research Building, University of Oxford, Old Road Campus, Headington, Oxford OX3 7FZ, England, UK
[2]Institute for Research in Biomedicine (IRB), 6-CH-6500 Bellinzona, Switzerland
[3]Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545
[4]Department of Microbiology & Immunology and [5]Department of Medicine, University of North Carolina, Chapel Hill, NC 27599

The preexisting HIV-1–specific T cell repertoire must influence both the immunodominance of T cells after infection and immunogenicity of vaccines. We directly compared two methods for measuring the preexisting CD4+ T cell repertoire in healthy HIV-1–negative volunteers, the HLA-peptide tetramer enrichment and T cell library technique, and show high concordance (r = 0.989). Using the library technique, we examined whether naive, central memory, and/or effector memory CD4+ T cells specific for overlapping peptides spanning the entire HIV-1 proteome were detectable in 10 HLA diverse, HIV-1–unexposed, seronegative donors. HIV-1–specific cells were detected in all donors at a mean of 55 cells/million naive cells and 38.9 and 34.1 cells/million in central and effector memory subsets. Remarkably, peptide mapping showed most epitopes recognized by naive (88%) and memory (56%) CD4+ T cells had been previously reported in natural HIV-1 infection. Furthermore, 83% of epitopes identified in preexisting memory subsets shared epitope length matches (8–12 amino acids) with human microbiome proteins, suggestive of a possible cross-reactive mechanism. These results underline the power of a proteome-wide analysis of peptide recognition by human T cells for the identification of dominant antigens and provide a baseline for optimizing HIV-1–specific helper cell responses by vaccination.

CORRESPONDENCE
Suzanne L. Campion:
Suzanne.Campion@ndm.ox.ac.uk

Abbreviations used: BLAST, basic local alignment sequence tool; gE, glycoprotein E; HMP, human microbiome project; IE63, immediate early phosphoprotein 63; LANL, Los Alamos National Laboratory; TT, tetanus toxoid; VZV, varicella zoster virus.

Only one candidate HIV vaccine, a canarypox vectored gp120 with a protein boost, has shown any efficacy (Rerks-Ngarm et al., 2009). The limited protection correlated with induction of nonneutralizing antibodies to the VI/V2 region of the virus Envelope protein (Env; Rerks-Ngarm et al., 2009; Haynes et al., 2012). This modest success has stimulated efforts to design vaccines that generate more efficient neutralizing antibodies, together with potent CD4+ T cell responses capable of providing help to B cells and cytotoxic T cells (Burton et al., 2012). Understanding how the magnitude and specificity of these helper T cells can be optimized will be critical to the design of an effective vaccine.

Primary immune responses are probably influenced strongly by the preexisting repertoire of B and T cells. However, characterization and quantification of these repertoires is difficult due to the extremely low number of circulating naive precursor cells (Jenkins et al., 2001; Su et al., 2013). Previous studies of naive CD4+ T cell repertoires in humans and mice have relied on magnetic beads to enrich MHC tetramer binding cells (Moon et al., 2007; Kwok et al., 2012; Su et al., 2013). However, although this approach gives precise information on responses to particular MHC-peptide epitopes, it does not measure the total repertoire and misses previously unknown epitopes. An alternative T cell library technique requires no prior knowledge of donor HLA type or epitope specificity (Geiger et al., 2009). The method presorts circulating T cells into naive and memory subsets which are seeded at limiting dilution before polyclonal

A.J. McMichael and F. Sallusto contributed equally to this paper.

The Rockefeller University Press $30.00
J. Exp. Med. 2014
www.jem.org/cgi/doi/10.1084/jem.20130555

Published June 23, 2014

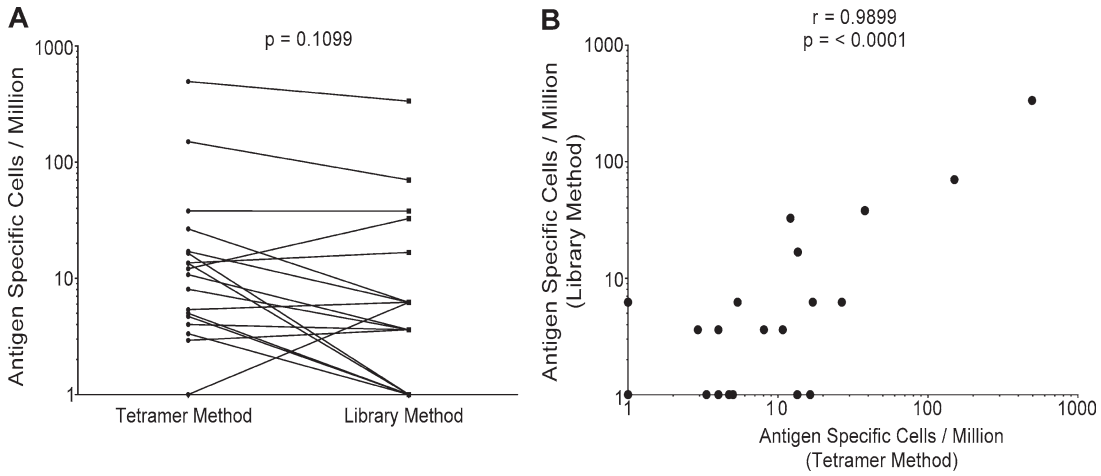


Figure 1. Comparison of the CD4+ T cell library and HLA tetramer enrichment methods. The naive and memory CD4+ T cell subsets of 5 HLADRB15*01 donors were screened in parallel for specificity to two known VZV epitopes (gE and IE63) using the T cell library technique and tetramer enrichment protocol. The precursor frequencies obtained from each methodology were subsequently compared using a paired Student's *t* test (P = 0.1099; A), and Pearson's correlation (r = 0.989; B). Given that the dataset included zero values, a log (x + 1) transformation was applied to all datasets.

expansion in the presence of PHA, allogeneic feeder cells, and IL-2. Individual cultures are then screened for proliferative responses to a protein or series of peptides representing the pathogen of interest (Geiger et al., 2009). Combined with epitope mapping and the Poisson distribution, the T cell library technique can provide quantitative data on the specificity of the entire preexisting naive and memory repertoire.

The existence of HIV-1–specific memory cells in seronegative donors was originally suggested by studies of highly exposed HIV-1 seronegative (HESN) donors. It has been shown that 25–61% of HESNs have demonstrable HIV-1–specific memory cells, probably primed by exposure to the virus. Surprisingly, HIV-1–specific $CD4^+$ T cells were also detected in 24–44% of unexposed donors (Ritchie et al., 2011), although it was not clear whether the latter came from cross-reactive memory T cells or naive T cells primed in vitro. More recently, the existence of low frequency (1–10/million) memory $CD4^+$ T cells, specific for a known HIV-1 Gag epitope, was demonstrated by HLA DR4 tetramers in 50% of HIV-1 unexposed HLA $DR4^+$ adults (Su et al., 2013), but it was not clear how generalizable the HIV-1 result was beyond the single epitope–HLA DR4 combination.

The present study first validates the library technique by direct comparison with the tetramer enrichment method for measuring precursor T cell frequencies. We then use the T cell library technique to provide the first proteome-wide analysis of the frequencies and specificities of preexposure HIV-1–specific naive and memory $CD4^+$ T cells in a HLA diverse population of HIV-1 unexposed donors.

RESULTS AND DISCUSSION

Comparison of the T cell library technique to tetramer enrichment

Before commencing a proteome-wide screen of the preexisting HIV-1–specific naive and memory $CD4^+$ repertories, we first performed a direct comparison of the T cell library (Geiger et al., 2009) and tetramer enrichment methods (Su et al., 2013) using varicella zoster virus (VZV) as a model antigen. We established $CD4^+$ T cell libraries of 192 wells per subset from naive ($CD45RA^+CCR7^+$), central memory ($CD45RA^-CCR7^+$), and effector memory ($CD45RA^-CCR7^-$) subsets of five anonymous $HLADRB1^*1501^+$ blood donors. Lines were polyclonally expanded for 16–20 d before screening for proliferative response to two VZV epitopes (glycoprotein E [gE] and immediate early phosphoprotein 63 [IE63]), which are known to be restricted by and commonly detected in HLADRB1*1501 donors (Jones et al., 2007; Malavige et al., 2008). In parallel 25–70 × 10^6 $CD4^+$ T cells from the same time point were screened using custom-made gE and IE63 HLADRB1*1501 tetramers. We show that the two techniques are highly comparable (Pearson's correlation, r = 0.999), with no significant difference in the precursor frequencies obtained by each technique (paired Student's *t* test, P = 0.1099; Fig. 1). These data support the T cell library technique for comprehensive analysis of naive and memory $CD4^+$ T cell repertoires.

HIV-1–specific T cells in HIV-1–unexposed, seronegative donors

Naive, central memory, and effector memory $CD4^+$ T cell subsets from 10 HLA diverse (Table S1), healthy, HIV-1 seronegative donors were screened for HIV-1–specific responses using the T cell library method (Geiger et al., 2009). As above, 192 cell lines per subset were established at limiting dilution and polyclonally expanded. Each cell line was screened for reactivity to peptides spanning the entire consensus clade C HIV-1 proteome, pooled by proteins Gag, Env, Pol, Nef, and a mix of Vpr, Vpu, Tat, Rev, and Vif (designated Nef/Acc). Responding cells were detected by ^3H thymidine incorporation.

HIV-1–specific $CD4^+$ T cells were detected in the naive subsets of all donors (Fig. 2 A) at frequencies ranging from

Published June 23, 2014

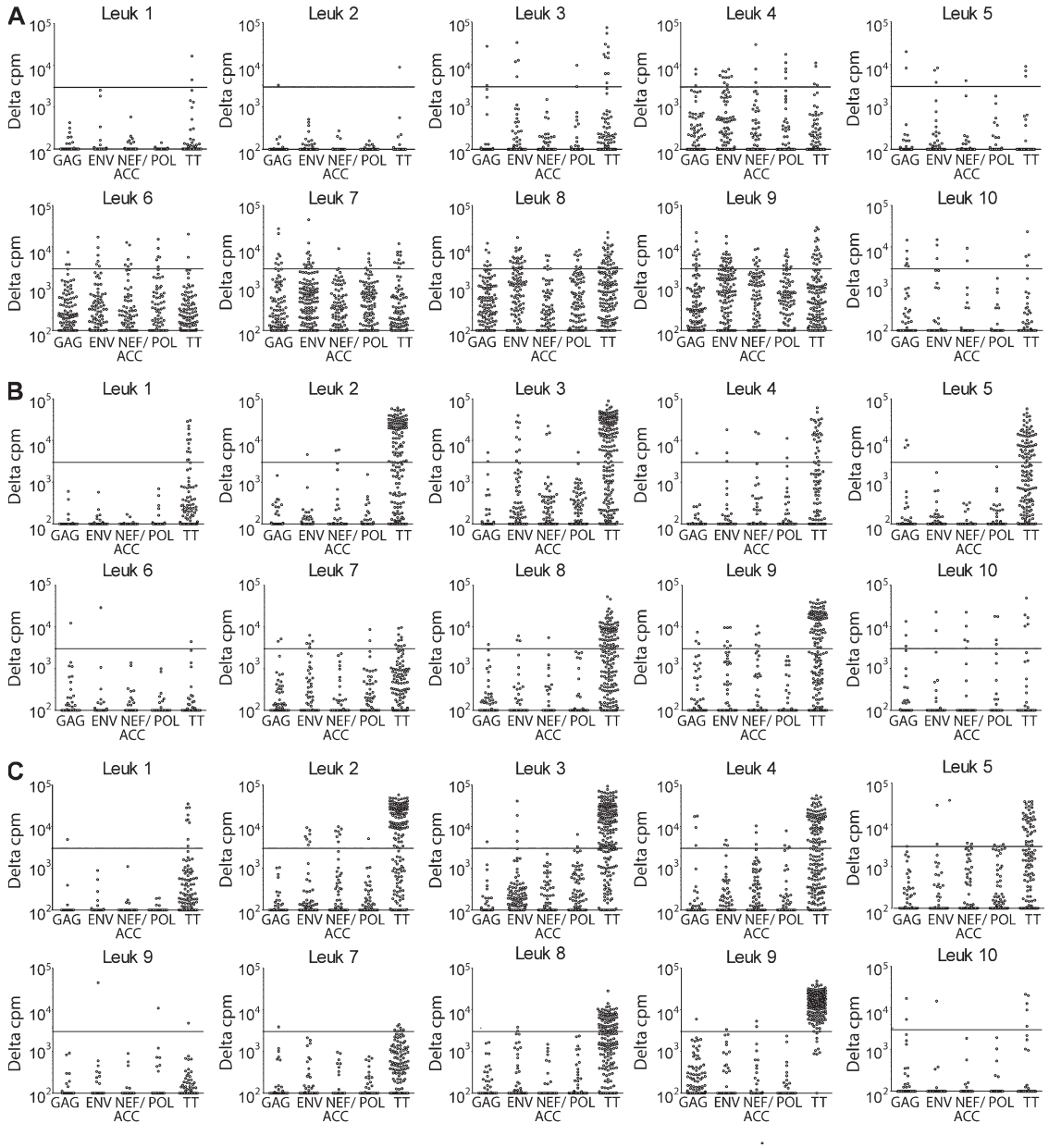
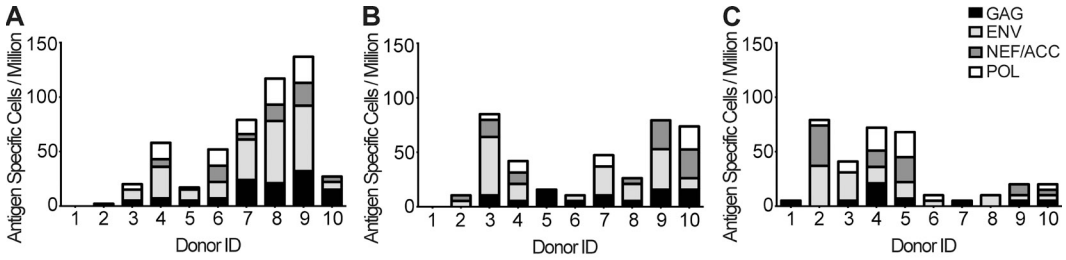


Figure 2. HIV-1–specific responses were detected in the circulating naive and memory CD4+ T cell subsets of healthy, HIV-1 seronegative donors. For each of 10 donors, a mean of 187 cultured cell lines (each represented by a single dot) per naive (A), central memory (B), and effector memory (C) subsets were screened against pools of overlapping peptides spanning the entire HIV-1 proteome. Peptide pools were split according to protein with Nef and Accessory proteins, Vpr, Vpu, Tat, Rev, and Vif included as a single pool referred to as Nef/Acc. All data presented are expressed as the counts per minute, after subtraction of background, nonspecific proliferation (delta cpm). Positive responses are defined as SI >5 and >3,000 delta cpm with cutoff criteria represented by a horizontal line. Cell lines with background counts of ≥3,000 cpm were excluded from analysis. Proliferative responses to known recall antigen TT are shown for each donor.

28 to 129 cells/million (Fig. 3, A and D), comparable to precursor frequencies of tetanus toxoid (TT)–specific naive CD4+ cells (Fig. 3 D). Although the use of a single consensus virus sequence allowed a comprehensive assessment of preexisting specificity, it must underestimate the total response specific for this highly variable virus. However, because 70% of new HIV-1 infections are established by a single founder virus (Keele et al., 2008), the frequencies presented herein should realistically predict an individual's CD4+ T cell response in acute infection.

In addition to the naive repertoire, HIV-1–specific T cells were also detected within the central memory (Fig. 2 B) and/or effector memory (Fig. 2 C) CD4+ T cell compartments of all donors. However, the observed frequencies were 15–20-fold lower (Fig. 3, B–D) than TT-specific memory CD4+ T cells (Fig. 3 D). The high circulating frequencies of TT-specific memory CD4+ T cells are consistent with previous reports (Geiger et al., 2009) and reflect successful prior immunization (Sallusto et al., 2010). In contrast, the low frequencies

Published June 23, 2014



D

	GAG	ENV	NEF/ACC	POL	TT
Naive	12.23	23.09	7.40	9.82	39.09
Central Memory	8.39	16.94	8.94	4.72	638.99
Effector Memory	5.51	12.17	9.24	7.15	877.19

E

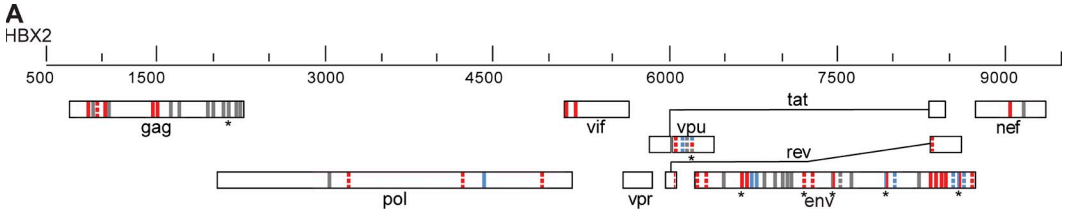
	GAG	ENV	NEF/ACC	POL
Naive	24.4	26.5	9.0	9.8
Central Memory	16.7	19.5	10.9	4.7
Effector Memory	11.0	14.0	11.3	7.1

Figure 3. Precursor frequencies of HIV-1 and TT–specific T cell responses detected within the naive, and memory CD4⁺ T cell subsets. These were calculated for naive (A), central memory (B) and effector memory (C) subsets using the Poisson distribution. The mean number of antigen-specific cells per million was calculated from all 10 donors studied (D) and the dataset normalized according to protein (defined as number of amino acids; E).

of HIV-1–specific memory CD4⁺ T cells in unexposed donors suggest that these T cells arose because of rare cross-reactivities with non–HIV-1 antigens.

Considerable inter-donor variation in both the specificity and frequency of naive and memory HIV-1–specific CD4⁺ T cells was observed (Fig. 3, A–C), probably contributing to the great variation in adaptive immune responses seen after natural infection and vaccination. Overall, the HIV-1–specific memory T cell frequencies detected herein were ∼10-fold higher than those recently reported using tetramer enrichment (Su et al., 2013). However, this is not a real discrepancy because we included peptides spanning the entire HIV-1 proteome, whereas they focused their analysis to a single known epitope; therefore, the findings are fully compatible.

After normalization of the dataset for viral protein size, HIV-1 Gag and Env proteins were the immunodominant targets of the preexposure naive and memory CD4⁺ T cell subsets (Fig. 3 E), similar to the known patterns of immunodominance in natural HIV-1 infection (Kaufmann et al., 2004; Ranasinghe et al., 2012). Consistent with the low frequency of Pol-specific responses observed after infection (Kaufmann et al., 2004; Ranasinghe et al., 2012), relatively



B

	Naive	Central Memory	Effector Memory
GAG	12/12	4/5	1/1
POL	1/1	0/3	1/1
ENV	10/11	10/16	4/7
NEF	1/1	1/1	0/0
VPU	0/2	0/2	0/2
REV	0/0	0/1	0/0
VIF	0/0	2/2	0/0
Total	**24/27**	**17/30**	**6/11**

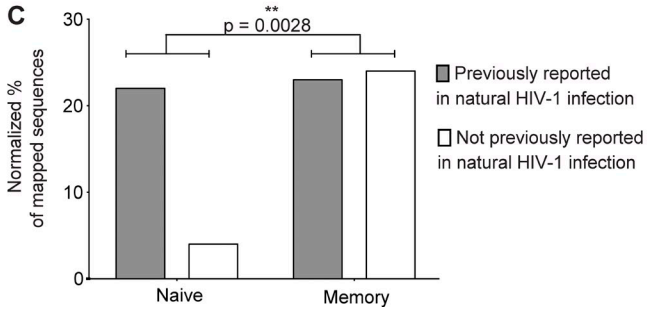
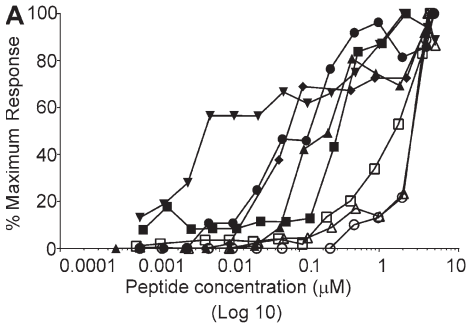
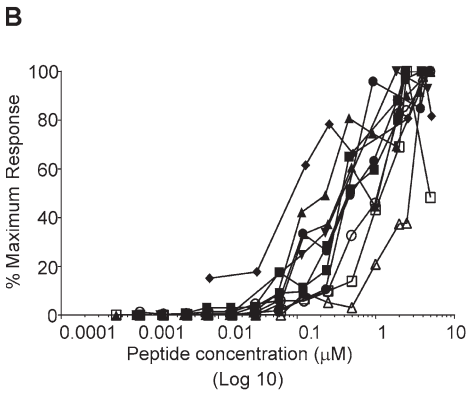


Figure 4. Pre-immune and post-infection CD4⁺ T cell responses frequently respond to the same peptide epitopes. The specificity of 68 cell lines from the naive (gray line), central memory (red line), and effector memory (blue line) CD4⁺ T cell subsets of 5 HIV-1 unexposed seronegative donors was determined using epitope mapping (A). Epitopes detected in two or more donors are represented with an *. The LANL database was screened to determine whether mapped epitopes had previously detected in natural HIV-1 infection. Epitopes that had previously been detected in HIV-1 infection are represented with a solid line, whereas those that had not previously been characterized are shown with a dashed line. The number of mapped epitopes that had previously been reported in natural HIV-1 infection as compared with the total number of mapped epitopes for each protein is presented in B. To minimize bias, the number of amino acids covered by epitopes identified in this study was normalized to the total number of amino acids in HIV-1 reference strain HBX2 covered by previously identified or as yet unknown CD4⁺ T cell epitopes and compared between memory and naive subsets using a Fisher's exact test (C). Significance (**) was defined as P = 0.0028.

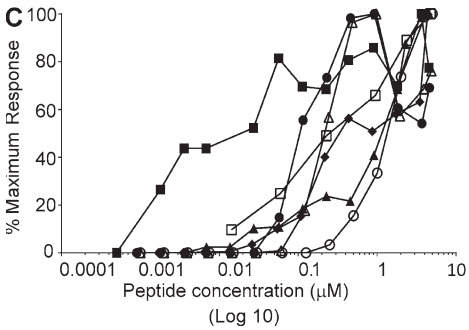
Published June 23, 2014



Donor ID	Protein	Region	HBX2 Location	Max Δcpm	EC50 (μM)
Leuk 7	GAG	p17	910 > 963	23829	0.0152
Leuk 7	GAG	p24	1663 > 1716	5838	0.0545
Leuk 6	GAG	p1	2074 > 2127	3685	2.72
Leuk 7	ENV	gp120	6472 > 6545	3007.5	1.259
Leuk 7	ENV	gp120	6861 > 6914	3805	2.14
Leuk 10	ENV	gp120	7029 > 7082	21681	0.0737
Leuk 7	ENV	gp120	7527 > 7580	25200	0.775
				Mean	1.006



Donor ID	Protein	Region	HBX2 Location	Max Δcpm	EC50 (μM)
Leuk 7	GAG	p17	886 > 939	3028	0.3815
Leuk 7	GAG	p17	958 > 1011	4587	0.3272
Leuk 9	GAG	p24	1471 > 1524	25200	0.0963
Leuk 7	POL	Integrase	4917 > 4970	7725	1.242
Leuk 9	ENV	gp120	6666 > 6719	8191	0.927
Leuk 9	ENV	gp120	6666 > 6719	9921	3.125
Leuk 9	ENV	gp120	7224 > 7277	13058	0.1701
Leuk 8	ENV	gp41	8592 > 8645	4747	1.09
Leuk 9	REV	REV	6018 > 6045 8379 > 8404	12988	0.5218
Leuk 10	REV	REV	8447 > 8500	6794	0.268
Leuk 9	VPU	VPU	6182 > 6235	8276	0.1524
				Mean	0.755



Donor ID	Protein	Region	HBX2 Location	Max Δcpm	EC50 (μM)
Leuk 7	GAG	p24	1567 > 1620	6401	0.1482
Leuk 6	POL	Integrase	4413 > 4466	6379.5	0.6726
Leuk 9	ENV	gp41	7917 > 7970	10015	0.8469
Leuk 6	ENV	gp41	8592 > 8645	9500	0.0829
Leuk 6	ENV	gp41	8616 > 8669	10796	0.1689
Leuk 9	VPU	VPU	6110 > 6193	3965	0.0016
Leuk 9	VPU	VPU	6182 > 6235	4012.5	1.7490
				Mean	0.524

Figure 5. The avidity of preexisting HIV-1–specific T cell responses detected in the naive and memory CD4+ T cell subsets of 5 healthy HIV-1 seronegative donors. Responses in naive (A), central memory (B), and effector memory (C) CD4+ T cells were assessed using peptide titration. EC_{50} values were determined for each epitope using interpolated dose–response curves and are presented for each cell line in the adjacent table, with epitopes defined according to their position within the HIV-1 reference strain HBX2. All data presented are expressed as the percentage of maximal counts per minute, after subtraction of background, nonspecific proliferation (delta cpm). The maximal delta cpm for each cell line is shown in the accompanying table.

few naive or memory CD4+ T cells specific for Pol were detected in the preexposure repertoire. Low expression of Pol protein in infected cells may account for the lack of Pol-specific responses after infection but cannot explain why these CD4+ T cells are rare in the naive repertoire. Because Pol is a relatively conserved protein, it is possible that T cells with this specificity are more likely to be deleted during thymic selection due to cross-reactivities with endogenous retrovirus sequences.

Peptide specificity and avidity of preexisting HIV-1–specific T cells

We next mapped the fine specificity of 68 preexisting naive, central memory, and effector memory HIV-1–specific CD4+ T cells in a total 68 T cell lines (Fig. 4 and Table S2). Epitopes were found to span the entire HIV-1 proteome (Fig. 4) with the majority of positive cell lines (76%) showing a single, unique peptide specificity, consistent with previous observations that preexisting memory CD4+ T cell responses are largely monoclonal (Geiger et al., 2009; Su et al., 2013).

The functional avidity of a subset of naive ($n = 7$; Fig. 5 A), central memory ($n = 11$; Fig. 5 B), and effector memory ($n = 7$; Fig. 5 C) CD4+ T cell lines was assessed by peptide titration. Overall, the avidities were comparable to those found in T cells responding to natural HIV-1 infection, with both high and low avidity responses (EC_{50} 3.125–0.001577 μM) detected, but no significant difference between subsets.

Published June 23, 2014

The immunogens that primed the preexisting HIV-1–specific memory CD4+ T cells detected within seronegative donors are unknown and likely to come from many sources. Sequence identity searches (Altschul et al., 1990; Edgar, 2010) performed using the reactive HIV-1 peptide sequences identified potential epitope-length (8–12 aa) subsequence matches to a variety of human (Tables S4 B and S5 B) and human microbiome proteins (Tables S4 and S5). Epitope matches ranged from 1 to 230 microbial sequences, with 83% of all HIV-1 epitopes mapped to the CD4+ memory subsets shown to have epitope length (8–12 aa) matches to human microbiome proteins (Tables S4 A and S5 A). These data suggest that microbial proteins could have contributed to T cell priming. Our list of potential cross-reactive epitopes (Tables S4 and S5) is unlikely to be exhaustive because the degree of sequence identity required to activate cross-reactive T cell responses is unpredictable, and highly divergent epitopes can elicit cross-reactive responses in mice (Birnbaum et al., 2014). Moreover, database searches are limited to sequenced organisms and gene prediction algorithms which may miss cryptic epitopes that could be processed and presented during the course of infection (Ho and Green, 2006).

Known immunoprevalent epitopes are strongly represented in the preexposure repertoire

We next compared the mapped epitopes identified in the preexisting repertoire against HIV-1 epitope data stored in the Los Alamos National Laboratory (LANL) database (LANL-Immunology-Database, http://www.hiv.lanl.gov/content/immunology). We found that 70% of epitopes detected had previously been reported in natural HIV-1 infection (Fig. 4 A and Table S3). Many of these epitopes are presented by multiple Class II HLA types, making them immunoprevalent across the population (Table S3). Indeed, 10% of the mapped epitopes detected within this study were recognized by more than one donor (Fig. 4 and Table S2), independent of HLA type (Table S1). These data imply that T cell precursor frequency and specificity in both the naive and memory subsets before infection could play a large part in determining what is immunodominant after infection.

Analysis of the mapped HIV-1 epitopes according to CD4+ T cell subset (Fig. 4, B and C) showed that independent of donor HLA or antigen sensitivity, 88% of the epitopes recognized by preexisting naive CD4+ T cells had previously been detected in natural infection, whereas a significantly lower proportion (56%) was found for epitopes recognized by memory CD4+ T cells ($P = 0.0028$). HIV-1 preferentially infects HIV-1–specific memory CD4+ T cells (Douek et al., 2002), and central memory CD4+ T cells in particular are rapidly depleted during acute infection leaving very few during chronic infection (Younes et al., 2003). Thus, preexisting memory HIV-1–specific T cells could be depleted during the acute stages of infection, allowing naive T cells to expand preferentially. Alternatively, the higher proportion of novel epitopes (42%) identified in the preexisting memory CD4+ T cell repertoire may help to promote diversity of the T cell response in natural infection and could be of clinical benefit to host (Rosenberg et al., 1997).

Detection of preexisting, cross-reactive memory CD4+ T cells in unexposed donors is not restricted to HIV-1

Finally, we asked whether detection of cross-reactive memory CD4+ T cells in unexposed donors is unique to HIV-1. We screened five of the same healthy donors for proliferative responses to peptides spanning the entire envelope protein of the Zaire reference strain of Ebola virus. Because Ebola infection is associated with an extremely high mortality rate (WHO, 1978), our donors cannot have been previously exposed. In addition to naive Ebola Env-specific cells (mean 25.26 specific cells/million; Fig. S1 A), Ebola-specific memory CD4+ T cells were detected in 4/5 donors tested (mean 142 specific cells/million; Fig. S1, B and C). The frequencies of Ebola Env-specific T cells were modestly higher than those observed for HIV-1 (Fig. S1 D) but still 10-fold lower in the memory subsets than TT-specific memory cells (Fig. S1 D and Fig. 3 B). These observations demonstrate that the findings on preexposure CD4+ repertoires reported herein are not limited to HIV-1.

Conclusion

The present dataset provides the first comprehensive, systematic, and quantitative analysis of the preexposure HIV-1–specific CD4+ T cell repertoire in HLA diverse seronegative donors. Using peptides spanning the entire HIV-1 proteome, we show that both specificity and avidity of the preexposure HIV-1–specific CD4+ T cell repertoire has considerable overlap with those of CD4+ T cells detected after natural HIV-1 infection. Furthermore, we suggest that some preexisting memory HIV-1–specific T cells may have been primed by microbial organisms present within the human microbiome. These data help explain immunodominance and immunoprevalence in natural HIV-1 infection and the variability in human immune responses to infection and vaccines.

MATERIALS AND METHODS

Study participants and approval. Leukapheresis samples were obtained from a total of 15 anonymous HIV-1 seronegative, healthy individuals recruited by the Basel Swiss Red Cross Blood Centre and National Blood Service (Bristol, UK). Informed, written consent was obtained from all donors and human primary cell protocols were approved by the Federal Office of Public Health (N. A000197/2 to F. Sallusto).

HLA typing. DNA was extracted using 5 PRIME Achieve Pure DNA kit (Prima Scientific) as per manufacturer's recommendations. HLA typing (Weatherall Institute of Molecular Medicine, Oxford, England, UK) was performed using the sequence-specific primer method adapted from Bunce (2003), which uses allele-specific primer combination in PCR amplification to provide absolute HLA resolution to two digits and high-probability resolution to four digits. HLA types for all donors are shown in Table S1.

Antigen preparation. Synthetic peptides were synthesized by Sigma-Aldrich, and/or the Medical Research Council Human Immunology Unit, WIMM (Oxford, UK), as 18mers overlapping by 10 aa. The Zaire Ebola Reference Strain (Zaire-strain Mayinga-76; FASTA ID VGP_EBOZM, UNIPROT Q05320; UniProt Consortium, 2012) was used to design peptides

Published June 23, 2014

($n = 82$) spanning the envelope glycoprotein of Ebola. HIV-1 peptides spanned the entire HIV-1 proteome and were designed according to the 2004 consensus clade C HIV-1 sequence (LANL-Sequence-Database, http://www.hiv.lanl.gov/). In total, 386 HIV-1–specific peptides were synthesized and pooled according to proteins Gag, Env, and Pol, with a single pool comprising Nef and Accessory proteins Vif, Tat, Vpr, and Vpu (referred to as NEF/ACC). In addition, known VZV epitopes gE (531–545 aa; TSPLLRYAAWTGGLA) and IE63 (229–243 aa; QRAIERYAGAETAEY) were synthesized as 15-aa peptides and complexed to DRB1*1501 tetramers conjugated to PE (NIH Core Tetramer Facility). Quality control checks ensured that the peptide of interest was one of the three major peaks by mass spectrometry and all peptide stocks were sterility tested before use. All peptides were used at a final concentration of 2 µg/ml, and the DMSO concentration in either peptide pools or negative control wells did not exceed 0.45% DMSO. Whole protein TT (provided by G. Galli, Novartis Vaccines, Siena, Italy) was used as a positive control at a final concentration of 5 µg/ml.

Tetramer enrichment protocol. Untouched $CD4^+$ T cells were isolated from PBMCs using magnetic microbeads (Miltenyi Biotec). Tetramer staining and enrichment was performed as described previously (Su et al., 2013). In brief, cells were incubated for 30 min with live/dead Aqua marker (Invitrogen), washed, and then labeled with either gE or IE63 tetramers at room temperature for 45 min (14 µg/ml). Surface markers AF700-labeled anti-CD3 (UCHT3; BD), FITC-labeled anti-CD4 (SK3; BD), Pacific blue–labeled anti-CD45RA (MHCD45RA28; BD), and PE-cyanine 7 (PECy7)–labeled anti-CD56 (B159; BD), anti-CD14 (M5E2; BD), and anti-CD8 (SK1; BD) were incubated at room temperature for 15 min. Before tetramer enrichment, 1/10th staining volume was removed and added to TruCount tubes (BD) to give an absolute count of the starting number of $CD4^+$ naive and memory T cells. The remaining staining volume was enriched for tetramer-positive cells using anti-PE microbeads (Miltenyi Biotec) and added to a separate TruCount tube. Samples were acquired using an LSR Fortessa (BD), and the frequency of tetramer-positive cells determined by dividing the absolute counts of tetramer positive cells by the starting number of $CD4^+$ naive/memory T cells.

Cell purification and sorting for T cell library. $CD14^+$ monocytes and $CD4^+$ T cells were isolated from PBMCs by positive selection with antibody-coated microbeads (Miltenyi Biotec). $CD14^+$ monocytes were immediately cryopreserved and stored in liquid nitrogen until required for use as antigen-presenting cells in subsequent stimulation assays. $CD4^+$ T cell subsets were cell sorted to 99% purity on a FACSAria (BD) after staining with FITC-labeled anti-CD45RA (ALB11; Beckman Coulter), allophycocyanin-labeled anti-CD4 (SK3; BD), and anti-CCR7 (150503; R&D Systems), followed by staining with biotinylated anti-IgG2a (SouthernBiotech) and streptavidin–Pacific blue (Invitrogen). PE-cyanine 5 (PC5)–labeled anti-CD56 (N901 [NHK-1]; Beckman Coulter), anti-CD25 (B1.49.9; Beckman Coulter), and anti-CD8 (B9.11; Beckman Coulter) were included as a dump channel to exclude natural killer, regulatory, and $CD8^+$ T cells.

T cell libraries. The medium used throughout was RPMI 1640 supplemented with 2 mM glutamine, 1% (vol/vol) nonessential amino acids, 1% (vol/vol) sodium pyruvate, 50 U/ml penicillin, 50 µg/ml streptomycin, and 5% human serum (Swiss Red Cross). Cell-sorted naive ($CD45RA^+CCR7^+$), central memory ($CD45RA^+CCR7^+$), and effector memory ($CD45RA^-CCR7^-$) $CD4^+$ T cell populations were seeded at 2,000 (naive) and 1,000 (memory) cells. Cell lines were polyclonally expanded in the presence of 2.5×10^4 irradiated (45Gy) allogeneic feeder cells, 500 IU/ml IL-2, and 1 µg/ml PHA (Thermo Fisher Scientific). Throughout the culture period, cells were supplemented with exogenous complete media containing 500 IU/ml IL-2 and progressively transferred from 96- to 24-well tissue culture plates. The full protocol has previously been published by Geiger et al. (2009).

Stimulation assays. After a 16–20-d culture period, T cells lines were counted and 2.5×10^5 T cells/line removed. T cells were washed 4× in 180 µl PBS and rested for a minimum of 4 h at 37°C 5% CO_2. Autologous $CD14^+$ monocytes (2.5×10^4/well) were pulsed for 2 h with appropriate peptide pools, or control antigen before co-culture with T cells. Proliferation was measured on day 4 after 16 h incubation with 1 µCi/ml [^{3}H] Thymidine (GE Healthcare). A stringent positivity criteria, defined as a stimulation index of >5 and a delta value (cpm in response to antigen-pulsed monocytes − cpm in response to unpulsed monocytes) of $>3 \times 10^3$ cpm was adopted, based upon observations made across multiple negative and positive samples assessed by T cell library technique and represented the 99th percentile of delta cpm obtained from unstimulated samples (Geiger et al., 2009). Overall, <3% of lines were excluded because of nonspecific proliferation. Positive cultures identified in the first screening assay were subsequently reanalyzed using a three dimensional matrix mapping approach (Roederer and Koup, 2003) to identify epitope specificity. Precursor frequencies were calculated based on the number of negative wells according to the Poisson distribution and expressed per million cells (Lefkovits and Waldmann, 1979). EC_{50} values were determined from interpolated dose–response curves using Prism (version 5.00 for Windows; GraphPad Software).

Epitope analysis. All mapped epitopes were screened against the LANL HIV-1 Molecular Immunology database (LANL-Immunology-Database) to determine whether they had previously been reported in natural HIV-1. HIV-1–specific epitopes mapped in the memory subsets were screened using the National Centre for Biotechnology and Information (NCBI) basic local alignment sequence tool (BLAST) search tool (Altschul et al., 1990) for epitope-length subsequence matches to human proteome and microbiome sequences. Short-sequence optimizations were used for BLAST, as the goal was detection of sequence similarity rather than of bona fide homology. Additional human proteome sequence data were downloaded from UniProt (UniProt Consortium, 2012), whereas further human microbiome sequences (shotgun sequences derived from body location samples) were obtained from the Human Microbiome Project (HMP) clustered gene indices catalog (http://www.hmpdacc.org/HMGC/). The UniProt and HMP sequences were processed with UBLAST (Edgar, 2010) using highly non-stringent criteria (id, 0.01; evalue, 10,000; maxaccepts, 500,000) to generate matches. Both NCBI-BLAST and UBLAST matches were then filtered on the number of amino acid identities and similarities (Tables S4 and S5). For NCBI, BLAST-included epitopes showed a minimum of 8 out of 9 matched amino acids with the highest sequence identities showing 12 of 12 matched amino acids. Sequences identified with UBLAST search algorithms had a minimum of 8 of 8 matched amino acids with the highest levels of sequence identity observed with 12 of 12 amino acids matched.

Statistics. Epitope maps obtained from the LANL immunology database show 60% of the HIV-1 proteome is covered by previously defined $CD4^+$ T cell epitopes (LANL-Immunology-Database). Because of this greater proportion of sequence with previously known epitopes, we normalized the number of amino acids covered by epitopes identified in this study to the number of amino acids from the HBX2 reference strain of HIV-1 which contain known $CD4^+$ T cell epitopes (1,875 aa) or which contained no epitopes (1,272 aa). Using a Fisher's exact test, the percentage of HIV-1 proteome covered by previously reported epitopes from the naive and memory $CD4^+$ T cell subsets could be compared. Significance was defined as $P \leq 0.005$.

Pearson's correlation was used to compare the precursor frequencies obtained using the T cell library technique and the tetramer enrichment protocol. Because the dataset included zero values, a log (x + 1) transformation was applied to all data points.

Online supplemental material. Fig. S1 shows data from the naive, central and effector memory subsets of five healthy leukapheresis donors who were screened for proliferative responses to a pool of peptides spanning the envelope protein of the Zaire reference strain of Ebola virus. Table S1 shows the HLA class I and II typing for all 10 donors studied within the context of this manuscript. Table S2 shows the amino acid sequence and delta cpm values for the 68 HIV-1–specific T cell responses detected in the naive, central memory, and effector memory $CD4^+$ T cell subsets of five healthy, HIV-1

seronegative donors. Table S3 shows the results of a Los Alamos database search to determine whether any of the mapped epitopes detected in the preexposure repertoire had previously been reported in natural infection, and lists the citation for instances where the epitope had previously been reported. Table S4 displays the sequence identity matches of HIV-1 epitopes detected in the memory $CD4^+$ T cell repertoire to the human microbiome, human sequences, and all GenBank/EMBL sequences excepting those classified as lentiviral, synthetic, or other. Table S5 shows the results of screening mapped HIV-1–specific epitopes in the memory $CD4^+$ T cell subsets of HIV-1 unexposed uninfected donors against predicted genes from the HMP, shotgun sequences and human sequences from Uniprot. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20130555/DC1.

The authors wish to thank Mrs. V.E. Whale and Miss Elena Brenna for technical assistance, Dr. D. Jarrossay for cell sorting, Professor G.S. Ogg for advice and guidance on the work with varicella zoster–specific epitopes, and MHC class II tetramers, and Dr. T. Rostron for HLA typing. In addition, we acknowledge the NIH Tetramer Core Facility (contract HHSN272201300006C) for provision of GE- and IE63-specific HLA Class II tetramers.

Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health, and by the Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, grant number UM1-AI100645-01, the Medical Research Council, the ERC, grant number ERC-2012-ADG-2012314, and the SNSF, grant number CRSII3_147662. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors declare no competing financial interests.

Submitted: 16 March 2013
Accepted: 29 May 2014

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2

Birnbaum, M.E., J.L. Mendoza, D.K. Sethi, S. Dong, J. Glanville, J. Dobbins, E. Ozkan, M.M. Davis, K.W. Wucherpfennig, and K.C. Garcia. 2014. Deconstructing the peptide-MHC specificity of T cell recognition. *Cell.* 157:1073–1087. http://dx.doi.org/10.1016/j.cell.2014.03.047

Bunce, M. 2003. PCR-sequence-specific primer typing of HLA class I and class II alleles. *Methods Mol. Biol.* 210:143–171.

Burton, D.R., R. Ahmed, D.H. Barouch, S.T. Butera, S. Crotty, A. Godzik, D.E. Kaufmann, M.J. McElrath, M.C. Nussenzweig, B. Pulendran, et al. 2012. A blueprint for HIV vaccine discovery. *Cell Host Microbe.* 12:396–407. http://dx.doi.org/10.1016/j.chom.2012.09.008

Douek, D.C., J.M. Brenchley, M.R. Betts, D.R. Ambrozak, B.J. Hill, Y. Okamoto, J.P. Casazza, J. Kuruppu, K. Kunstman, S. Wolinsky, et al. 2002. HIV preferentially infects HIV-specific $CD4^+$ T cells. *Nature.* 417:95–98. http://dx.doi.org/10.1038/417095a

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 26:2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461

Geiger, R., T. Duhen, A. Lanzavecchia, and F. Sallusto. 2009. Human naive and memory $CD4^+$ T cell repertoires specific for naturally processed antigens analyzed using libraries of amplified T cells. *J. Exp. Med.* 206:1525–1534. http://dx.doi.org/10.1084/jem.20090504

Haynes, B.F., P.B. Gilbert, M.J. McElrath, S. Zolla-Pazner, G.D. Tomaras, S.M. Alam, D.T. Evans, D.C. Montefiori, C. Karnasuta, R. Sutthent, et al. 2012. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.* 366:1275–1286. http://dx.doi.org/10.1056/NEJMoa1113425

Ho, O., and W.R. Green. 2006. Alternative translational products and cryptic T cell epitopes: expecting the unexpected. *J. Immunol.* 177:8283–8289. http://dx.doi.org/10.4049/jimmunol.177.12.8283

Jenkins, M.K., A. Khoruts, E. Ingulli, D.L. Mueller, S.J. McSorley, R.L. Reinhardt, A. Itano, and K.A. Pape. 2001. In vivo activation of antigen-specific CD4 T cells. *Annu. Rev. Immunol.* 19:23–45. http://dx.doi.org/10.1146/annurev.immunol.19.1.23

Jones, L., A.P. Black, G.N. Malavige, and G.S. Ogg. 2007. Phenotypic analysis of human $CD4^+$ T cells specific for immediate-early 63 protein of varicella-zoster virus. *Eur. J. Immunol.* 37:3393–3403. http://dx.doi.org/10.1002/eji.200737648

Kaufmann, D.E., P.M. Bailey, J. Sidney, B. Wagner, P.J. Norris, M.N. Johnston, L.A. Cosimi, M.M. Addo, M. Lichterfeld, M. Altfeld, et al. 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. *J. Virol.* 78:4463–4477. http://dx.doi.org/10.1128/JVI.78.9.4463-4477.2004

Keele, B.F., E.E. Giorgi, J.F. Salazar-Gonzalez, J.M. Decker, K.T. Pham, M.G. Salazar, C. Sun, T. Grayson, S. Wang, H. Li, et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA.* 105:7552–7557. http://dx.doi.org/10.1073/pnas.0802203105

Kwok, W.W., V. Tan, L. Gillette, C.T. Littell, M.A. Soltis, R.B. LaFond, J. Yang, E.A. James, and J.H. DeLong. 2012. Frequency of epitope-specific naive $CD4(^+)$ T cells correlates with immunodominance in the human memory repertoire. *J. Immunol.* 188:2537–2544. http://dx.doi.org/10.4049/jimmunol.1102190

Lefkovits, I., and H. Waldmann. 1979. Limiting Dilution Analysis of Cells of the Immune System. Cambridge University Press, Cambridge. 262 pp.

Malavige, G.N., L. Jones, A.P. Black, and G.S. Ogg. 2008. Varicella zoster virus glycoprotein E-specific $CD4^+$ T cells show evidence of recent activation and effector differentiation, consistent with frequent exposure to replicative cycle antigens in healthy immune donors. *Clin. Exp. Immunol.* 152:522–531. http://dx.doi.org/10.1111/j.1365-2249.2008.03633.x

Moon, J.J., H.H. Chu, M. Pepper, S.J. McSorley, S.C. Jameson, R.M. Kedl, and M.K. Jenkins. 2007. Naive $CD4(^+)$ T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity.* 27:203–213. http://dx.doi.org/10.1016/j.immuni.2007.07.007

Ranasinghe, S., M. Flanders, S. Cutler, D.Z. Soghoian, M. Ghebremichael, I. Davis, M. Lindqvist, F. Pereyra, B.D. Walker, D. Heckerman, and H. Streeck. 2012. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol.* 86:277–283. http://dx.doi.org/10.1128/JVI.05577-11

Rerks-Ngarm, S., P. Pitisuttithum, S. Nitayaphan, J. Kaewkungwal, J. Chiu, R. Paris, N. Premsri, C. Namwat, M. de Souza, E. Adams, et al. MOPH-TAVEG Investigators. 2009. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* 361:2209–2220. http://dx.doi.org/10.1056/NEJMoa0908492

Ritchie, A.J., S.L. Campion, J. Kopycinski, Z. Moodie, Z.M. Wang, K. Pandya, S. Moore, M.K. Liu, S. Brackenridge, K. Kuldanek, et al. 2011. Differences in HIV-specific T cell responses between HIV-exposed and -unexposed HIV-seronegative individuals. *J. Virol.* 85:3507–3516. http://dx.doi.org/10.1128/JVI.02444-10

Roederer, M., and R.A. Koup. 2003. Optimized determination of T cell epitope responses. *J. Immunol. Methods.* 274:221–228. http://dx.doi.org/10.1016/S0022-1759(02)00423-4

Rosenberg, E.S., J.M. Billingsley, A.M. Caliendo, S.L. Boswell, P.E. Sax, S.A. Kalams, and B.D. Walker. 1997. Vigorous HIV-1-specific $CD4^+$ T cell responses associated with control of viremia. *Science.* 278:1447–1450. http://dx.doi.org/10.1126/science.278.5342.1447

Sallusto, F., A. Lanzavecchia, K. Araki, and R. Ahmed. 2010. From vaccines to memory and back. *Immunity.* 33:451–463. http://dx.doi.org/10.1016/j.immuni.2010.10.008

Su, L.F., B.A. Kidd, A. Han, J.J. Kotzin, and M.M. Davis. 2013. Virus-specific $CD4(^+)$ memory-phenotype T cells are abundant in unexposed adults. *Immunity.* 38:373–383. http://dx.doi.org/10.1016/j.immuni.2012.10.021

UniProt Consortium. 2012. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 40:D71–D75. http://dx.doi.org/10.1093/nar/gkr981

W.H.O. 1978. Ebola haemorrhagic fever in Zaire, 1976. *Bull. World Health Organ.* 56:271–293.

Younes, S.A., B. Yassine-Diab, A.R. Dumont, M.R. Boulassel, Z. Grossman, J.P. Routy, and R.P. Sekaly. 2003. HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory $CD4^+$ T cells endowed with proliferative capacity. *J. Exp. Med.* 198:1909–1922. http://dx.doi.org/10.1084/jem.20031598

Table S3. HIV-1 specific epitopes mapped in the pre-exposure repertoire were compared against the Los Alamos database to establish whether they had previously been reported in natural infection

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 33→50	Leuk 7	Central Memory	**HLVWASRELERFALNPGL**	
			----------L	Ramduth et al., 2009
			K-I----------V	Chevalier et al., 2011
			K-I----------V	Kaufmann et al., 2004
			-I----------VN	Wahren et al., 1989
			-I----------VN	Koeppe et al., 2006
			-I----------VN	Jones et al., 2009
			-------VN	Kaufmann et al., 2004
			----------	Ranasinghe et al., 2012
Gag 41→58	Leuk 7	Naïve	**LERFALNPGLLETSEGCK**	
			----------	Kaushik et al., 2005
			----------	Ramduth et al., 2009
Gag 57→74	Leuk 7	Central Memory	**CKQIIKQLQPALQTGTEE**	
Gag 73→90	Leuk 9	Central Memory	**EELRSLYNTVATLYCVHE**	
			TGS----------	Chevalier et al., 2011
			TGS----------	Kaufmann et al., 2004
			----------	Jones et al., 2009
			----------	Fonseca et al., 2006
Gag 81→98	Leuk 10	Naïve	**TVATLYCVHEKIEVRDTK**	
			SLYN---------QR---	Chevalier et al., 2011
			SLYN---------QR---	Kaufmann et al., 2004
			SLYN---------QR---	Geels et al., 2006
Gag 228→245	Leuk 9	Central Memory	**MREPRGSDIAGTTSTLQE**	
			----------QIGWM	Boritz et al., 2007
			----------QI	Koeppe et al., 2006
Gag 236→253	Leuk 9	Central Memory	**IAGTTSTLQEQIAWMTSN**	
			PRGSD----------	Boritz et al., 2007
			GSD----------	Koeppe et al., 2006
			---------PPVPVG	Kaushik et al., 2005

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 260→277	Leuk 7	Effector Memory	**DIYKRWIILGLNKIVRMY**	
			PVG----------	Ramduth et al., 2009
			VGE----------SPV	De Groot et al., 2005
			----------	Boaz et al., 2003
			----------	Wilson et al., 2001
			----------SPTS	Vingert et al., 2010
			----------SPTSILD	Rosenberg et al., 1997
			----------SP	Koeppe et al., 2006
			----------SP	Jones et al., 2009
			----------SPTSI	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------SPTSILDIR	Adams et al., 1997
			----------SPVSILDIRQGP	De Groot et al., 2005
Gag 268→285	Leuk 7	Naïve	**LGLNKIVRMYSPVSILDI**	
			----------T-----	Ritchie et al., 2011
			I----------RQGP	De Groot et al., 2005
			PVGDIYKRWI--------	Ramduth et al., 2009
			VGEDIYKRWI----------	De Groot et al., 2005
			YKRWI----------	Boaz et al., 2003
			YKRWI----------	Wilson et al., 2001
			YKRWI----------T-	Vingert et al., 2010
			YKRWI----------T----	Rosenberg et al., 1997
			I----------	Koeppe et al., 2006
			----------	Jones et al., 2009
			----------T--	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------T----R	Adams et al., 1997
Gag 276→293	Leuk 7	Naïve	**MYSPVSILDIKQGPKEPF**	
			ILGLNKIVR----------	De Groot et al., 2005
			LGLNKIVR----------SPTSILD	De Groot et al., 2005
			----------RDYV	Chevalier et al., 2011
			----------RDYVDRFY	Schrier et al., 1989
			--------RDYVDR	Younes et al., 2003
Gag 292→309	Leuk 7	Naïve	**PFRDYVDRFFKTLRAEQA**	
			ILDIRQGPKE---------Y	Schrier et al., 1989
			IRQGPKE--------	Younes et al., 2003
			GPKE---------Y-	Wahren et al., 1989
			GPKE---------Y----	Kaufmann et al., 2004
			QPKE----------	Adams et al., 1997
			E----------S	Boritz et al., 2007

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Supplementary Table 3 (Continued):

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Gag 397→414	Leuk 10	Naïve	**KEGHIARNCRAPRKKGCW**	
			---K----------K	Kaufmann et al., 2004
			---KN----------K	Chevalier et al., 2011
Gag 429→446	Leuk 7	Naïve	**RQANFLGKIWPSHKGRPG**	
			MKDCTE---------	Wahren et al., 1989
			----------	Kaufmann et al., 2004
			----------	Chevalier et al., 2011
			----------NFLQSR	Kaufmann et al., 2004
			----------NFLQSR	Ramduth et al., 2009
			----------NFLQSR	Chevalier et al., 2011
Gag 445→462	Leuk 7	Naïve	**PGNFLQNRPEPTAPPAES**	
	Leuk 7	Naïve	---S----------FRF	Chevalier et al., 2011
	Leuk 10	Naïve	-----S---------	
Gag 453→472	Leuk 7	Naïve	**PEPTAPPAESFRFEETTP**	
			FLQSR-------E-----	Chevalier et al., 2011
			----E-----GEE---PSQK	Kaufmann et al., 2004
Gag 461→480	Leuk 7	Naïve	**ESFRFEETTPAPKQEPKD**	
			TAPPE----------S--	Kaufmann et al., 2004
			----------	Kaushik et al., 2005
Pol 313→330	Leuk 7	Naïve	**AIFQSSMTKILEPFRAQN**	
			SP----------	Wilson et al., 2001
			SP----------	Boaz et al., 2003
Pol 377→394	Leuk 7	Central Memory	**QKEPPFLWMGYELHPDKW**	
Pol 713→730	Leuk 10	Central Memory	**KVLFLDGIDKAQEEHEKY**	
Pol 777→794	Leuk 6	Effector Memory	**QLDCTHLEGKIILVAVHV**	
			----------ASGYI	Fonseca et al., 2006
Pol 945→962	Leuk 7	Central Memory	**SRDPIWKGPAKLLWKGEG**	
Env 1→19	Leuk 9	Central Memory	**MRVRGILRNCQQWWIWGI**	
Env 9→27	Leuk 9	Central Memory	**NCQQWWIWGILGFWMLMI**	
Env 90→107	Leuk 7	Naïve	**TENFNMWKNDMVDQMHED**	
			PQEVVLVNV----------	Geretti et al., 1994
Env 148→165	Leuk 9	Central Memory	**TNTMGEIKNCSFNITTEL**	
	Leuk 9	Central Memory	TNPTSSSWGMMEK-------	Mirano-Bascos et al., 2008
			SSSGRMIMEK----------	Geretti et al., 1994
			MEK----------SIRNK	Mirano-Bascos et al., 2008
			----------SIRGKVQK	Geretti et al., 1994

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Env 156→173	Leuk 9	Central Memory	**NCSFNITTELRDKKQKVY** GEIK-------TSIRG-V-- MEKGEIK----Y---SIRNK K----------I---	 Geretti et al., 1994 Mirano-Bascos et al., 2008 Gaudebout et al., 1997
Env 164→181	Leuk 8	Effector Memory	**ELRDKKQKVYALFYRLDI** ----------VPLTK	 Harari et al., 2008
Env 172→189	Leuk 8	Effector Memory	**VYALFYRLDIVPLNENNS** ----------TK	 Harari et al., 2008
Env 213→230	Leuk 7	Naïve	**IPIHYCAPAGYAILKCNN** P----------F-------K -----P----------F-	 Geretti et al., 1994 Ranasinghe et al., 2012
Env 269→286	Leuk 10	Naïve	**EIIIRSENLTNNAKTIIV** -DIV----F-D-------Q -VV-----F---------QLNES	 Mirano-Bascos et al., 2008 De Groot et al., 2004
Env 277→294	Leuk 7	Naïve	**LTNNAKTIIVHLNESVEI** NF-D----------QIN IRSVNF-D-------Q---T- SANF-D-------Q-	 Malhotra et al., 2003 Geretti et al., 1994 Wahren et al., 1989
Env 285→302	Leuk 7	Naïve	**IVHLNESVEIVCTRPNNN** NF-D----------QIN IRSVNF-D-------Q---T-	 Malhotra et al., 2003 Geretti et al., 1994
Env 334→351	Leuk 7 Leuk 9	Central Memory Central Memory	**SEDKWNKTLQKVSKKLKE**	
Env 342→360	Leuk 7	Central Memory	**LQKVSKKLKEHFPNKTIK**	
Env 411→428	Leuk 7 Leuk 10	Central Memory Naïve	**NSTITLPCRIKQIINMWQ** D----------KVG D----------KVG ----------KVGKA ----------KVGKA	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 419→436	Leuk 7 Leuk 7 Leuk 10	Naïve Central Memory Naïve	**RIKQIINMWQEVGRAMYA** DTITLPC---------- DTITLPC---------- ITLQC----------K- ----------K-	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 435→452	Leuk 7	Naïve	**YAPPIAGNITCKSNITGL**	

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Env 477→494	Leuk 7	Naïve	**DNWRSELYKYKVVEIKPL** GGDMR---------- DMR---------- -------K-----GVAPTKA -------R----GVAPTRAK	 Geretti et al., 1994. Wahren et al., 1989 Geretti et al., 1994. Mirano-Bascos et al., 2008
Env 565→582	Leuk 9 Leuk 8	Effector Memory Central Memory	**MLQLTVWGIKQLQTRVLA** ----------VERYLK QQHL----------	 Malhotra et al., 2003 Wahren et al., 1989
Env 613→630	Leuk 8	Effector Memory	**SWSNKSQEDIWDNMTWMQ**	
Env 669→686	Leuk 7	Naïve	**LWNWFDITNWLWYIKIFI** AS-----N-------	 Wahren et al., 1989
Env 725→742	Leuk 9	Central Memory	**RGPDRLGRIEEEGGEQDR** ---------R--DR	 Schrier et al., 1989
Env 733→750	Leuk 9	Central Memory	**IEEEGGEQDRDRSIRLVS** GR---------R--DR	 Schrier et al., 1989
Env 773→790	Leuk 8	Effector Memory	**DFILIAARAVELLGRSSL**	
Env 790→807	Leuk 6 Leuk 8	Effector Memory Central Memory	**WEALKYLGSLVQYWGLEL** RIVELLGRRG----------KNSAVS RIVELLGRRG----------KNSAVS	 Berzofsky et al., 1991 Berzofsky et al., 1991
Env 798→815	Leuk 6	Effector Memory	**SLVQYWGLELKKSAISLL**	
Env 830→847	Leuk 7	Central Memory	**IELIQRICRAIRNIPRRI**	
Nef 73→90	Leuk 9	Central Memory	**QVPLRPMTYKAAFDLSFF** VGFPVRP---------- ---------H-LKEKGGL	 De Groot et al., 2005. Ranasinghe et al., 2012
Nef 129→146	Leuk 10	Naïve	**PGPGVRYPLTFGWCFKLV** ----------Y---PVEPDKVEEANKG	 Pancré et al., 2007
Rev 17→34	Leuk 9	Central Memory	**RIIKILYQSNPYPKPEGT**	
Rev 49→66	Leuk 10	Central Memory	**QRQIHSISERILSTCLGR** RRRRWRER---------	 Blazevic et al., 1995
Vif 1→18	Leuk 9	Central Memory	**MENRWQVLIVWQVDRMKI** ----------RTWNSLVK	 De Groot et al., 2005.
Vif 81→98	Leuk 10	Central Memory	**LGHGVSIEWRLRRYSTQV** ----------KQ----	 Ranki et al., 1997
Vpu 2 →20	Leuk 8	Central Memory	**ARVDYRLGVGALIVALII**	
Vpu 17→34	Leuk 9	Central Memory	**IIAIVVWTIVYIEYRKLL**	
Vpu 33→50	Leuk 10	Naïve	**LLRQRKIDWLIKRIRERA**	
Vpu 41→58	Leuk 9	Central Memory	**WLIKRIRERAEDSGNESE**	

	Leuk 9	Effector Memory		

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

References

Adams, S.L., R.A. Biti, and G.J. Stewart. 1997. T-cell response to HIV in natural infection: optimized culture conditions for detecting responses to gag peptides. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 15:257–263.

Berzofsky, J.A., C.D. Pendleton, M. Clerici, J. Ahlers, D.R. Lucey, S.D. Putney, and G.M. Shearer. 1991. Peptides containing multideterminant clusters of human immunodeficiency virus envelope induce murine and human T-cell responses in diverse histocompatibility types. *Trans. Assoc. Am. Physicians.* 104:69–77.

Blazevic, V., A. Ranki, and K.J. Krohn. 1995. Helper and cytotoxic T cell responses of HIV type 1-infected individuals to synthetic peptides of HIV type 1 Rev. *AIDS Res. Hum. Retroviruses.* 11:1335–1342.

Boaz, M.J., A. Waters, S. Murad, P.J. Easterbrook, E. D'Sousa, C. van Wheeley, and A. Vyakarnam. 2003. CD4 responses to conserved HIV-1 T helper epitopes show both negative and positive associations with virus load in chronically infected subjects. *Clin. Exp. Immunol.* 134:454–463.

Boritz, E., E.L. Rapaport, T.B. Campbell, J.R. Koeppe, and C.C. Wilson. 2007. $CD4^+$ T cell targeting of human immunodeficiency virus type 1 (HIV-1) peptide sequences present in vivo during chronic, progressive HIV-1 disease. *Virology.* 361:34–44.

Chevalier, M.F., B. Jülg, A. Pyo, M. Flanders, S. Ranasinghe, D.Z. Soghoian, D.S. Kwon, J. Rychert, J. Lian, M.I. Muller, et al. 2011. HIV-1-specific interleukin-21^+ $CD4^+$ T cell responses contribute to durable viral control through the modulation of HIV-specific $CD8^+$ T cell function. *J. Virol.* 85:733–741.

De Groot, A.S., E.A. Bishop, B. Khan, M. Lally, L. Marcon, J. Franco, K.H. Mayer, C.C. Carpenter, and W. Martin. 2004. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods.* 34:476–487.

De Groot, A.S., L. Marcon, E.A. Bishop, D. Rivera, M. Kutzler, D.B. Weiner, and W. Martin. 2005. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine.* 23:2136–2148.

Fonseca, S.G., A. Coutinho-Silva, L.A. Fonseca, A.C. Segurado, S.L. Moraes, H. Rodrigues, J. Hammer, E.G. Kallás, J. Sidney, A. Sette, et al. 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS.* 20:2263–2273.

Gaudebout, P., D. Zeliszewski, J.J. Golvano, C. Pignal, S. Le Gac, F. Borras-Cuesta, and G. Sterkers. 1997. Binding analysis of 95 HIV gp120 peptides to HLA-DR1101 and –DR0401 evidenced many HLA-class II binding regions on gp120 and suggested several promiscuous regions. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 14:91–101.

Geels, M.J., C.A. Jansen, E. Baan, I.M. De Cuyper, G.J. van Schijndel, H. Schuitemaker, J. Goudsmit, G. Pollakis, F. Miedema, W.A. Paxton, and D. van Baarle. 2006. CTL escape and increased viremia irrespective of HIV-specific CD4+ T-helper responses in two HIV-infected individuals. *Virology.* 345:209–219.

Geretti, A.M., C.A. Van Baalen, J.C. Borleffs, C.A. Van Els, and A.D. Osterhaus. 1994. Kinetics and specificities of the T helper-cell response to gp120 in the asymptomatic stage of HIV-1 infection. *Scand. J. Immunol.* 39:355–362.

Harari, A., P.A. Bart, W. Stöhr, G. Tapia, M. Garcia, E. Medjitna-Rais, S. Burnet, C. Cellerai, O. Erlwein, T. Barber, et al. 2008. An HIV-1 clade C DNA prime, NYVAC boost vaccine regimen induces reliable, polyfunctional, and long-lasting T cell responses. *J. Exp. Med.* 205:63–77.

Jones, R.B., F.Y. Yue, X.X. Gu, D.V. Hunter, S. Mujib, G. Gyenes, R.D. Mason, R. Mohamed, K.S. MacDonald, C. Kovacs, and M.A. Ostrowski. 2009. Human immunodeficiency virus type 1 escapes from interleukin-2-producing $CD4^+$ T-cell responses without high-frequency fixation of mutations. *J. Virol.* 83:8722–8732.

Kaufmann, D.E., P.M. Bailey, J. Sidney, B. Wagner, P.J. Norris, M.N. Johnston, L.A. Cosimi, M.M. Addo, M. Lichterfeld, M. Altfeld, et al. 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. *J. Virol.* 78:4463–4477.

Kaushik, S., M. Vajpayee, N. Wig, and P. Seth. 2005. Characterization of HIV-1 Gag-specific T cell responses in chronically infected Indian population. *Clin. Exp. Immunol.* 142:388–397.

Koeppe, J.R., T.B. Campbell, E.L. Rapaport, and C.C. Wilson. 2006. HIV-1-specific $CD4^+$ T-cell responses are not associated with significant viral epitope variation in persons with persistent plasma viremia. *J. Acquir. Immune Defic. Syndr.* 41:140–148.

Koup, R.A., M. Roederer, L. Lamoreaux, J. Fischer, L. Novik, M.C. Nason, B.D. Larkin, M.E. Enama, J.E. Ledgerwood, R.T. Bailer, et al., VRC 010 Study Team. 2010. Priming immunization with DNA augments immunogenicity of recombinant adenoviral vectors for both HIV-1 specific antibody and T-cell responses. *PLoS ONE.* 5:e9015.

Malhotra, U., S. Holte, T. Zhu, E. Delpit, C. Huntsberry, A. Sette, R. Shankarappa, J. Maenza, L. Corey, and M.J. McElrath. 2003. Early induction and maintenance of Env-specific T-helper cells following human immunodeficiency virus type 1 infection. *J. Virol.* 77:2663–2674.

Mathiesen, T., A. Sönnerborg, and B. Wahren. 1989. Detection of antibodies against myelin basic protein and increased levels of HIV-IgG antibodies and HIV antigen after solubilization of immune complexes in sera and CSF of HIV infected patients. *Viral Immunol.* 2:1–9.

Mirano-Bascos, D., M. Tary-Lehmann, and S.J. Landry. 2008. Antigen structure influences helper T-cell epitope dominance in the human immune response to HIV envelope glycoprotein gp120. *Eur. J. Immunol.* 38:1231–1237.

Pancré, V., N. Delhem, Y. Yazdanpanah, A. Delanoye, M. Delacre, S. Depil, O. Moralès, Y. Mouton, and C. Auriault. 2007. Presence of HIV-1 Nef specific CD4 T cell response is associated with non-progression in HIV-1 infection. *Vaccine.* 25:5927–5937.

Ramduth, D., C.L. Day, C.F. Thobakgale, N.P. Mkhwanazi, C. de Pierres, S. Reddy, M. van der Stok, Z. Mncube, K. Nair, E.S. Moodley, et al. 2009. Immunodominant HIV-1 Cd4^{+} T cell epitopes in chronic untreated clade C HIV-1 infection. *PLoS ONE*. 4:e5013.

Ranasinghe, S., M. Flanders, S. Cutler, D.Z. Soghoian, M. Ghebremichael, I. Davis, M. Lindqvist, F. Pereyra, B.D. Walker, D. Heckerman, and H. Streeck. 2012. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol*. 86:277–283.

Ranki, A., J. Suni, V. Blazevic, P. Holmström, S. Mattinen, K. Krohn, and S.L. Valle. 1997. T-cell recognition of HIV antigens in HIV-seroreverted persons. *AIDS*. 11:132–133.

Ritchie, A.J., S.L. Campion, J. Kopycinski, Z. Moodie, Z.M. Wang, K. Pandya, S. Moore, M.K. Liu, S. Brackenridge, K. Kuldanek, K. Legg, M.S. Cohen, E.L. Delwart, B.F. Haynes, S. Fidler, A.J. McMichael, and N. Goonetilleke. 2011. Differences in HIV-specific T cell responses between HIV-exposed and –unexposed HIV-seronegative individuals. *J. Virol*. 85:3507–3516.

Rosenberg, E.S., J.M. Billingsley, A.M. Caliendo, S.L. Boswell, P.E. Sax, S.A. Kalams, and B.D. Walker. 1997. Vigorous HIV-1-specific CD4^{+} T cell responses associated with control of viremia. *Science*. 278:1447–1450.

Schrier, R.D., J.W. Gnann Jr., R. Landes, C. Lockshin, D. Richman, A. McCutchan, C. Kennedy, M.B. Oldstone, and J.A. Nelson. 1989. T cell recognition of HIV synthetic peptides in a natural infection. *J. Immunol*. 142:1166–1176.

Vingert, B., S. Perez-Patrigeon, P. Jeannin, O. Lambotte, F. Boufassa, F. Lemaître, W.W. Kwok, I. Theodorou, J.F. Delfraissy, J. Thèze, and L.A. Chakrabarti; ANRS EP36 HIV Controllers Study Group. 2010. HIV controller CD4^{+} T cells respond to minimal amounts of Gag antigen due to high TCR avidity. *PLoS Pathog*. 6:e1000780.

Wahren, B., J. Rosen, E. Sandström, T. Mathiesen, S. Modrow, and H. Wigzell. 1989. HIV-1 peptides induce a proliferative response in lymphocytes from infected persons. *J. Acquir. Immune Defic. Syndr*. 2:448–456.

Wilson, C.C., B. Palmer, S. Southwood, J. Sidney, Y. Higashimoto, E. Appella, R. Chesnut, A. Sette, and B.D. Livingston. 2001. Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J. Virol*. 75:4195–4207.

Younes, S.A., B. Yassine-Diab, A.R. Dumont, M.R. Boulassel, Z. Grossman, J.P. Routy, and R.P. Sekaly. 2003. HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory CD4^{+} T cells endowed with proliferative capacity. *J. Exp. Med*. 198:1909–1922.

Table S3. HIV-1 specific epitopes mapped in the pre-exposure repertoire were compared against the Los Alamos database to establish whether they had previously been reported in natural infection

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 33→50	Leuk 7	Central Memory	**HLVWASRELERFALNPGL**	
			----------L	Ramduth et al., 2009
			K-I----------V	Chevalier et al., 2011
			K-I----------V	Kaufmann et al., 2004
			-I----------VN	Wahren et al., 1989
			-I----------VN	Koeppe et al., 2006
			-I----------VN	Jones et al., 2009
			-------VN	Kaufmann et al., 2004
			----------	Ranasinghe et al., 2012
Gag 41→58	Leuk 7	Naïve	**LERFALNPGLLETSEGCK**	
			----------	Kaushik et al., 2005
			----------	Ramduth et al., 2009
Gag 57→74	Leuk 7	Central Memory	**CKQIIKQLQPALQTGTEE**	
Gag 73→90	Leuk 9	Central Memory	**EELRSLYNTVATLYCVHE**	
			TGS----------	Chevalier et al., 2011
			TGS----------	Kaufmann et al., 2004
			----------	Jones et al., 2009
			----------	Fonseca et al., 2006
Gag 81→98	Leuk 10	Naïve	**TVATLYCVHEKIEVRDTK**	
			SLYN---------QR---	Chevalier et al., 2011
			SLYN---------QR---	Kaufmann et al., 2004
			SLYN---------QR---	Geels et al., 2006
Gag 228→245	Leuk 9	Central Memory	**MREPRGSDIAGTTSTLQE**	
			----------QIGWM	Boritz et al., 2007
			----------QI	Koeppe et al., 2006
Gag 236→253	Leuk 9	Central Memory	**IAGTTSTLQEQIAWMTSN**	
			PRGSD----------	Boritz et al., 2007
			GSD----------	Koeppe et al., 2006
			---------PPVPVG	Kaushik et al., 2005

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 260→277	Leuk 7	Effector Memory	**DIYKRWIILGLNKIVRMY**	
			PVG----------	Ramduth et al., 2009
			VGE----------SPV	De Groot et al., 2005
			----------	Boaz et al., 2003
			----------	Wilson et al., 2001
			----------SPTS	Vingert et al., 2010
			----------SPTSILD	Rosenberg et al., 1997
			----------SP	Koeppe et al., 2006
			----------SP	Jones et al., 2009
			----------SPTSI	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------SPTSILDIR	Adams et al., 1997
			----------SPVSILDIRQGP	De Groot et al., 2005
Gag 268→285	Leuk 7	Naïve	**LGLNKIVRMYSPVSILDI**	
			----------T-----	Ritchie et al., 2011
			I----------RQGP	De Groot et al., 2005
			PVGDIYKRWI--------	Ramduth et al., 2009
			VGEDIYKRWI----------	De Groot et al., 2005
			YKRWI----------	Boaz et al., 2003
			YKRWI----------	Wilson et al., 2001
			YKRWI----------T-	Vingert et al., 2010
			YKRWI----------T----	Rosenberg et al., 1997
			I----------	Koeppe et al., 2006
			----------	Jones et al., 2009
			----------T--	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------T----R	Adams et al., 1997
Gag 276→293	Leuk 7	Naïve	**MYSPVSILDIKQGPKEPF**	
			ILGLNKIVR----------	De Groot et al., 2005
			LGLNKIVR----------SPTSILD	De Groot et al., 2005
			----------RDYV	Chevalier et al., 2011
			----------RDYVDRFY	Schrier et al., 1989
			--------RDYVDR	Younes et al., 2003
Gag 292→309	Leuk 7	Naïve	**PFRDYVDRFFKTLRAEQA**	
			ILDIRQGPKE---------Y	Schrier et al., 1989
			IRQGPKE--------	Younes et al., 2003
			GPKE---------Y-	Wahren et al., 1989
			GPKE---------Y----	Kaufmann et al., 2004
			QPKE----------	Adams et al., 1997
			E----------S	Boritz et al., 2007

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Supplementary Table 3 (Continued):

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Gag 397→414	Leuk 10	Naïve	**KEGHIARNCRAPRKKGCW**	
			---K----------K	Kaufmann et al., 2004
			---KN----------K	Chevalier et al., 2011
Gag 429→446	Leuk 7	Naïve	**RQANFLGKIWPSHKGRPG**	
			MKDCTE---------	Wahren et al., 1989
			----------	Kaufmann et al., 2004
			----------	Chevalier et al., 2011
			----------NFLQSR	Kaufmann et al., 2004
			----------NFLQSR	Ramduth et al., 2009
			----------NFLQSR	Chevalier et al., 2011
Gag 445→462	Leuk 7	Naïve	**PGNFLQNRPEPTAPPAES**	
	Leuk 7	Naïve	---S----------FRF	Chevalier et al., 2011
	Leuk 10	Naïve	-----S---------	
Gag 453→472	Leuk 7	Naïve	**PEPTAPPAESFRFEETTP**	
			FLQSR-------E-----	Chevalier et al., 2011
			----E-----GEE---PSQK	Kaufmann et al., 2004
Gag 461→480	Leuk 7	Naïve	**ESFRFEETTPAPKQEPKD**	
			TAPPE----------S--	Kaufmann et al., 2004
			----------	Kaushik et al., 2005
Pol 313→330	Leuk 7	Naïve	**AIFQSSMTKILEPFRAQN**	
			SP----------	Wilson et al., 2001
			SP----------	Boaz et al., 2003
Pol 377→394	Leuk 7	Central Memory	**QKEPPFLWMGYELHPDKW**	
Pol 713→730	Leuk 10	Central Memory	**KVLFLDGIDKAQEEHEKY**	
Pol 777→794	Leuk 6	Effector Memory	**QLDCTHLEGKIILVAVHV**	
			----------ASGYI	Fonseca et al., 2006
Pol 945→962	Leuk 7	Central Memory	**SRDPIWKGPAKLLWKGEG**	
Env 1→19	Leuk 9	Central Memory	**MRVRGILRNCQQWWIWGI**	
Env 9→27	Leuk 9	Central Memory	**NCQQWWIWGILGFWMLMI**	
Env 90→107	Leuk 7	Naïve	**TENFNMWKNDMVDQMHED**	
			PQEVVLVNV----------	Geretti et al., 1994
Env 148→165	Leuk 9	Central Memory	**TNTMGEIKNCSFNITTEL**	
	Leuk 9	Central Memory	TNPTSSSWGMMEK-------	Mirano-Bascos et al., 2008
			SSSGRMIMEK----------	Geretti et al., 1994
			MEK----------SIRNK	Mirano-Bascos et al., 2008
			----------SIRGKVQK	Geretti et al., 1994

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Env 156→173	Leuk 9	Central Memory	**NCSFNITTELRDKKQKVY** GEIK-------TSIRG-V-- MEKGEIK----Y---SIRNK K----------I---	 Geretti et al., 1994 Mirano-Bascos et al., 2008 Gaudebout et al., 1997
Env 164→181	Leuk 8	Effector Memory	**ELRDKKQKVYALFYRLDI** ----------VPLTK	 Harari et al., 2008
Env 172→189	Leuk 8	Effector Memory	**VYALFYRLDIVPLNENNS** ----------TK	 Harari et al., 2008
Env 213→230	Leuk 7	Naïve	**IPIHYCAPAGYAILKCNN** P----------F-------K -----P----------F-	 Geretti et al., 1994 Ranasinghe et al., 2012
Env 269→286	Leuk 10	Naïve	**EIIIRSENLTNNAKTIIV** -DIV----F-D-------Q -VV-----F---------QLNES	 Mirano-Bascos et al., 2008 De Groot et al., 2004
Env 277→294	Leuk 7	Naïve	**LTNNAKTIIVHLNESVEI** NF-D----------QIN IRSVNF-D-------Q---T- SANF-D-------Q-	 Malhotra et al., 2003 Geretti et al., 1994 Wahren et al., 1989
Env 285→302	Leuk 7	Naïve	**IVHLNESVEIVCTRPNNN** NF-D----------QIN IRSVNF-D-------Q---T-	 Malhotra et al., 2003 Geretti et al., 1994
Env 334→351	Leuk 7 Leuk 9	Central Memory Central Memory	**SEDKWNKTLQKVSKKLKE**	
Env 342→360	Leuk 7	Central Memory	**LQKVSKKLKEHFPNKTIK**	
Env 411→428	Leuk 7 Leuk 10	Central Memory Naïve	**NSTITLPCRIKQIINMWQ** D----------KVG D----------KVG ----------KVGKA ----------KVGKA	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 419→436	Leuk 7 Leuk 7 Leuk 10	Naïve Central Memory Naïve	**RIKQIINMWQEVGRAMYA** DTITLPC---------- DTITLPC---------- ITLQC----------K- ----------K-	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 435→452	Leuk 7	Naïve	**YAPPIAGNITCKSNITGL**	

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Env 477→494	Leuk 7	Naïve	**DNWRSELYKYKVVEIKPL** GGDMR---------- DMR---------- -------K-----GVAPTKA -------R----GVAPTRAK	 Geretti et al., 1994. Wahren et al., 1989 Geretti et al., 1994. Mirano-Bascos et al., 2008
Env 565→582	Leuk 9 Leuk 8	Effector Memory Central Memory	**MLQLTVWGIKQLQTRVLA** ----------VERYLK QQHL----------	 Malhotra et al., 2003 Wahren et al., 1989
Env 613→630	Leuk 8	Effector Memory	**SWSNKSQEDIWDNMTWMQ**	
Env 669→686	Leuk 7	Naïve	**LWNWFDITNWLWYIKIFI** AS-----N-------	 Wahren et al., 1989
Env 725→742	Leuk 9	Central Memory	**RGPDRLGRIEEEGGEQDR** ---------R--DR	 Schrier et al., 1989
Env 733→750	Leuk 9	Central Memory	**IEEEGGEQDRDRSIRLVS** GR---------R--DR	 Schrier et al., 1989
Env 773→790	Leuk 8	Effector Memory	**DFILIAARAVELLGRSSL**	
Env 790→807	Leuk 6 Leuk 8	Effector Memory Central Memory	**WEALKYLGSLVQYWGLEL** RIVELLGRRG----------KNSAVS RIVELLGRRG----------KNSAVS	 Berzofsky et al., 1991 Berzofsky et al., 1991
Env 798→815	Leuk 6	Effector Memory	**SLVQYWGLELKKSAISLL**	
Env 830→847	Leuk 7	Central Memory	**IELIQRICRAIRNIPRRI**	
Nef 73→90	Leuk 9	Central Memory	**QVPLRPMTYKAAFDLSFF** VGFPVRP---------- ---------H-LKEKGGL	 De Groot et al., 2005. Ranasinghe et al., 2012
Nef 129→146	Leuk 10	Naïve	**PGPGVRYPLTFGWCFKLV** ----------Y---PVEPDKVEEANKG	 Pancré et al., 2007
Rev 17→34	Leuk 9	Central Memory	**RIIKILYQSNPYPKPEGT**	
Rev 49→66	Leuk 10	Central Memory	**QRQIHSISERILSTCLGR** RRRRWRER---------	 Blazevic et al., 1995
Vif 1→18	Leuk 9	Central Memory	**MENRWQVLIVWQVDRMKI** ----------RTWNSLVK	 De Groot et al., 2005.
Vif 81→98	Leuk 10	Central Memory	**LGHGVSIEWRLRRYSTQV** ----------KQ----	 Ranki et al., 1997
Vpu 2 →20	Leuk 8	Central Memory	**ARVDYRLGVGALIVALII**	
Vpu 17→34	Leuk 9	Central Memory	**IIAIVVWTIVYIEYRKLL**	
Vpu 33→50	Leuk 10	Naïve	**LLRQRKIDWLIKRIRERA**	
Vpu 41→58	Leuk 9	Central Memory	**WLIKRIRERAEDSGNESE**	

	Leuk 9	Effector Memory		

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

References

Adams, S.L., R.A. Biti, and G.J. Stewart. 1997. T-cell response to HIV in natural infection: optimized culture conditions for detecting responses to gag peptides. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 15:257–263.

Berzofsky, J.A., C.D. Pendleton, M. Clerici, J. Ahlers, D.R. Lucey, S.D. Putney, and G.M. Shearer. 1991. Peptides containing multideterminant clusters of human immunodeficiency virus envelope induce murine and human T-cell responses in diverse histocompatibility types. *Trans. Assoc. Am. Physicians.* 104:69–77.

Blazevic, V., A. Ranki, and K.J. Krohn. 1995. Helper and cytotoxic T cell responses of HIV type 1-infected individuals to synthetic peptides of HIV type 1 Rev. *AIDS Res. Hum. Retroviruses.* 11:1335–1342.

Boaz, M.J., A. Waters, S. Murad, P.J. Easterbrook, E. D'Sousa, C. van Wheeley, and A. Vyakarnam. 2003. CD4 responses to conserved HIV-1 T helper epitopes show both negative and positive associations with virus load in chronically infected subjects. *Clin. Exp. Immunol.* 134:454–463.

Boritz, E., E.L. Rapaport, T.B. Campbell, J.R. Koeppe, and C.C. Wilson. 2007. $CD4^+$ T cell targeting of human immunodeficiency virus type 1 (HIV-1) peptide sequences present in vivo during chronic, progressive HIV-1 disease. *Virology.* 361:34–44.

Chevalier, M.F., B. Jülg, A. Pyo, M. Flanders, S. Ranasinghe, D.Z. Soghoian, D.S. Kwon, J. Rychert, J. Lian, M.I. Muller, et al. 2011. HIV-1-specific interleukin-21^+ $CD4^+$ T cell responses contribute to durable viral control through the modulation of HIV-specific $CD8^+$ T cell function. *J. Virol.* 85:733–741.

De Groot, A.S., E.A. Bishop, B. Khan, M. Lally, L. Marcon, J. Franco, K.H. Mayer, C.C. Carpenter, and W. Martin. 2004. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods.* 34:476–487.

De Groot, A.S., L. Marcon, E.A. Bishop, D. Rivera, M. Kutzler, D.B. Weiner, and W. Martin. 2005. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine.* 23:2136–2148.

Fonseca, S.G., A. Coutinho-Silva, L.A. Fonseca, A.C. Segurado, S.L. Moraes, H. Rodrigues, J. Hammer, E.G. Kallás, J. Sidney, A. Sette, et al. 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS.* 20:2263–2273.

Gaudebout, P., D. Zeliszewski, J.J. Golvano, C. Pignal, S. Le Gac, F. Borras-Cuesta, and G. Sterkers. 1997. Binding analysis of 95 HIV gp120 peptides to HLA-DR1101 and –DR0401 evidenced many HLA-class II binding regions on gp120 and suggested several promiscuous regions. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 14:91–101.

Geels, M.J., C.A. Jansen, E. Baan, I.M. De Cuyper, G.J. van Schijndel, H. Schuitemaker, J. Goudsmit, G. Pollakis, F. Miedema, W.A. Paxton, and D. van Baarle. 2006. CTL escape and increased viremia irrespective of HIV-specific CD4+ T-helper responses in two HIV-infected individuals. *Virology.* 345:209–219.

Geretti, A.M., C.A. Van Baalen, J.C. Borleffs, C.A. Van Els, and A.D. Osterhaus. 1994. Kinetics and specificities of the T helper-cell response to gp120 in the asymptomatic stage of HIV-1 infection. *Scand. J. Immunol.* 39:355–362.

Harari, A., P.A. Bart, W. Stöhr, G. Tapia, M. Garcia, E. Medjitna-Rais, S. Burnet, C. Cellerai, O. Erlwein, T. Barber, et al. 2008. An HIV-1 clade C DNA prime, NYVAC boost vaccine regimen induces reliable, polyfunctional, and long-lasting T cell responses. *J. Exp. Med.* 205:63–77.

Jones, R.B., F.Y. Yue, X.X. Gu, D.V. Hunter, S. Mujib, G. Gyenes, R.D. Mason, R. Mohamed, K.S. MacDonald, C. Kovacs, and M.A. Ostrowski. 2009. Human immunodeficiency virus type 1 escapes from interleukin-2-producing $CD4^+$ T-cell responses without high-frequency fixation of mutations. *J. Virol.* 83:8722–8732.

Kaufmann, D.E., P.M. Bailey, J. Sidney, B. Wagner, P.J. Norris, M.N. Johnston, L.A. Cosimi, M.M. Addo, M. Lichterfeld, M. Altfeld, et al. 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. *J. Virol.* 78:4463–4477.

Kaushik, S., M. Vajpayee, N. Wig, and P. Seth. 2005. Characterization of HIV-1 Gag-specific T cell responses in chronically infected Indian population. *Clin. Exp. Immunol.* 142:388–397.

Koeppe, J.R., T.B. Campbell, E.L. Rapaport, and C.C. Wilson. 2006. HIV-1-specific $CD4^+$ T-cell responses are not associated with significant viral epitope variation in persons with persistent plasma viremia. *J. Acquir. Immune Defic. Syndr.* 41:140–148.

Koup, R.A., M. Roederer, L. Lamoreaux, J. Fischer, L. Novik, M.C. Nason, B.D. Larkin, M.E. Enama, J.E. Ledgerwood, R.T. Bailer, et al., VRC 010 Study Team. 2010. Priming immunization with DNA augments immunogenicity of recombinant adenoviral vectors for both HIV-1 specific antibody and T-cell responses. *PLoS ONE.* 5:e9015.

Malhotra, U., S. Holte, T. Zhu, E. Delpit, C. Huntsberry, A. Sette, R. Shankarappa, J. Maenza, L. Corey, and M.J. McElrath. 2003. Early induction and maintenance of Env-specific T-helper cells following human immunodeficiency virus type 1 infection. *J. Virol.* 77:2663–2674.

Mathiesen, T., A. Sönnerborg, and B. Wahren. 1989. Detection of antibodies against myelin basic protein and increased levels of HIV-IgG antibodies and HIV antigen after solubilization of immune complexes in sera and CSF of HIV infected patients. *Viral Immunol.* 2:1–9.

Mirano-Bascos, D., M. Tary-Lehmann, and S.J. Landry. 2008. Antigen structure influences helper T-cell epitope dominance in the human immune response to HIV envelope glycoprotein gp120. *Eur. J. Immunol.* 38:1231–1237.

Pancré, V., N. Delhem, Y. Yazdanpanah, A. Delanoye, M. Delacre, S. Depil, O. Moralès, Y. Mouton, and C. Auriault. 2007. Presence of HIV-1 Nef specific CD4 T cell response is associated with non-progression in HIV-1 infection. *Vaccine.* 25:5927–5937.

Ramduth, D., C.L. Day, C.F. Thobakgale, N.P. Mkhwanazi, C. de Pierres, S. Reddy, M. van der Stok, Z. Mncube, K. Nair, E.S. Moodley, et al. 2009. Immunodominant HIV-1 Cd4^{+} T cell epitopes in chronic untreated clade C HIV-1 infection. *PLoS ONE*. 4:e5013.

Ranasinghe, S., M. Flanders, S. Cutler, D.Z. Soghoian, M. Ghebremichael, I. Davis, M. Lindqvist, F. Pereyra, B.D. Walker, D. Heckerman, and H. Streeck. 2012. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol*. 86:277–283.

Ranki, A., J. Suni, V. Blazevic, P. Holmström, S. Mattinen, K. Krohn, and S.L. Valle. 1997. T-cell recognition of HIV antigens in HIV-seroreverted persons. *AIDS*. 11:132–133.

Ritchie, A.J., S.L. Campion, J. Kopycinski, Z. Moodie, Z.M. Wang, K. Pandya, S. Moore, M.K. Liu, S. Brackenridge, K. Kuldanek, K. Legg, M.S. Cohen, E.L. Delwart, B.F. Haynes, S. Fidler, A.J. McMichael, and N. Goonetilleke. 2011. Differences in HIV-specific T cell responses between HIV-exposed and –unexposed HIV-seronegative individuals. *J. Virol*. 85:3507–3516.

Rosenberg, E.S., J.M. Billingsley, A.M. Caliendo, S.L. Boswell, P.E. Sax, S.A. Kalams, and B.D. Walker. 1997. Vigorous HIV-1-specific CD4^{+} T cell responses associated with control of viremia. *Science*. 278:1447–1450.

Schrier, R.D., J.W. Gnann Jr., R. Landes, C. Lockshin, D. Richman, A. McCutchan, C. Kennedy, M.B. Oldstone, and J.A. Nelson. 1989. T cell recognition of HIV synthetic peptides in a natural infection. *J. Immunol*. 142:1166–1176.

Vingert, B., S. Perez-Patrigeon, P. Jeannin, O. Lambotte, F. Boufassa, F. Lemaître, W.W. Kwok, I. Theodorou, J.F. Delfraissy, J. Thèze, and L.A. Chakrabarti; ANRS EP36 HIV Controllers Study Group. 2010. HIV controller CD4^{+} T cells respond to minimal amounts of Gag antigen due to high TCR avidity. *PLoS Pathog*. 6:e1000780.

Wahren, B., J. Rosen, E. Sandström, T. Mathiesen, S. Modrow, and H. Wigzell. 1989. HIV-1 peptides induce a proliferative response in lymphocytes from infected persons. *J. Acquir. Immune Defic. Syndr*. 2:448–456.

Wilson, C.C., B. Palmer, S. Southwood, J. Sidney, Y. Higashimoto, E. Appella, R. Chesnut, A. Sette, and B.D. Livingston. 2001. Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J. Virol*. 75:4195–4207.

Younes, S.A., B. Yassine-Diab, A.R. Dumont, M.R. Boulassel, Z. Grossman, J.P. Routy, and R.P. Sekaly. 2003. HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory CD4^{+} T cells endowed with proliferative capacity. *J. Exp. Med*. 198:1909–1922.

Table S3. HIV-1 specific epitopes mapped in the pre-exposure repertoire were compared against the Los Alamos database to establish whether they had previously been reported in natural infection

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 33→50	Leuk 7	Central Memory	**HLVWASRELERFALNPGL**	
			----------L	Ramduth et al., 2009
			K-I----------V	Chevalier et al., 2011
			K-I----------V	Kaufmann et al., 2004
			-I----------VN	Wahren et al., 1989
			-I----------VN	Koeppe et al., 2006
			-I----------VN	Jones et al., 2009
			-------VN	Kaufmann et al., 2004
			----------	Ranasinghe et al., 2012
Gag 41→58	Leuk 7	Naïve	**LERFALNPGLLETSEGCK**	
			----------	Kaushik et al., 2005
			----------	Ramduth et al., 2009
Gag 57→74	Leuk 7	Central Memory	**CKQIIKQLQPALQTGTEE**	
Gag 73→90	Leuk 9	Central Memory	**EELRSLYNTVATLYCVHE**	
			TGS----------	Chevalier et al., 2011
			TGS----------	Kaufmann et al., 2004
			----------	Jones et al., 2009
			----------	Fonseca et al., 2006
Gag 81→98	Leuk 10	Naïve	**TVATLYCVHEKIEVRDTK**	
			SLYN---------QR---	Chevalier et al., 2011
			SLYN---------QR---	Kaufmann et al., 2004
			SLYN---------QR---	Geels et al., 2006
Gag 228→245	Leuk 9	Central Memory	**MREPRGSDIAGTTSTLQE**	
			----------QIGWM	Boritz et al., 2007
			----------QI	Koeppe et al., 2006
Gag 236→253	Leuk 9	Central Memory	**IAGTTSTLQEQIAWMTSN**	
			PRGSD----------	Boritz et al., 2007
			GSD----------	Koeppe et al., 2006
			---------PPVPVG	Kaushik et al., 2005

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 260→277	Leuk 7	Effector Memory	**DIYKRWIILGLNKIVRMY**	
			PVG----------	Ramduth et al., 2009
			VGE----------SPV	De Groot et al., 2005
			----------	Boaz et al., 2003
			----------	Wilson et al., 2001
			----------SPTS	Vingert et al., 2010
			----------SPTSILD	Rosenberg et al., 1997
			----------SP	Koeppe et al., 2006
			----------SP	Jones et al., 2009
			----------SPTSI	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------SPTSILDIR	Adams et al., 1997
			----------SPVSILDIRQGP	De Groot et al., 2005
Gag 268→285	Leuk 7	Naïve	**LGLNKIVRMYSPVSILDI**	
			----------T-----	Ritchie et al., 2011
			I----------RQGP	De Groot et al., 2005
			PVGDIYKRWI--------	Ramduth et al., 2009
			VGEDIYKRWI----------	De Groot et al., 2005
			YKRWI----------	Boaz et al., 2003
			YKRWI----------	Wilson et al., 2001
			YKRWI----------T-	Vingert et al., 2010
			YKRWI----------T----	Rosenberg et al., 1997
			I----------	Koeppe et al., 2006
			----------	Jones et al., 2009
			----------T--	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------T----R	Adams et al., 1997
Gag 276→293	Leuk 7	Naïve	**MYSPVSILDIKQGPKEPF**	
			ILGLNKIVR----------	De Groot et al., 2005
			LGLNKIVR----------SPTSILD	De Groot et al., 2005
			----------RDYV	Chevalier et al., 2011
			----------RDYVDRFY	Schrier et al., 1989
			--------RDYVDR	Younes et al., 2003
Gag 292→309	Leuk 7	Naïve	**PFRDYVDRFFKTLRAEQA**	
			ILDIRQGPKE---------Y	Schrier et al., 1989
			IRQGPKE--------	Younes et al., 2003
			GPKE---------Y-	Wahren et al., 1989
			GPKE---------Y----	Kaufmann et al., 2004
			QPKE----------	Adams et al., 1997
			E----------S	Boritz et al., 2007

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Supplementary Table 3 (Continued):

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Gag 397→414	Leuk 10	Naïve	**KEGHIARNCRAPRKKGCW**	
			---K----------K	Kaufmann et al., 2004
			---KN----------K	Chevalier et al., 2011
Gag 429→446	Leuk 7	Naïve	**RQANFLGKIWPSHKGRPG**	
			MKDCTE---------	Wahren et al., 1989
			----------	Kaufmann et al., 2004
			----------	Chevalier et al., 2011
			----------NFLQSR	Kaufmann et al., 2004
			----------NFLQSR	Ramduth et al., 2009
			----------NFLQSR	Chevalier et al., 2011
Gag 445→462	Leuk 7	Naïve	**PGNFLQNRPEPTAPPAES**	
	Leuk 7	Naïve	---S----------FRF	Chevalier et al., 2011
	Leuk 10	Naïve	-----S---------	
Gag 453→472	Leuk 7	Naïve	**PEPTAPPAESFRFEETTP**	
			FLQSR-------E-----	Chevalier et al., 2011
			----E-----GEE---PSQK	Kaufmann et al., 2004
Gag 461→480	Leuk 7	Naïve	**ESFRFEETTPAPKQEPKD**	
			TAPPE----------S--	Kaufmann et al., 2004
			----------	Kaushik et al., 2005
Pol 313→330	Leuk 7	Naïve	**AIFQSSMTKILEPFRAQN**	
			SP----------	Wilson et al., 2001
			SP----------	Boaz et al., 2003
Pol 377→394	Leuk 7	Central Memory	**QKEPPFLWMGYELHPDKW**	
Pol 713→730	Leuk 10	Central Memory	**KVLFLDGIDKAQEEHEKY**	
Pol 777→794	Leuk 6	Effector Memory	**QLDCTHLEGKIILVAVHV**	
			----------ASGYI	Fonseca et al., 2006
Pol 945→962	Leuk 7	Central Memory	**SRDPIWKGPAKLLWKGEG**	
Env 1→19	Leuk 9	Central Memory	**MRVRGILRNCQQWWIWGI**	
Env 9→27	Leuk 9	Central Memory	**NCQQWWIWGILGFWMLMI**	
Env 90→107	Leuk 7	Naïve	**TENFNMWKNDMVDQMHED**	
			PQEVVLVNV----------	Geretti et al., 1994
Env 148→165	Leuk 9	Central Memory	**TNTMGEIKNCSFNITTEL**	
	Leuk 9	Central Memory	TNPTSSSWGMMEK-------	Mirano-Bascos et al., 2008
			SSSGRMIMEK----------	Geretti et al., 1994
			MEK----------SIRNK	Mirano-Bascos et al., 2008
			----------SIRGKVQK	Geretti et al., 1994

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Env 156→173	Leuk 9	Central Memory	**NCSFNITTELRDKKQKVY** GEIK-------TSIRG-V-- MEKGEIK----Y---SIRNK K----------I---	 Geretti et al., 1994 Mirano-Bascos et al., 2008 Gaudebout et al., 1997
Env 164→181	Leuk 8	Effector Memory	**ELRDKKQKVYALFYRLDI** ----------VPLTK	 Harari et al., 2008
Env 172→189	Leuk 8	Effector Memory	**VYALFYRLDIVPLNENNS** ----------TK	 Harari et al., 2008
Env 213→230	Leuk 7	Naïve	**IPIHYCAPAGYAILKCNN** P----------F-------K -----P----------F-	 Geretti et al., 1994 Ranasinghe et al., 2012
Env 269→286	Leuk 10	Naïve	**EIIIRSENLTNNAKTIIV** -DIV----F-D-------Q -VV-----F---------QLNES	 Mirano-Bascos et al., 2008 De Groot et al., 2004
Env 277→294	Leuk 7	Naïve	**LTNNAKTIIVHLNESVEI** NF-D----------QIN IRSVNF-D-------Q---T- SANF-D-------Q-	 Malhotra et al., 2003 Geretti et al., 1994 Wahren et al., 1989
Env 285→302	Leuk 7	Naïve	**IVHLNESVEIVCTRPNNN** NF-D----------QIN IRSVNF-D-------Q---T-	 Malhotra et al., 2003 Geretti et al., 1994
Env 334→351	Leuk 7 Leuk 9	Central Memory Central Memory	**SEDKWNKTLQKVSKKLKE**	
Env 342→360	Leuk 7	Central Memory	**LQKVSKKLKEHFPNKTIK**	
Env 411→428	Leuk 7 Leuk 10	Central Memory Naïve	**NSTITLPCRIKQIINMWQ** D----------KVG D----------KVG ----------KVGKA ----------KVGKA	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 419→436	Leuk 7 Leuk 7 Leuk 10	Naïve Central Memory Naïve	**RIKQIINMWQEVGRAMYA** DTITLPC---------- DTITLPC---------- ITLQC----------K- ----------K-	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 435→452	Leuk 7	Naïve	**YAPPIAGNITCKSNITGL**	

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Env 477→494	Leuk 7	Naïve	**DNWRSELYKYKVVEIKPL** GGDMR---------- DMR---------- -------K-----GVAPTKA -------R----GVAPTRAK	 Geretti et al., 1994. Wahren et al., 1989 Geretti et al., 1994. Mirano-Bascos et al., 2008
Env 565→582	Leuk 9 Leuk 8	Effector Memory Central Memory	**MLQLTVWGIKQLQTRVLA** ----------VERYLK QQHL----------	 Malhotra et al., 2003 Wahren et al., 1989
Env 613→630	Leuk 8	Effector Memory	**SWSNKSQEDIWDNMTWMQ**	
Env 669→686	Leuk 7	Naïve	**LWNWFDITNWLWYIKIFI** AS-----N-------	 Wahren et al., 1989
Env 725→742	Leuk 9	Central Memory	**RGPDRLGRIEEEGGEQDR** ---------R--DR	 Schrier et al., 1989
Env 733→750	Leuk 9	Central Memory	**IEEEGGEQDRDRSIRLVS** GR---------R--DR	 Schrier et al., 1989
Env 773→790	Leuk 8	Effector Memory	**DFILIAARAVELLGRSSL**	
Env 790→807	Leuk 6 Leuk 8	Effector Memory Central Memory	**WEALKYLGSLVQYWGLEL** RIVELLGRRG----------KNSAVS RIVELLGRRG----------KNSAVS	 Berzofsky et al., 1991 Berzofsky et al., 1991
Env 798→815	Leuk 6	Effector Memory	**SLVQYWGLELKKSAISLL**	
Env 830→847	Leuk 7	Central Memory	**IELIQRICRAIRNIPRRI**	
Nef 73→90	Leuk 9	Central Memory	**QVPLRPMTYKAAFDLSFF** VGFPVRP---------- ---------H-LKEKGGL	 De Groot et al., 2005. Ranasinghe et al., 2012
Nef 129→146	Leuk 10	Naïve	**PGPGVRYPLTFGWCFKLV** ----------Y---PVEPDKVEEANKG	 Pancré et al., 2007
Rev 17→34	Leuk 9	Central Memory	**RIIKILYQSNPYPKPEGT**	
Rev 49→66	Leuk 10	Central Memory	**QRQIHSISERILSTCLGR** RRRRWRER---------	 Blazevic et al., 1995
Vif 1→18	Leuk 9	Central Memory	**MENRWQVLIVWQVDRMKI** ----------RTWNSLVK	 De Groot et al., 2005.
Vif 81→98	Leuk 10	Central Memory	**LGHGVSIEWRLRRYSTQV** ----------KQ----	 Ranki et al., 1997
Vpu 2 →20	Leuk 8	Central Memory	**ARVDYRLGVGALIVALII**	
Vpu 17→34	Leuk 9	Central Memory	**IIAIVVWTIVYIEYRKLL**	
Vpu 33→50	Leuk 10	Naïve	**LLRQRKIDWLIKRIRERA**	
Vpu 41→58	Leuk 9	Central Memory	**WLIKRIRERAEDSGNESE**	

	Leuk 9	Effector Memory		

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

References

Adams, S.L., R.A. Biti, and G.J. Stewart. 1997. T-cell response to HIV in natural infection: optimized culture conditions for detecting responses to gag peptides. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 15:257–263.

Berzofsky, J.A., C.D. Pendleton, M. Clerici, J. Ahlers, D.R. Lucey, S.D. Putney, and G.M. Shearer. 1991. Peptides containing multideterminant clusters of human immunodeficiency virus envelope induce murine and human T-cell responses in diverse histocompatibility types. *Trans. Assoc. Am. Physicians.* 104:69–77.

Blazevic, V., A. Ranki, and K.J. Krohn. 1995. Helper and cytotoxic T cell responses of HIV type 1-infected individuals to synthetic peptides of HIV type 1 Rev. *AIDS Res. Hum. Retroviruses.* 11:1335–1342.

Boaz, M.J., A. Waters, S. Murad, P.J. Easterbrook, E. D'Sousa, C. van Wheeley, and A. Vyakarnam. 2003. CD4 responses to conserved HIV-1 T helper epitopes show both negative and positive associations with virus load in chronically infected subjects. *Clin. Exp. Immunol.* 134:454–463.

Boritz, E., E.L. Rapaport, T.B. Campbell, J.R. Koeppe, and C.C. Wilson. 2007. $CD4^+$ T cell targeting of human immunodeficiency virus type 1 (HIV-1) peptide sequences present in vivo during chronic, progressive HIV-1 disease. *Virology.* 361:34–44.

Chevalier, M.F., B. Jülg, A. Pyo, M. Flanders, S. Ranasinghe, D.Z. Soghoian, D.S. Kwon, J. Rychert, J. Lian, M.I. Muller, et al. 2011. HIV-1-specific interleukin-21^+ $CD4^+$ T cell responses contribute to durable viral control through the modulation of HIV-specific $CD8^+$ T cell function. *J. Virol.* 85:733–741.

De Groot, A.S., E.A. Bishop, B. Khan, M. Lally, L. Marcon, J. Franco, K.H. Mayer, C.C. Carpenter, and W. Martin. 2004. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods.* 34:476–487.

De Groot, A.S., L. Marcon, E.A. Bishop, D. Rivera, M. Kutzler, D.B. Weiner, and W. Martin. 2005. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine.* 23:2136–2148.

Fonseca, S.G., A. Coutinho-Silva, L.A. Fonseca, A.C. Segurado, S.L. Moraes, H. Rodrigues, J. Hammer, E.G. Kallás, J. Sidney, A. Sette, et al. 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS.* 20:2263–2273.

Gaudebout, P., D. Zeliszewski, J.J. Golvano, C. Pignal, S. Le Gac, F. Borras-Cuesta, and G. Sterkers. 1997. Binding analysis of 95 HIV gp120 peptides to HLA-DR1101 and –DR0401 evidenced many HLA-class II binding regions on gp120 and suggested several promiscuous regions. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 14:91–101.

Geels, M.J., C.A. Jansen, E. Baan, I.M. De Cuyper, G.J. van Schijndel, H. Schuitemaker, J. Goudsmit, G. Pollakis, F. Miedema, W.A. Paxton, and D. van Baarle. 2006. CTL escape and increased viremia irrespective of HIV-specific CD4+ T-helper responses in two HIV-infected individuals. *Virology.* 345:209–219.

Geretti, A.M., C.A. Van Baalen, J.C. Borleffs, C.A. Van Els, and A.D. Osterhaus. 1994. Kinetics and specificities of the T helper-cell response to gp120 in the asymptomatic stage of HIV-1 infection. *Scand. J. Immunol.* 39:355–362.

Harari, A., P.A. Bart, W. Stöhr, G. Tapia, M. Garcia, E. Medjitna-Rais, S. Burnet, C. Cellerai, O. Erlwein, T. Barber, et al. 2008. An HIV-1 clade C DNA prime, NYVAC boost vaccine regimen induces reliable, polyfunctional, and long-lasting T cell responses. *J. Exp. Med.* 205:63–77.

Jones, R.B., F.Y. Yue, X.X. Gu, D.V. Hunter, S. Mujib, G. Gyenes, R.D. Mason, R. Mohamed, K.S. MacDonald, C. Kovacs, and M.A. Ostrowski. 2009. Human immunodeficiency virus type 1 escapes from interleukin-2-producing $CD4^+$ T-cell responses without high-frequency fixation of mutations. *J. Virol.* 83:8722–8732.

Kaufmann, D.E., P.M. Bailey, J. Sidney, B. Wagner, P.J. Norris, M.N. Johnston, L.A. Cosimi, M.M. Addo, M. Lichterfeld, M. Altfeld, et al. 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. *J. Virol.* 78:4463–4477.

Kaushik, S., M. Vajpayee, N. Wig, and P. Seth. 2005. Characterization of HIV-1 Gag-specific T cell responses in chronically infected Indian population. *Clin. Exp. Immunol.* 142:388–397.

Koeppe, J.R., T.B. Campbell, E.L. Rapaport, and C.C. Wilson. 2006. HIV-1-specific $CD4^+$ T-cell responses are not associated with significant viral epitope variation in persons with persistent plasma viremia. *J. Acquir. Immune Defic. Syndr.* 41:140–148.

Koup, R.A., M. Roederer, L. Lamoreaux, J. Fischer, L. Novik, M.C. Nason, B.D. Larkin, M.E. Enama, J.E. Ledgerwood, R.T. Bailer, et al., VRC 010 Study Team. 2010. Priming immunization with DNA augments immunogenicity of recombinant adenoviral vectors for both HIV-1 specific antibody and T-cell responses. *PLoS ONE.* 5:e9015.

Malhotra, U., S. Holte, T. Zhu, E. Delpit, C. Huntsberry, A. Sette, R. Shankarappa, J. Maenza, L. Corey, and M.J. McElrath. 2003. Early induction and maintenance of Env-specific T-helper cells following human immunodeficiency virus type 1 infection. *J. Virol.* 77:2663–2674.

Mathiesen, T., A. Sönnerborg, and B. Wahren. 1989. Detection of antibodies against myelin basic protein and increased levels of HIV-IgG antibodies and HIV antigen after solubilization of immune complexes in sera and CSF of HIV infected patients. *Viral Immunol.* 2:1–9.

Mirano-Bascos, D., M. Tary-Lehmann, and S.J. Landry. 2008. Antigen structure influences helper T-cell epitope dominance in the human immune response to HIV envelope glycoprotein gp120. *Eur. J. Immunol.* 38:1231–1237.

Pancré, V., N. Delhem, Y. Yazdanpanah, A. Delanoye, M. Delacre, S. Depil, O. Moralès, Y. Mouton, and C. Auriault. 2007. Presence of HIV-1 Nef specific CD4 T cell response is associated with non-progression in HIV-1 infection. *Vaccine.* 25:5927–5937.

Ramduth, D., C.L. Day, C.F. Thobakgale, N.P. Mkhwanazi, C. de Pierres, S. Reddy, M. van der Stok, Z. Mncube, K. Nair, E.S. Moodley, et al. 2009. Immunodominant HIV-1 Cd4^{+} T cell epitopes in chronic untreated clade C HIV-1 infection. *PLoS ONE*. 4:e5013.

Ranasinghe, S., M. Flanders, S. Cutler, D.Z. Soghoian, M. Ghebremichael, I. Davis, M. Lindqvist, F. Pereyra, B.D. Walker, D. Heckerman, and H. Streeck. 2012. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol*. 86:277–283.

Ranki, A., J. Suni, V. Blazevic, P. Holmström, S. Mattinen, K. Krohn, and S.L. Valle. 1997. T-cell recognition of HIV antigens in HIV-seroreverted persons. *AIDS*. 11:132–133.

Ritchie, A.J., S.L. Campion, J. Kopycinski, Z. Moodie, Z.M. Wang, K. Pandya, S. Moore, M.K. Liu, S. Brackenridge, K. Kuldanek, K. Legg, M.S. Cohen, E.L. Delwart, B.F. Haynes, S. Fidler, A.J. McMichael, and N. Goonetilleke. 2011. Differences in HIV-specific T cell responses between HIV-exposed and –unexposed HIV-seronegative individuals. *J. Virol*. 85:3507–3516.

Rosenberg, E.S., J.M. Billingsley, A.M. Caliendo, S.L. Boswell, P.E. Sax, S.A. Kalams, and B.D. Walker. 1997. Vigorous HIV-1-specific CD4^{+} T cell responses associated with control of viremia. *Science*. 278:1447–1450.

Schrier, R.D., J.W. Gnann Jr., R. Landes, C. Lockshin, D. Richman, A. McCutchan, C. Kennedy, M.B. Oldstone, and J.A. Nelson. 1989. T cell recognition of HIV synthetic peptides in a natural infection. *J. Immunol*. 142:1166–1176.

Vingert, B., S. Perez-Patrigeon, P. Jeannin, O. Lambotte, F. Boufassa, F. Lemaître, W.W. Kwok, I. Theodorou, J.F. Delfraissy, J. Thèze, and L.A. Chakrabarti; ANRS EP36 HIV Controllers Study Group. 2010. HIV controller CD4^{+} T cells respond to minimal amounts of Gag antigen due to high TCR avidity. *PLoS Pathog*. 6:e1000780.

Wahren, B., J. Rosen, E. Sandström, T. Mathiesen, S. Modrow, and H. Wigzell. 1989. HIV-1 peptides induce a proliferative response in lymphocytes from infected persons. *J. Acquir. Immune Defic. Syndr*. 2:448–456.

Wilson, C.C., B. Palmer, S. Southwood, J. Sidney, Y. Higashimoto, E. Appella, R. Chesnut, A. Sette, and B.D. Livingston. 2001. Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J. Virol*. 75:4195–4207.

Younes, S.A., B. Yassine-Diab, A.R. Dumont, M.R. Boulassel, Z. Grossman, J.P. Routy, and R.P. Sekaly. 2003. HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory CD4^{+} T cells endowed with proliferative capacity. *J. Exp. Med*. 198:1909–1922.

Table S3. HIV-1 specific epitopes mapped in the pre-exposure repertoire were compared against the Los Alamos database to establish whether they had previously been reported in natural infection

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 33→50	Leuk 7	Central Memory	**HLVWASRELERFALNPGL**	
			----------L	Ramduth et al., 2009
			K-I----------V	Chevalier et al., 2011
			K-I----------V	Kaufmann et al., 2004
			-I----------VN	Wahren et al., 1989
			-I----------VN	Koeppe et al., 2006
			-I----------VN	Jones et al., 2009
			-------VN	Kaufmann et al., 2004
			----------	Ranasinghe et al., 2012
Gag 41→58	Leuk 7	Naïve	**LERFALNPGLLETSEGCK**	
			----------	Kaushik et al., 2005
			----------	Ramduth et al., 2009
Gag 57→74	Leuk 7	Central Memory	**CKQIIKQLQPALQTGTEE**	
Gag 73→90	Leuk 9	Central Memory	**EELRSLYNTVATLYCVHE**	
			TGS----------	Chevalier et al., 2011
			TGS----------	Kaufmann et al., 2004
			----------	Jones et al., 2009
			----------	Fonseca et al., 2006
Gag 81→98	Leuk 10	Naïve	**TVATLYCVHEKIEVRDTK**	
			SLYN---------QR---	Chevalier et al., 2011
			SLYN---------QR---	Kaufmann et al., 2004
			SLYN---------QR---	Geels et al., 2006
Gag 228→245	Leuk 9	Central Memory	**MREPRGSDIAGTTSTLQE**	
			----------QIGWM	Boritz et al., 2007
			----------QI	Koeppe et al., 2006
Gag 236→253	Leuk 9	Central Memory	**IAGTTSTLQEQIAWMTSN**	
			PRGSD----------	Boritz et al., 2007
			GSD----------	Koeppe et al., 2006
			---------PPVPVG	Kaushik et al., 2005

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Gag 260→277	Leuk 7	Effector Memory	**DIYKRWIILGLNKIVRMY**	
			PVG----------	Ramduth et al., 2009
			VGE----------SPV	De Groot et al., 2005
			----------	Boaz et al., 2003
			----------	Wilson et al., 2001
			----------SPTS	Vingert et al., 2010
			----------SPTSILD	Rosenberg et al., 1997
			----------SP	Koeppe et al., 2006
			----------SP	Jones et al., 2009
			----------SPTSI	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------SPTSILDIR	Adams et al., 1997
			----------SPVSILDIRQGP	De Groot et al., 2005
Gag 268→285	Leuk 7	Naïve	**LGLNKIVRMYSPVSILDI**	
			----------T-----	Ritchie et al., 2011
			I----------RQGP	De Groot et al., 2005
			PVGDIYKRWI--------	Ramduth et al., 2009
			VGEDIYKRWI----------	De Groot et al., 2005
			YKRWI----------	Boaz et al., 2003
			YKRWI----------	Wilson et al., 2001
			YKRWI----------T-	Vingert et al., 2010
			YKRWI----------T----	Rosenberg et al., 1997
			I----------	Koeppe et al., 2006
			----------	Jones et al., 2009
			----------T--	Kaufmann et al., 2004
			----------	Kaufmann et al., 2004
			----------T----R	Adams et al., 1997
Gag 276→293	Leuk 7	Naïve	**MYSPVSILDIKQGPKEPF**	
			ILGLNKIVR----------	De Groot et al., 2005
			LGLNKIVR----------SPTSILD	De Groot et al., 2005
			----------RDYV	Chevalier et al., 2011
			----------RDYVDRFY	Schrier et al., 1989
			--------RDYVDR	Younes et al., 2003
Gag 292→309	Leuk 7	Naïve	**PFRDYVDRFFKTLRAEQA**	
			ILDIRQGPKE---------Y	Schrier et al., 1989
			IRQGPKE--------	Younes et al., 2003
			GPKE---------Y-	Wahren et al., 1989
			GPKE---------Y----	Kaufmann et al., 2004
			QPKE----------	Adams et al., 1997
			E----------S	Boritz et al., 2007

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Supplementary Table 3 (Continued):

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Gag 397→414	Leuk 10	Naïve	**KEGHIARNCRAPRKKGCW**	
			---K----------K	Kaufmann et al., 2004
			---KN----------K	Chevalier et al., 2011
Gag 429→446	Leuk 7	Naïve	**RQANFLGKIWPSHKGRPG**	
			MKDCTE---------	Wahren et al., 1989
			----------	Kaufmann et al., 2004
			----------	Chevalier et al., 2011
			----------NFLQSR	Kaufmann et al., 2004
			----------NFLQSR	Ramduth et al., 2009
			----------NFLQSR	Chevalier et al., 2011
Gag 445→462	Leuk 7	Naïve	**PGNFLQNRPEPTAPPAES**	
	Leuk 7	Naïve	---S----------FRF	Chevalier et al., 2011
	Leuk 10	Naïve	-----S---------	
Gag 453→472	Leuk 7	Naïve	**PEPTAPPAESFRFEETTP**	
			FLQSR-------E-----	Chevalier et al., 2011
			----E-----GEE---PSQK	Kaufmann et al., 2004
Gag 461→480	Leuk 7	Naïve	**ESFRFEETTPAPKQEPKD**	
			TAPPE----------S--	Kaufmann et al., 2004
			----------	Kaushik et al., 2005
Pol 313→330	Leuk 7	Naïve	**AIFQSSMTKILEPFRAQN**	
			SP----------	Wilson et al., 2001
			SP----------	Boaz et al., 2003
Pol 377→394	Leuk 7	Central Memory	**QKEPPFLWMGYELHPDKW**	
Pol 713→730	Leuk 10	Central Memory	**KVLFLDGIDKAQEEHEKY**	
Pol 777→794	Leuk 6	Effector Memory	**QLDCTHLEGKIILVAVHV**	
			----------ASGYI	Fonseca et al., 2006
Pol 945→962	Leuk 7	Central Memory	**SRDPIWKGPAKLLWKGEG**	
Env 1→19	Leuk 9	Central Memory	**MRVRGILRNCQQWWIWGI**	
Env 9→27	Leuk 9	Central Memory	**NCQQWWIWGILGFWMLMI**	
Env 90→107	Leuk 7	Naïve	**TENFNMWKNDMVDQMHED**	
			PQEVVLVNV----------	Geretti et al., 1994
Env 148→165	Leuk 9	Central Memory	**TNTMGEIKNCSFNITTEL**	
	Leuk 9	Central Memory	TNPTSSSWGMMEK-------	Mirano-Bascos et al., 2008
			SSSGRMIMEK----------	Geretti et al., 1994
			MEK----------SIRNK	Mirano-Bascos et al., 2008
			----------SIRGKVQK	Geretti et al., 1994

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	$CD4^+$ T cell subset	Sequence	Reference
Env 156→173	Leuk 9	Central Memory	**NCSFNITTELRDKKQKVY** GEIK-------TSIRG-V-- MEKGEIK----Y---SIRNK K----------I---	 Geretti et al., 1994 Mirano-Bascos et al., 2008 Gaudebout et al., 1997
Env 164→181	Leuk 8	Effector Memory	**ELRDKKQKVYALFYRLDI** ----------VPLTK	 Harari et al., 2008
Env 172→189	Leuk 8	Effector Memory	**VYALFYRLDIVPLNENNS** ----------TK	 Harari et al., 2008
Env 213→230	Leuk 7	Naïve	**IPIHYCAPAGYAILKCNN** P----------F-------K -----P----------F-	 Geretti et al., 1994 Ranasinghe et al., 2012
Env 269→286	Leuk 10	Naïve	**EIIIRSENLTNNAKTIIV** -DIV----F-D-------Q -VV-----F---------QLNES	 Mirano-Bascos et al., 2008 De Groot et al., 2004
Env 277→294	Leuk 7	Naïve	**LTNNAKTIIVHLNESVEI** NF-D----------QIN IRSVNF-D-------Q---T- SANF-D-------Q-	 Malhotra et al., 2003 Geretti et al., 1994 Wahren et al., 1989
Env 285→302	Leuk 7	Naïve	**IVHLNESVEIVCTRPNNN** NF-D----------QIN IRSVNF-D-------Q---T-	 Malhotra et al., 2003 Geretti et al., 1994
Env 334→351	Leuk 7 Leuk 9	Central Memory Central Memory	**SEDKWNKTLQKVSKKLKE**	
Env 342→360	Leuk 7	Central Memory	**LQKVSKKLKEHFPNKTIK**	
Env 411→428	Leuk 7 Leuk 10	Central Memory Naïve	**NSTITLPCRIKQIINMWQ** D----------KVG D----------KVG ----------KVGKA ----------KVGKA	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 419→436	Leuk 7 Leuk 7 Leuk 10	Naïve Central Memory Naïve	**RIKQIINMWQEVGRAMYA** DTITLPC---------- DTITLPC---------- ITLQC----------K- ----------K-	 Geretti et al., 1994. Harari et al., 2008 Mirano-Bascos et al., 2008 Koup et al., 2010
Env 435→452	Leuk 7	Naïve	**YAPPIAGNITCKSNITGL**	

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

Table S3 (Continued):

HBX2 location	Subject ID	CD4$^+$ T cell subset	Sequence	Reference
Env 477→494	Leuk 7	Naïve	**DNWRSELYKYKVVEIKPL** GGDMR---------- DMR---------- -------K-----GVAPTKA -------R----GVAPTRAK	 Geretti et al., 1994. Wahren et al., 1989 Geretti et al., 1994. Mirano-Bascos et al., 2008
Env 565→582	Leuk 9 Leuk 8	Effector Memory Central Memory	**MLQLTVWGIKQLQTRVLA** ----------VERYLK QQHL----------	 Malhotra et al., 2003 Wahren et al., 1989
Env 613→630	Leuk 8	Effector Memory	**SWSNKSQEDIWDNMTWMQ**	
Env 669→686	Leuk 7	Naïve	**LWNWFDITNWLWYIKIFI** AS-----N-------	 Wahren et al., 1989
Env 725→742	Leuk 9	Central Memory	**RGPDRLGRIEEEGGEQDR** ---------R--DR	 Schrier et al., 1989
Env 733→750	Leuk 9	Central Memory	**IEEEGGEQDRDRSIRLVS** GR---------R--DR	 Schrier et al., 1989
Env 773→790	Leuk 8	Effector Memory	**DFILIAARAVELLGRSSL**	
Env 790→807	Leuk 6 Leuk 8	Effector Memory Central Memory	**WEALKYLGSLVQYWGLEL** RIVELLGRRG----------KNSAVS RIVELLGRRG----------KNSAVS	 Berzofsky et al., 1991 Berzofsky et al., 1991
Env 798→815	Leuk 6	Effector Memory	**SLVQYWGLELKKSAISLL**	
Env 830→847	Leuk 7	Central Memory	**IELIQRICRAIRNIPRRI**	
Nef 73→90	Leuk 9	Central Memory	**QVPLRPMTYKAAFDLSFF** VGFPVRP---------- ---------H-LKEKGGL	 De Groot et al., 2005. Ranasinghe et al., 2012
Nef 129→146	Leuk 10	Naïve	**PGPGVRYPLTFGWCFKLV** ----------Y---PVEPDKVEEANKG	 Pancré et al., 2007
Rev 17→34	Leuk 9	Central Memory	**RIIKILYQSNPYPKPEGT**	
Rev 49→66	Leuk 10	Central Memory	**QRQIHSISERILSTCLGR** RRRRWRER---------	 Blazevic et al., 1995
Vif 1→18	Leuk 9	Central Memory	**MENRWQVLIVWQVDRMKI** ----------RTWNSLVK	 De Groot et al., 2005.
Vif 81→98	Leuk 10	Central Memory	**LGHGVSIEWRLRRYSTQV** ----------KQ----	 Ranki et al., 1997
Vpu 2 →20	Leuk 8	Central Memory	**ARVDYRLGVGALIVALII**	
Vpu 17→34	Leuk 9	Central Memory	**IIAIVVWTIVYIEYRKLL**	
Vpu 33→50	Leuk 10	Naïve	**LLRQRKIDWLIKRIRERA**	
Vpu 41→58	Leuk 9	Central Memory	**WLIKRIRERAEDSGNESE**	

	Leuk 9	Effector Memory		

Sequence identified in T cell library assay is presented in bold font. Previously reported epitopes containing shared sequence homology (represented by dashed line (-)) to mapped epitopes are displayed with accompanying literature reference. Gray highlighting indicates peptide epitopes that have not been identified before.

References

Adams, S.L., R.A. Biti, and G.J. Stewart. 1997. T-cell response to HIV in natural infection: optimized culture conditions for detecting responses to gag peptides. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 15:257–263.

Berzofsky, J.A., C.D. Pendleton, M. Clerici, J. Ahlers, D.R. Lucey, S.D. Putney, and G.M. Shearer. 1991. Peptides containing multideterminant clusters of human immunodeficiency virus envelope induce murine and human T-cell responses in diverse histocompatibility types. *Trans. Assoc. Am. Physicians.* 104:69–77.

Blazevic, V., A. Ranki, and K.J. Krohn. 1995. Helper and cytotoxic T cell responses of HIV type 1-infected individuals to synthetic peptides of HIV type 1 Rev. *AIDS Res. Hum. Retroviruses.* 11:1335–1342.

Boaz, M.J., A. Waters, S. Murad, P.J. Easterbrook, E. D'Sousa, C. van Wheeley, and A. Vyakarnam. 2003. CD4 responses to conserved HIV-1 T helper epitopes show both negative and positive associations with virus load in chronically infected subjects. *Clin. Exp. Immunol.* 134:454–463.

Boritz, E., E.L. Rapaport, T.B. Campbell, J.R. Koeppe, and C.C. Wilson. 2007. $CD4^+$ T cell targeting of human immunodeficiency virus type 1 (HIV-1) peptide sequences present in vivo during chronic, progressive HIV-1 disease. *Virology.* 361:34–44.

Chevalier, M.F., B. Jülg, A. Pyo, M. Flanders, S. Ranasinghe, D.Z. Soghoian, D.S. Kwon, J. Rychert, J. Lian, M.I. Muller, et al. 2011. HIV-1-specific interleukin-21^+ $CD4^+$ T cell responses contribute to durable viral control through the modulation of HIV-specific $CD8^+$ T cell function. *J. Virol.* 85:733–741.

De Groot, A.S., E.A. Bishop, B. Khan, M. Lally, L. Marcon, J. Franco, K.H. Mayer, C.C. Carpenter, and W. Martin. 2004. Engineering immunogenic consensus T helper epitopes for a cross-clade HIV vaccine. *Methods.* 34:476–487.

De Groot, A.S., L. Marcon, E.A. Bishop, D. Rivera, M. Kutzler, D.B. Weiner, and W. Martin. 2005. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine.* 23:2136–2148.

Fonseca, S.G., A. Coutinho-Silva, L.A. Fonseca, A.C. Segurado, S.L. Moraes, H. Rodrigues, J. Hammer, E.G. Kallás, J. Sidney, A. Sette, et al. 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS.* 20:2263–2273.

Gaudebout, P., D. Zeliszewski, J.J. Golvano, C. Pignal, S. Le Gac, F. Borras-Cuesta, and G. Sterkers. 1997. Binding analysis of 95 HIV gp120 peptides to HLA-DR1101 and –DR0401 evidenced many HLA-class II binding regions on gp120 and suggested several promiscuous regions. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 14:91–101.

Geels, M.J., C.A. Jansen, E. Baan, I.M. De Cuyper, G.J. van Schijndel, H. Schuitemaker, J. Goudsmit, G. Pollakis, F. Miedema, W.A. Paxton, and D. van Baarle. 2006. CTL escape and increased viremia irrespective of HIV-specific CD4+ T-helper responses in two HIV-infected individuals. *Virology.* 345:209–219.

Geretti, A.M., C.A. Van Baalen, J.C. Borleffs, C.A. Van Els, and A.D. Osterhaus. 1994. Kinetics and specificities of the T helper-cell response to gp120 in the asymptomatic stage of HIV-1 infection. *Scand. J. Immunol.* 39:355–362.

Harari, A., P.A. Bart, W. Stöhr, G. Tapia, M. Garcia, E. Medjitna-Rais, S. Burnet, C. Cellerai, O. Erlwein, T. Barber, et al. 2008. An HIV-1 clade C DNA prime, NYVAC boost vaccine regimen induces reliable, polyfunctional, and long-lasting T cell responses. *J. Exp. Med.* 205:63–77.

Jones, R.B., F.Y. Yue, X.X. Gu, D.V. Hunter, S. Mujib, G. Gyenes, R.D. Mason, R. Mohamed, K.S. MacDonald, C. Kovacs, and M.A. Ostrowski. 2009. Human immunodeficiency virus type 1 escapes from interleukin-2-producing $CD4^+$ T-cell responses without high-frequency fixation of mutations. *J. Virol.* 83:8722–8732.

Kaufmann, D.E., P.M. Bailey, J. Sidney, B. Wagner, P.J. Norris, M.N. Johnston, L.A. Cosimi, M.M. Addo, M. Lichterfeld, M. Altfeld, et al. 2004. Comprehensive analysis of human immunodeficiency virus type 1-specific CD4 responses reveals marked immunodominance of gag and nef and the presence of broadly recognized peptides. *J. Virol.* 78:4463–4477.

Kaushik, S., M. Vajpayee, N. Wig, and P. Seth. 2005. Characterization of HIV-1 Gag-specific T cell responses in chronically infected Indian population. *Clin. Exp. Immunol.* 142:388–397.

Koeppe, J.R., T.B. Campbell, E.L. Rapaport, and C.C. Wilson. 2006. HIV-1-specific $CD4^+$ T-cell responses are not associated with significant viral epitope variation in persons with persistent plasma viremia. *J. Acquir. Immune Defic. Syndr.* 41:140–148.

Koup, R.A., M. Roederer, L. Lamoreaux, J. Fischer, L. Novik, M.C. Nason, B.D. Larkin, M.E. Enama, J.E. Ledgerwood, R.T. Bailer, et al., VRC 010 Study Team. 2010. Priming immunization with DNA augments immunogenicity of recombinant adenoviral vectors for both HIV-1 specific antibody and T-cell responses. *PLoS ONE.* 5:e9015.

Malhotra, U., S. Holte, T. Zhu, E. Delpit, C. Huntsberry, A. Sette, R. Shankarappa, J. Maenza, L. Corey, and M.J. McElrath. 2003. Early induction and maintenance of Env-specific T-helper cells following human immunodeficiency virus type 1 infection. *J. Virol.* 77:2663–2674.

Mathiesen, T., A. Sönnerborg, and B. Wahren. 1989. Detection of antibodies against myelin basic protein and increased levels of HIV-IgG antibodies and HIV antigen after solubilization of immune complexes in sera and CSF of HIV infected patients. *Viral Immunol.* 2:1–9.

Mirano-Bascos, D., M. Tary-Lehmann, and S.J. Landry. 2008. Antigen structure influences helper T-cell epitope dominance in the human immune response to HIV envelope glycoprotein gp120. *Eur. J. Immunol.* 38:1231–1237.

Pancré, V., N. Delhem, Y. Yazdanpanah, A. Delanoye, M. Delacre, S. Depil, O. Moralès, Y. Mouton, and C. Auriault. 2007. Presence of HIV-1 Nef specific CD4 T cell response is associated with non-progression in HIV-1 infection. *Vaccine.* 25:5927–5937.

Ramduth, D., C.L. Day, C.F. Thobakgale, N.P. Mkhwanazi, C. de Pierres, S. Reddy, M. van der Stok, Z. Mncube, K. Nair, E.S. Moodley, et al. 2009. Immunodominant HIV-1 Cd4^{+} T cell epitopes in chronic untreated clade C HIV-1 infection. *PLoS ONE*. 4:e5013.

Ranasinghe, S., M. Flanders, S. Cutler, D.Z. Soghoian, M. Ghebremichael, I. Davis, M. Lindqvist, F. Pereyra, B.D. Walker, D. Heckerman, and H. Streeck. 2012. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol*. 86:277–283.

Ranki, A., J. Suni, V. Blazevic, P. Holmström, S. Mattinen, K. Krohn, and S.L. Valle. 1997. T-cell recognition of HIV antigens in HIV-seroreverted persons. *AIDS*. 11:132–133.

Ritchie, A.J., S.L. Campion, J. Kopycinski, Z. Moodie, Z.M. Wang, K. Pandya, S. Moore, M.K. Liu, S. Brackenridge, K. Kuldanek, K. Legg, M.S. Cohen, E.L. Delwart, B.F. Haynes, S. Fidler, A.J. McMichael, and N. Goonetilleke. 2011. Differences in HIV-specific T cell responses between HIV-exposed and –unexposed HIV-seronegative individuals. *J. Virol*. 85:3507–3516.

Rosenberg, E.S., J.M. Billingsley, A.M. Caliendo, S.L. Boswell, P.E. Sax, S.A. Kalams, and B.D. Walker. 1997. Vigorous HIV-1-specific CD4^{+} T cell responses associated with control of viremia. *Science*. 278:1447–1450.

Schrier, R.D., J.W. Gnann Jr., R. Landes, C. Lockshin, D. Richman, A. McCutchan, C. Kennedy, M.B. Oldstone, and J.A. Nelson. 1989. T cell recognition of HIV synthetic peptides in a natural infection. *J. Immunol*. 142:1166–1176.

Vingert, B., S. Perez-Patrigeon, P. Jeannin, O. Lambotte, F. Boufassa, F. Lemaître, W.W. Kwok, I. Theodorou, J.F. Delfraissy, J. Thèze, and L.A. Chakrabarti; ANRS EP36 HIV Controllers Study Group. 2010. HIV controller CD4^{+} T cells respond to minimal amounts of Gag antigen due to high TCR avidity. *PLoS Pathog*. 6:e1000780.

Wahren, B., J. Rosen, E. Sandström, T. Mathiesen, S. Modrow, and H. Wigzell. 1989. HIV-1 peptides induce a proliferative response in lymphocytes from infected persons. *J. Acquir. Immune Defic. Syndr*. 2:448–456.

Wilson, C.C., B. Palmer, S. Southwood, J. Sidney, Y. Higashimoto, E. Appella, R. Chesnut, A. Sette, and B.D. Livingston. 2001. Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J. Virol*. 75:4195–4207.

Younes, S.A., B. Yassine-Diab, A.R. Dumont, M.R. Boulassel, Z. Grossman, J.P. Routy, and R.P. Sekaly. 2003. HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory CD4^{+} T cells endowed with proliferative capacity. *J. Exp. Med*. 198:1909–1922.