nature medicine

# CD8+ T-cell responses to different HIV proteins have discordant associations with viral load

Photini Kiepiela[1], Kholiswa Ngumbela[1], Christina Thobakgale[1], Dhanwanthie Ramduth[1], Isobella Honeyborne[2], Eshia Moodley[1], Shabashini Reddy[1], Chantal de Pierres[1], Zenele Mncube[1], Nompumelelo Mkhwanazi[1], Karen Bishop[1], Mary van der Stok[1], Kriebashnie Nair[1], Nasreen Khan[1], Hayley Crawford[2], Rebecca Payne[2], Alasdair Leslie[2], Julia Prado[2], Andrew Prendergast[2], John Frater[2], Noel McCarthy[3], Christian Brander[4], Gerald H Learn[5], David Nickle[5], Christine Rousseau[5], Hoosen Coovadia[1], James I Mullins[5], David Heckerman[6], Bruce D Walker[1,4,7] & Philip Goulder[1,2,4]

**Selection of T-cell vaccine antigens for chronic persistent viral infections has been largely empirical. To define the relationship, at the population level, between the specificity of the cellular immune response and viral control for a relevant human pathogen, we performed a comprehensive analysis of the 160 dominant CD8+ T-cell responses in 578 untreated HIV-infected individuals from KwaZulu-Natal, South Africa. Of the HIV proteins targeted, only Gag-specific responses were associated with lowering viremia. Env-specific and Accessory/Regulatory protein–specific responses were associated with higher viremia. Increasing breadth of Gag-specific responses was associated with decreasing viremia and increasing Env breadth with increasing viremia. Association of the specific CD8+ T-cell response with low viremia was independent of *HLA* type and unrelated to epitope sequence conservation. These population-based data, suggesting the existence of both effective immune responses and responses lacking demonstrable biological impact in chronic HIV infection, are of relevance to HIV vaccine design and evaluation.**



HIV vaccine strategies focused on the generation of virus-specific T-cell responses have the goal of protecting from disease progression subsequent to infection, as sterilizing immunity through the induction of neutralizing antibodies is not currently feasible[1]. However, although CD8+ T cells play a central role in immune control of HIV (ref. 2), high-frequency responses are also detectable in individuals progressing to AIDS (ref. 3), suggesting that CD8+ T cells differ in their ability to inhibit viral replication.

To determine the specificity and breadth of CD8+ T-cell responses associated with immune control, we performed high-resolution *HLA* typing, plasma viral load quantitation, and comprehensive mapping of CD8+ T-cell responses to all expressed HIV proteins, in a large cohort of untreated individuals in KwaZulu-Natal, South Africa, where infection rates are among the highest in the world. For each individual, recognition by peripheral blood mononuclear cells (PBMC) of a panel of 410 overlapping 18-mer peptides spanning the entire C-clade consensus sequence was assessed in interferon (IFN)-γ enzyme-linked immunosorbent spot (ELISPOT) assays that preferentially detect CD8+ T-cell responses[4].

## RESULTS

We identified 160 significant ($P < 0.05$ in each case) associations between the recognition of a particular peptide and the expression of a specific *HLA* class I allele, based on the analysis of all 67 alleles present in this cohort at a phenotypic frequency of >0.5% (**Fig. 1a**, **Supplementary Table 1** online). These responses, defined through strong statistical associations after correction for multiple comparisons[5], included exactly 100 in which the optimal epitope and *HLA* restriction was experimentally defined (http://www.hiv.lanl.gov, and data not shown); in no case was there any discrepancy between the predicted and experimentally defined *HLA* restriction.

### Gag associated with low viremia, Env with high viremia

For the initial analysis of associations of *HLA* alleles with immune control, we determined the dominant epitope targeted for each allele and then compared the viral load in responders and nonresponders expressing the designated allele. Two *HLA* class I alleles, *HLA-B*57* and *B*5801*, have previously been associated with low viremia, and two, *HLA-B*18* and *B*5802*, with high viremia (ref. 5 and **Fig. 1b**).

[1]HIV Pathogenesis Programme, Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban 4013, South Africa. [2]Department of Paediatrics, Nuffield Department of Medicine, and [3]Department of Zoology, Peter Medawar Building, South Parks Road, Oxford OX1 3SY, UK. [4]Partners AIDS Research Center, Massachusetts General Hospital, 13th Street, Building 149, Charlestown, Boston, Massachusetts 02129, USA. [5]Department of Microbiology, University of Washington School of Medicine, Seattle, Washington 98195, USA. [6]Miscrosoft Research, One Microsoft Way, Redmond 98052, Washington, USA. [7]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, Maryland 20815, USA. Correspondence should be addressed to P.G. (philip.goulder@paediatrics.ox.ac.uk).

Received 8 August; accepted 8 November; published online 17 December 2006; doi:10.1038/nm1520

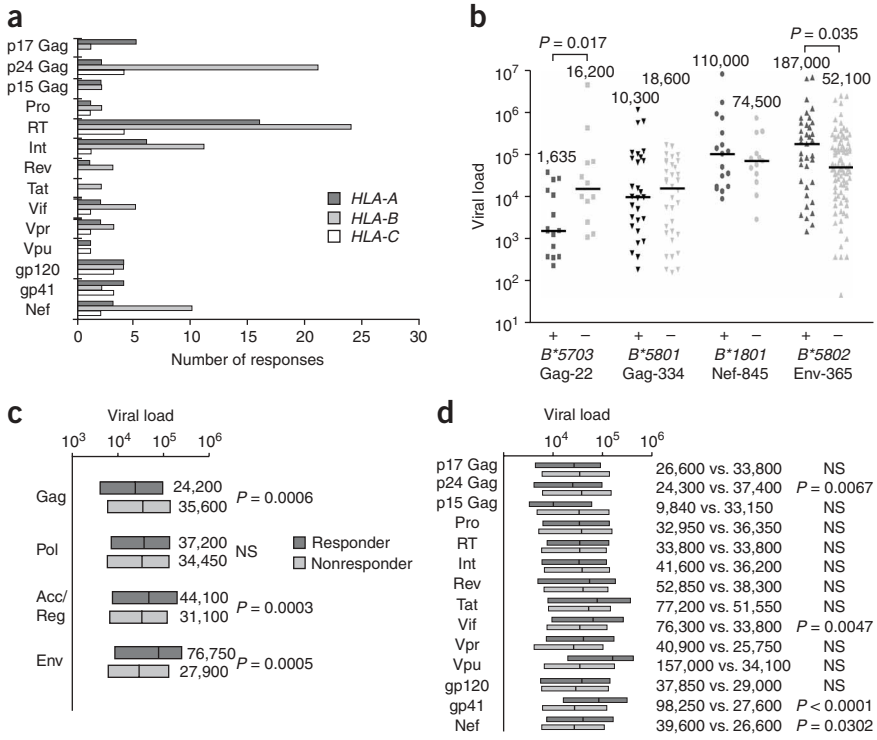


**Figure 1** Location of the 160 dominant CD8+ T-cell epitopes targeted in the study population, and viral load associations of responses to these epitopes. (**a**) Distribution of the 160 HIV-specific CD8+ T-cell responses, in terms of HIV protein and *HLA* restriction. (**b**) Viral loads of responder and nonresponder individuals to the dominant epitope presented by *HLA-B*5703*, *B*5801*, *B*1801* and *B*5802*. Horizontal lines indicate median viral loads. Significant differences are indicated (Mann-Whitney test). (**c**,**d**) Viral loads (interquartile range and median value) of responders and nonresponders to (**c**) epitopes identified in Gag, Pol, Env and the accessory and regulatory proteins Rev, Tat, Vif, Vpr, Vpu and Nef ('Acc/Reg'), and (**d**) individual HIV proteins.

In the case of *HLA-B*5703*, the most prevalent subtype of *HLA-B*57* in this population, the dominant response was directed toward the p24 Gag 18-mer peptide WVKVIEEKAFSPEVIPMF, which contains the previously defined *B*5703*-restricted optimal epitope KAFSPEVIPMF (Gag residues 162–172). We observed recognition of this peptide in 14 *B*5703*-positive individuals and nonrecognition in 12 *B*5703*-positive individuals (median viral load, in terms of HIV RNA copies per ml, was 1,635 and 16,200, respectively; $P = 0.017$, Mann-Whitney test). We observed a similar trend for the dominant *B*5801*-restricted response, also directed against a p24 Gag epitope (TSTLQEQIAW, Gag 240–249). For the dominant *B*18*- and *B*5802*-restricted epitopes, we observed the opposite trend: responders to these epitopes (YPLTFGWCF, Nef 136–144; and QTRVLAIERYL, Env gp41 residues 577–587; respectively) had higher viral loads than nonresponders expressing these alleles.

We expanded this analysis to all 160 epitopes, including a total of 37 Gag, 66 Pol, 20 Env and 37 epitopes within the accessory and regulatory proteins Rev, Tat, Vif, Vpr, Vpu and Nef (collectively termed 'Acc/Reg' proteins; **Supplementary Fig. 1** and **Supplementary Table 1** online). Grouping the CD8+ T-cell responses to individual proteins (**Fig. 1c**) demonstrated that Gag-specific responses were associated with a significant reduction in viremia ($P = 0.0006$, Mann-Whitney), whereas Env-specific and Acc/Reg-specific responses were associated with significantly increased viral loads ($P = 0.0005$ and $P = 0.0003$, respectively). The same data analyzed according to individual proteins targeted (**Fig. 1d**) indicated that the impact of Gag-specific responses on enhanced control was predominantly mediated by p24-specific responses. For responses associated with higher viral loads, the main effects were associated with targeting of gp41, Vif and Nef. Together, these data indicate that most CD8+ T-cell responses are not associated with effective control of HIV in this setting of chronic infection.

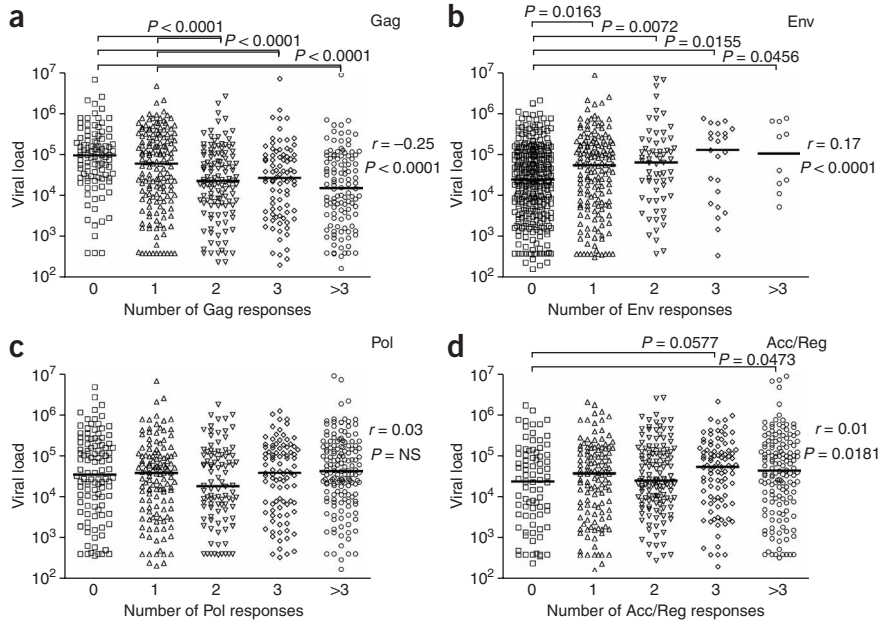


**Figure 2** Breadth of protein-specific CD8+ T-cell responses in relation to viral load. (**a**–**d**) Viral load of individuals in terms of the number of Gag-, Env-, Pol- and Acc/Reg-specific responses detected.

## Only Gag breadth associated with decreasing viremia

Given the discordant association between Gag- and Env-specific CD8+ T-cell responses and viral load, we next assessed the influence of the number of epitopes targeted within these proteins on viral load (**Fig. 2**). Targeting a single Gag epitope was not significantly better than targeting no Gag epitopes ($P > 0.05$), but viral loads were significantly lower when there was >1 Gag-specific CD8+

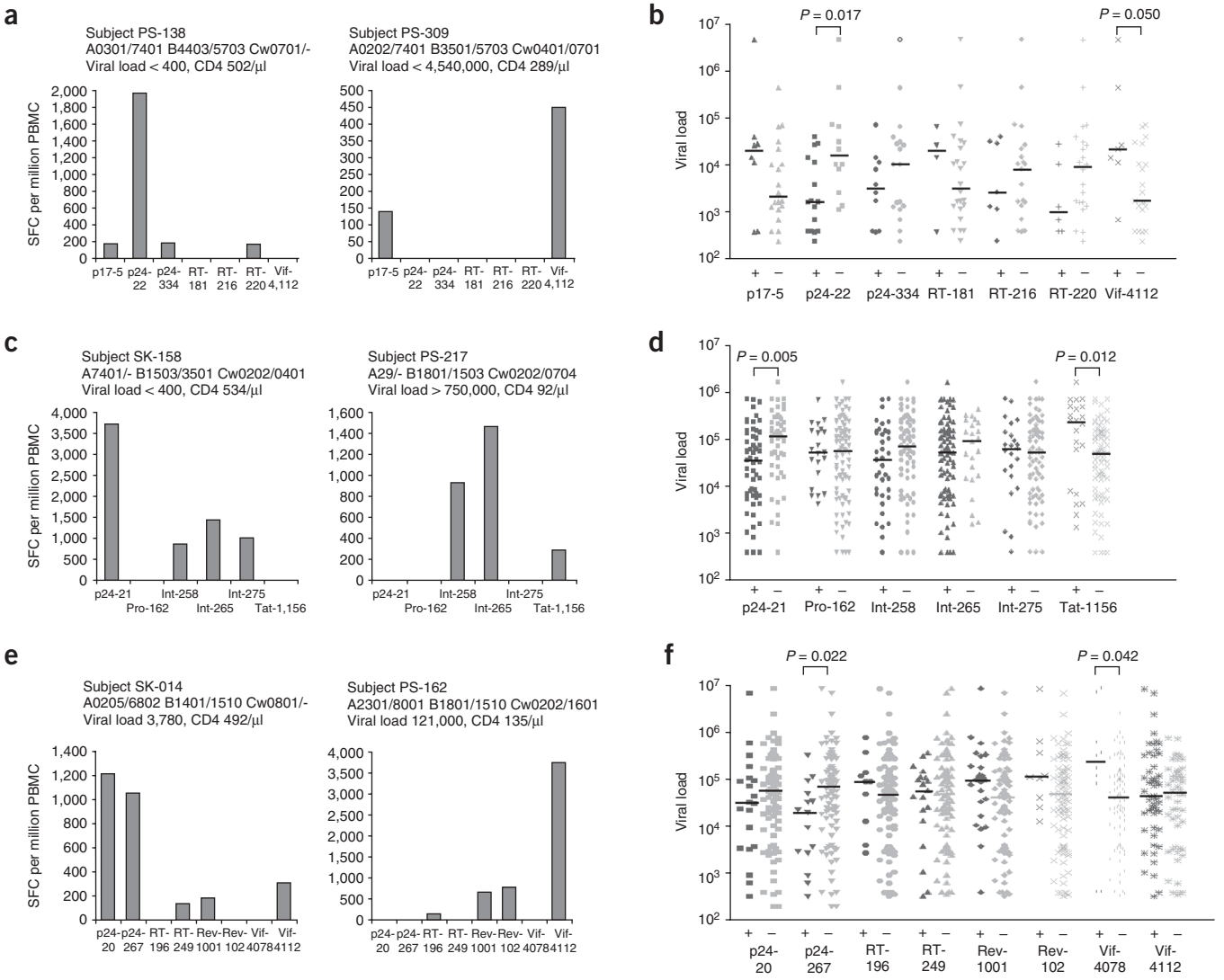


**Figure 3** Comparison of viral loads in responders and *HLA*-matched nonresponders to epitopes in Gag and non-Gag proteins. (**a**) *HLA-B*5703*–restricted responses in two selected *B*5703*-positive individuals showing full HLA class I type. (**b**) Viral loads of responders and nonresponders to *HLA-B*5703*–restricted epitopes. (**c**–**f**) As for **a**,**b** but showing data from *B*1503*- and *B*1510*-positive study subjects and for *B*1503*- and *B*1510*-restricted epitopes, respectively. Significant differences between responders and nonresponders for each epitope are shown (Mann-Whitney test).

T-cell response compared to no Gag-specific $CD8^+$ T-cell responses (median viral load: 21,900 and 94,400, respectively; $P < 0.0001$). In contrast, the number of Env-specific responses was directly correlated with viral load ($r = 0.17$, $P < 0.0001$). These data illustrate that increasing breadth of the HIV-specific $CD8^+$ T-cell response is not associated with the control of viremia, as, with the exception of Gag- and Pol-specific responses, the opposite is the case.

The discordance of the associations between the breadth of protein-specific anti-HIV $CD8^+$ T-cell response and viral load was somewhat more marked when we took into account that the numbers of responses to the different proteins were positively correlated (**Supplementary Table 2** online). Thus, individuals with several Gag responses also had Env responses, which served to dampen the association of Gag with low viremia. We therefore conducted a multivariate linear regression to estimate the independent associations between breadth of response for each protein and viral load. These analyses indicated, on average, a 0.21-$\log_{10}$ drop in viral load for each additional Gag response.

### Protein-specific viremia associations are *HLA* independent

We next assessed whether *HLA* associations with extremes of viral load, such as were observed for *B*5703*, *B*5801*, *B*18* and *B*5802*, are dependent on the protein targeted or on the allele itself. For individuals expressing *B*5703*, a total of seven epitopes were defined in four proteins (p17 Gag, p24 Gag, RT and Vif) but only the Vif-associated epitope was significantly ($P = 0.05$) targeted in association with high viral loads (**Fig. 3a**,**b**).

For the two alleles significantly[5] associated with high viremia, *B*18* and *B*5802* (median viral loads: 84,900 and 63,700, respectively; cohort median viral load: 35,250), no Gag-specific epitopes were significantly targeted in this population. However, *B*1503* and *B*1510*, tending to an association with high viremia (median viral loads: 58,550 and 51,375) have subdominant p24 Gag–specific responses. In these cases, targeting Gag-specific epitopes was associated with lower viral loads, whereas Vif- and Tat-specific responses were associated with higher viral loads (**Fig. 3c**–**f**). These data suggest that for individuals expressing a given *HLA* allele,

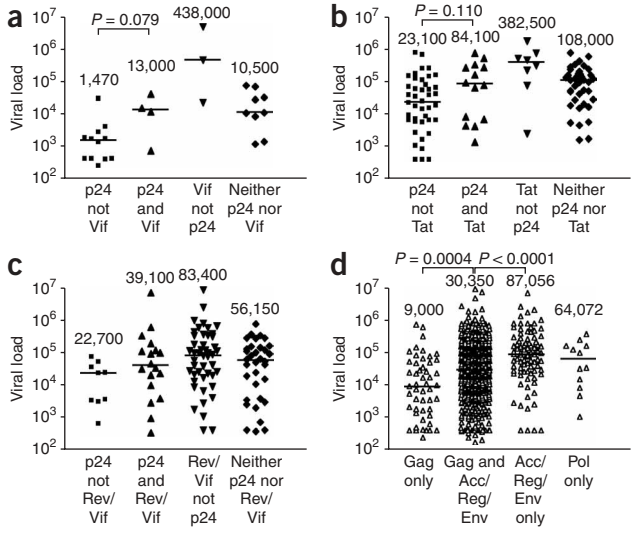


**Figure 4** Comparison of viral loads in study subjects with responses to only Gag, to only Acc/Reg/Env protein epitopes, to both or to neither. (**a**) Viral loads of *B*5703*-positive individuals. The HIV proteins targeted were the p24 Gag epitopes alone; the p24 Gag epitopes and the Vif epitope; the Vif epitope alone; and neither the p24 Gag nor the Vif epitopes. (**b**) As in **a** but for *B*1503*, comparing responses to the p24 Gag epitope alone; the p24 Gag epitope and the Tat epitope; the Tat epitope alone; and neither. (**c**) As in **a** but for *B*1510*, comparing responses to the p24 Gag epitopes alone; the p24 Gag epitopes and the Rev/Vif epitopes; the Rev/Vif epitopes alone; and neither. (**d**) As in **a**–**c** but comparing responses to only Gag, to only Acc/Reg/Env, to both or to neither.

viral load is primarily related to the protein targeted as opposed to the *HLA* allele itself.

### Gag-only CD8$^+$ responses are associated with lower viremia

We next investigated whether CD8$^+$ T-cell activity toward Env and Acc/Reg proteins might negatively influence the effectiveness of Gag-specific activity. We first separated the individuals with *B*5703*, *B*1503* and *B*1510* into four groups: those responding only to the p24 Gag epitopes (shown in **Fig. 3**) restricted by these alleles; those responding to the p24 Gag epitopes and Acc/Reg epitopes restricted by these alleles; those responding to the Acc/Reg epitopes alone; and those responding to neither the Gag nor the Acc/Reg epitopes (**Fig. 4**). In each case, we found that the presence of Acc/Reg protein responses in addition to p24 Gag responses was associated with a higher viral load, although these individual allele-specific differences did not reach statistical significance ($P > 0.05$).

Extending this analysis to the entire cohort of 578 individuals revealed a higher viral load in those with responses to both Gag and the Env/Acc/Reg epitopes than in those with responses to Gag alone (median: 30,350 versus 9,000; $P = 0.0004$). These data also showed that responses to Env and Acc/Reg epitopes are not dependent on the absence of Gag responses but frequently coexist with a Gag response.

### Discordant sequence-independent *HLA*-B and *HLA*-C impact

We next addressed the influence of the restricting *HLA* alleles on the associations observed between CD8$^+$ T-cell response specificity and control of viremia. Previous studies in this population have shown that the expression of particular *HLA-B* alleles has the strongest influence on outcome (in terms of the viral set point, CD4 count and selection pressure on the virus)[5]. When stratified by protein, the *HLA-B*–restricted Gag-specific responses were most strongly associated with the effective control of viremia, and the *HLA-B*–restricted Env- and Acc/Reg-specific responses with a lack of control (**Fig. 5**). In contrast, *HLA-C*–restricted responses were consistently associated with high viremia, even when within Gag. The effect of *HLA-A*–restricted responses was similar to that of *HLA-B*–restricted responses, but had a weaker and less significant impact (for Gag: $P = 0.0186$ for *HLA-A*, $P < 0.0001$ for *HLA-B*; for Acc/Reg and Env: $P > 0.05$ for *HLA-A*, $P = 0.0025$ and 0.0061, respectively, for *HLA-B*). These data indicate that within-protein differences exist between CD8$^+$ T-cell specificities that are related to the *HLA-A*, *HLA-B* or *HLA-C* restriction of the response.

We next addressed the question of whether Gag-specific CD8$^+$ T cells are more likely to be effective primarily because of sequence conservation within Gag, as epitope escape within Gag epitopes might be expected to result in a fitness cost to the virus[6–11]. We compared the viral load of responders to any of these 410 overlapping peptides (4,547 responses in total) with the average entropy of the overlapping peptide targeted, and found, overall, that there was no significant correlation between viral load and peptide entropy ($r = 0.018$, $P > 0.05$). Similar analysis for each individual protein likewise yielded no significant correlation ($P > 0.05$ in each case; data for Gag, Pol, Nef, Vif and Env are shown in **Supplementary Fig. 2** online, $r = 0.026$–$0.042$). Thus, targeting regions of sequence conservation is not a sufficient requirement for a CD8$^+$ T-cell response to be associated with low viremia.

### Stronger Gag- than Env-mediated selection pressure on HIV

The observation that Gag-specific CD8$^+$ T-cell responses are associated with effective control of viremia, whereas Env-specific responses are associated with ineffective control, prompts the question of causality. We hypothesized that effective CD8$^+$ T-cell responses are more likely than noneffective ones to drive selection pressure on the virus, and that greater allele-specific HIV amino acid sequence variation should be observed in Gag compared to Env, despite the fact that Gag is more conserved than Env and therefore less likely to accommodate sequence change without significant fitness cost[6–11]. We

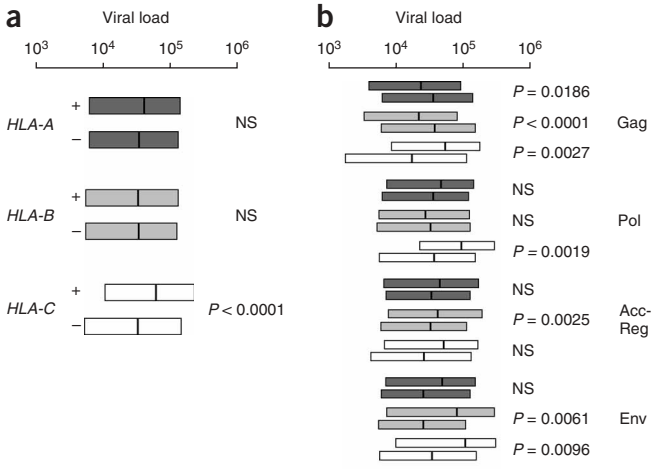


**Figure 5** Comparison of the impact of *HLA-A*–, *HLA-B*– and *HLA-C*–restricted responses on viral load. (**a**) Viral loads (median and interquartile range) of responders and nonresponders to *HLA-A*–, *HLA-B*– and *HLA-C*–restricted epitopes ($n = 49$, 90 and 21, respectively). Significance tested by Mann-Whitney test. (**b**) As in **a** but with data now grouped by protein.

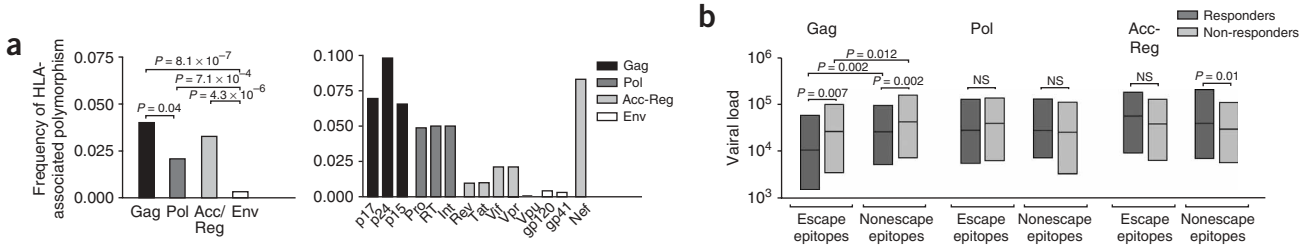


**Figure 6** Protein-specific and epitope-specific CD8+ T-cell responses mediate selection pressure on HIV. (**a**) Frequency of occurrence of *HLA*-associated HIV amino acid polymorphisms within Gag, Pol, Acc/Reg and Env. Data are shown for associations with a *q*-value < 0.05 (Methods). Statistical comparisons by Fisher's exact test. Significant differences were found between (i) Gag, Pol, Acc/Reg and Env; (ii) any one of Rev, Tat, Vpu, gp120 and gp41, and any one of p17, p24, p15, Pro, RT, Int and Nef ($P < 0.05$); and (iii) either Vif or Vpr and either p24 or Nef (not shown). (**b**) Viral loads for 'escape epitopes' (that is, the targets were *HLA-B*–restricted epitopes associated with escape mutations) and 'nonescape epitopes' (targets were *HLA-B*–restricted epitopes not associated with escape mutations). Escape epitopes are listed in **Supplementary Table 3**.

sequenced single full-length virus clones from a subset of 272 study subjects. As hypothesized, *HLA*-associated variation within Gag was significantly higher than that within Env ($P = 8.1 \times 10^{-7}$, Fisher's exact test; **Fig. 6a**) or Pol ($P = 0.04$), but not significantly more than that within the Acc/Reg proteins ($P > 0.05$). Analysis of the individual proteins indicated that Nef was exceptional among the Acc-Reg proteins in that it contained a relatively high level of *HLA*-associated sequence polymorphism, consistent with the fact that Nef is highly immunogenic and that escape within Nef epitopes may be achieved at little fitness cost[12].

To relate these findings to the 160 epitopes identified as dominating the anti-HIV CD8+ T-cell response in this study population, we conducted a similar analysis, except that we sought amino acid sequence polymorphism within each of these epitopes only in association with the expression of the particular MHC class I alleles restricting that response (**Supplementary Table 3** online). We found evidence of escape (defined here as: significant *HLA*-associated sequence variation, after correction for multiple tests, within a defined epitope restricted by the respective allele) in 37 of the 160 epitopes. We initially focused on the 90 *HLA-B*–restricted responses, 25 of which showed evidence of escape. There were no *HLA-B*–restricted epitopes within Env at which there was evidence of escape; thus, the analyses were limited to *HLA-B*–restricted responses within Gag, Pol and Acc/Reg.

Within Gag, we compared the eight *HLA-B*–restricted 'escape epitopes' with the 16 *HLA-B*–restricted 'nonescape epitopes' (**Fig. 6b**). For the eight escape epitopes, responders had lower viral loads than nonresponders (13,700 versus 26,600; $P = 0.007$), consistent with the above data showing a benefit of targeting Gag. Likewise, among the 16 nonescape epitopes, responders had lower viral loads than nonresponders (26,100 versus 42,650; $P = 0.002$), but this analysis also revealed that the responders to the escape Gag epitopes had a lower viral load than the responders to the nonescape Gag epitopes (13,700 versus 26,100; $P = 0.002$). In contrast, these distinctions in terms of viral load between the targeting of escape and nonescape epitopes were not observed in the analyses of Pol and Acc/Reg epitopes. We were unable to undertake meaningful analyses of the *HLA-A*– and *HLA-C*–restricted epitopes because of the very small number of these epitopes at which there was evidence of escape (**Supplementary Table 3**).

These data add to previous studies[6–11] supporting the hypothesis that targeting certain epitopes within Gag drives strong selection pressure on the virus and that escape within these critical epitopes may incur some fitness cost, as suggested by the lower viral load in the targeting epitopes that can escape. Together with the viral load associations, these data suggest that HIV replication is driven by Gag-specific CD8+ T cells, rather than vice versa, and that Env-specific CD8+ T cells, at least in the setting of natural HIV infection, are driven by high levels of viremia.

## DISCUSSION

This large population-based study, linking immune responses, viral load and *HLA* expression, demonstrates that targeting of the HIV-1 Gag protein and the breadth of Gag-specific responses are critical factors associated with effectiveness of an HIV-specific CD8+ T-cell response in the setting of natural HIV infection. This would equate to an expected time-to-AIDS of 2–3 years in the absence of any Gag responses, compared to an expected time of >8–10 years in individuals with ≥2 Gag responses (**Supplementary Table 2** and refs. 13,14). Moreover, these data do not support the hypothesis that non–Gag-specific CD8+ T-cell responses, particularly Env-specific responses, contribute to immune control.

The potential importance of Gag as an immunogen has been suggested in previous smaller studies[4–12,15–35]. A strength of the present study is the cohort size, which enables the impact of individual epitope-specific responses and their restriction by *HLA* to be analyzed at the population level. This allows the effectiveness of individual responses to be assessed by comparing viral loads in responders and *HLA*-matched nonresponders to each epitope. Such a study also effectively controls for potentially confounding variables, as reflected by the wide variation in viral load among individuals generating apparently similar CD8+ T-cell responses. These variables include the extraordinary diversity of *HLA* class I type, the diversity of infecting viral sequence and coinfections with other pathogens. The conclusions of the present study could not have been reached were it not for the availability of the population-based data.

The reasons for these unanticipated protein-specific differences between anti-HIV CD8+ T-cell responses are unknown. Because Gag protein accompanies the incoming virus in an infected cell, intracellular processing of viral capsid protein (>1,000 capsid proteins per virus[36]) would begin immediately on viral entry, whereas intracellular processing of Env epitopes would await *de novo* synthesis of Env. Moreover, CD8+ T cell–mediated Gag mutations may extract a fitness cost resulting from the selection of escape mutations within regions of high sequence conservation[6–11]. Although p24 Gag is highly conserved, conservation alone is insufficient to account for this, as targeting the conserved Pol protein is not associated with effective immune control. There are also conserved regions within the central part of Nef, but viral loads when these regions are targeted are

no lower than those when the variable regions of Nef are targeted. A further factor of potential relevance is the impact of neutralizing antibody escape on the Env-specific $CD8^+$ T-cell responses. However, whether this interaction contributes to the associations observed is unknown.

It is important to distinguish between the implications of these data, obtained from a cohort of HIV-infected individuals, for control of natural chronic infection and the implications for immunogens that may constitute an effective or ineffective vaccine. Although these data demonstrate a lack of evidence for the efficacy of non–Gag-specific $CD8^+$ T-cell responses in chronic HIV infection, it is possible that such responses may mediate successful control of viremia when vaccinated HIV-negative individuals subsequently become HIV infected. However, these data do provide evidence that Gag-specific $CD8^+$ T-cell responses, and in particular broad Gag-specific responses, are effective in the control of chronic HIV infection. As vaccinated individuals who subsequently become infected will also become chronically infected, these data suggest that a broad Gag-specific $CD8^+$ T-cell response is likely to contribute to the maintenance of immune control of HIV in this setting also and that it will be important to monitor breadth and specificity of responses in vaccine recipients who may become infected.

The discordance observed in the associations between the breadth of the Gag-specific response and viral load on the one hand, and that of the non-Gag (particularly Env) response and viral load on the other, suggests two possibilities. The first, as described above, is that Env-specific $CD8^+$ T-cell responses are as effective as Gag-specific responses in controlling viral replication in vaccinees, but are recruited only later in natural infection, in the context of high viral loads. An alternative explanation, also consistent with these data, is that Env-specific responses are inherently less effective than Gag-specific responses. This latter possibility is a critically important consideration for vaccine design and requires further investigation.

Data from the nonhuman primate (NHP) model provide some evidence that non-Gag proteins, in addition to Gag, as $CD8^+$ T-cell targets may contribute to the control of simian immunodeficiency virus (SIV) in vaccinated animals[37–40]. Studies from the NHP model also suggest the potential for Env vaccination to enhance SIV replication and disease[41]. However, it can be problematic to extrapolate from NHP vaccine studies to humans. First, in most NHP vaccine studies, the Env used to vaccinate is very similar to the Env in the challenge virus. Second, the genetic background and, particularly in relation to $CD8^+$ T-cell response, the MHC class I genetic background of the animals vaccinated are likely to be critical in determining the relative success of individual vaccine regimens and challenges. For example, there are substantial differences in the outcome of SIV infection in the Indian, Burmese and Chinese rhesus macaques, which are members of the same *Macaca mulatta* species but differ in the expression of the genes encoding MHC class I (ref. 42). Thus the SIV-macaque model, although invaluable, is not definitive in dictating which $CD8^+$ T-cell responses may be effective in HIV control in humans.

The data here apply to the genetic diversity in a relevant human population, and the evidence presented—that broad Gag-specific, $CD8^+$ T-cell responses are effective in chronic HIV infection—is independent of the particular *HLA* type within the studied population. This supports the hypothesis that, notwithstanding the existence of individual epitope-specific differences, overall $CD8^+$ T cell–mediated control of HIV is protein-specific rather than *HLA*-specific. The lack of a demonstrable biological effect of non–Gag-specific $CD8^+$ T-cell responses in our study does not exclude the possibility that these responses would be effective in vaccinated HIV-uninfected individuals, but challenges the assumption that the more $CD8^+$ T-cell responses that can be induced by a vaccine, the better, irrespective of specificity.

## METHODS

**Study subjects.** We characterized HIV-specific $CD8^+$ T-cell responses in 578 adults, naive to antiretroviral therapy, from Durban, KwaZulu-Natal, South Africa. Of these, 219 were asymptomatic women, identified while pregnant, and were studied after being diagnosed; the remaining 359 subjects were recruited from out-patient follow-up clinics. The median viral load of the entire cohort was 35,259 HIV RNA copies per ml plasma (interquartile range (IQR), 6,115–139,000). Viral load was measured using the Roche Amplicor version 1.5 assay. We obtained informed consent from all participating individuals; the study was approved by institutional review boards at the University of KwaZuluNatal, Massachusetts General Hospital and the University of Oxford.

**ELISPOT, *HLA* restriction assays.** We synthesized 410 18-mer peptides, spanning the HIV proteome, based on the 2001 C-clade consensus sequence. We used these peptides in a matrix system of 11–12 peptides per pool to screen individuals for HIV-specific T-cell responses by IFN-$\gamma$ ELISPOT assay, as previously described[5]. To experimentally define the *HLA* restriction of the responses, we conducted either an intracellular cytokine staining assay or an ELISPOT assay using PBMC or cytotoxic T lymphocyte (CTL) lines as effectors[5,43].

***HLA* class I typing.** Genomic DNA samples were initially typed to an oligo-alleleic level using Dynal RELITM reverse sequence-specific oligonucleotide (SSO) kits for the *HLA-A*, *HLA-B* and *HLA-C* loci (Dynal Biotech). Refining the genotype to the allele level was performed using Dynal Biotech sequence-specific priming (SSP) kits in conjunction with the previous SSO type.

**Sequencing of viral RNA.** Viral sequencing from plasma RNA was undertaken as previously described[44]. RNA was extracted from blood plasma, reverse transcribed, and amplified using nested PCR. The amplified HIV-1 genome was then purified, cloned and sequenced. Near full-length genomes were sequenced from 272 study subjects.

**Statistical methods to identify HLA-peptide associations.** We used Fisher's exact test to identify significant ($P < 0.05$) associations between the recognition of individual 18-mer peptides and the expression of particular *HLA* class I alleles. All 67 *HLA* class I alleles expressed at a phenotypic frequency of $>0.5\%$ were included in the analysis (25 *HLA-A*, 25 *HLA-B* and 17 *HLA-C* alleles). As previously described[5], for each 18-mer peptide recognized, we first looked for the strongest association with an individual *HLA* class I allele, and then sought associations between additional *HLA* class I alleles and targeting of the same peptide. No *HLA* associations were identified for 275 of the 410 peptides (78% of which were targeted by $\leq 3$ individuals); 2 *HLA* associations were identified for 21 peptides, 3 for 7 peptides, and 4 for 5 peptides. To correct for multiple comparisons, we used the conservative Bonferroni correction; thus, associations of $0.05 > P > 0.00075$ were lost following correction for 67 alleles analyzed. Overall, we identified 160 significant associations (in each case, $P < 0.00075$) between the expression of particular class I molecules and recognition (**Supplementary Table 1**). We defined these as the 'dominant' $CD8^+$ T-cell responses in this population.

**Identification of *HLA* class I associations with HIV amino acid sequence polymorphisms.** We used a previously described method[45], with modifications, to identify associations between *HLA* alleles and HIV amino acid sequences (**Fig. 6a**). For every *HLA* allele, the position of the amino acid, and the amino acid at that position, we used Fisher's exact test to analyze the number of patients with and without the allele and the sequences with and without the amino acid (in ref. 45, only the consensus amino acid at each position was studied). In contrast with ref. 45, we used $q$-values to correct for multiple tests[46] (because the counts were finite, $q$-values could be and were computed efficiently in closed form). The number of tests for each protein were 207,766 for Gag; 372,373 for Pol; 48,998 for Rev; 45,124 for Tat; 56,529 for Vif; 45,648 for Vpr; 60,406 for Vpu; 541,188 for Env; and 96,165 for Nef. Triplets of *HLA* allele, amino acid position and amino acid having a $q$-value $< 0.05$ were

deemed significant. Associations of *HLA* and HIV polymorphism corresponding to the same adaptation event were then counted as a single association. Thus, two apparent *HLA* associations with the same HIV sequence polymorphism resulting from linkage disequilibrium between those two alleles were counted as a single association.

Sequences removed before determining the counts comprised five hypermutated sequences, two sequences containing recombinants, and two child sequences. HLA frequencies in the study subjects whose viral sequence was determined did not differ from the frequencies in subjects whose $CD8^+$ T-cell responses were analyzed (**Supplementary Fig. 3** online). Sequences were aligned using ClustalW version 1.83 (ref. 47) and MacClade[48], and regions of uncertain homology were removed.

To identify evidence of escape in the 160 epitopes, we used Fisher's exact test, with correction for multiple tests (18–21 tests per epitope for 8- to 11-mer epitopes) to compare the frequency of amino acid polymorphisms either within the optimal epitope or within the five amino acids on either side flanking the epitope (to accommodate the possibility of processing mutations) in sequences from individuals expressing the relevant restriction element compared to individuals not expressing the relevant *HLA* class I allele (**Supplementary Table 3**).

**Statistical methods to evaluate number of protein-specific associations with viral load.** The association of number of responses to different proteins was estimated and tested statistically using Spearman correlation coefficients (**Supplementary Table 2**). We performed a linear regression to estimate the association between breadth of response and log viral load for each protein individually and in a multivariate model including all four proteins.

*Note: Supplementary information is available on the Nature Medicine website.*

ACKNOWLEDGMENTS

This work is funded by the US National Institutes of Health (contract NO1-A1-15422, 2RO1AI46995-06 and R01AI067073), the Wellcome Trust (A.L. and P.G.), and the Mark and Lisa Schwartz Foundation. B.D.W. is a Doris Duke Distinguished Clinical Science Professor. P.G. is an Elizabeth Glaser Pediatric AIDS Foundation Scientist.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturemedicine
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/



1. Burton, D.R. *et al.* HIV vaccine design and the neutralising antibody problem. *Nat. Immunol.* **5**, 233–236 (2004).
2. Goulder, P.J. & Watkins, D.I. HIV and SIV CTL escape: implications for vaccine design. *Nat. Rev. Immunol.* **4**, 630–640 (2004).
3. Draenert, R. *et al.* Persistent recognition of autologous virus by high-avidity CD8 T cells in chronic, progressive human immunodeficiency virus type 1 infection. *J. Virol.* **78**, 630–641 (2004).
4. Addo, M.M. *et al.* Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* **77**, 2081–2092 (2003).
5. Kiepiela, P. *et al.* Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**, 769–775 (2004).
6. Leslie, A.J. *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* **10**, 282–289 (2004).
7. Friedrich, T.C. *et al.* Reversion of CTL escape-variant immunodeficiency viruses in vivo. *Nat. Med.* **10**, 275–281 (2004).
8. Matano, T. *et al.* Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial. *J. Exp. Med.* **199**, 1709–1718 (2004).
9. Kobayashi, M. *et al.* Reversion in vivo after inoculation of a molecular proviral DNA clone of simian immunodeficiency virus with a cytotoxic-T-lymphocyte escape mutation. *J. Virol.* **79**, 11529–11532 (2005).
10. Fernandez, C.S. *et al.* Rapid viral escape at an immunodominant simian-human immunodeficiency virus cytotoxic T-lymphocyte epitope exacts a dramatic fitness cost. *J. Virol.* **79**, 5721–5731 (2005).
11. Martinez-Picado, J. *et al.* Fitness cost of escape mutation in p24 Gag in association with control of HIV-1. *J. Virol.* **80**, 3617–3623 (2006).
12. Leslie, A. *et al.* Transmission and accumulation of CTL escape variants drives negative associations between HIV polymorphisms and HLA. *J. Exp. Med.* **201**, 891–902 (2005).
13. Mellors, J.W. *et al.* Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**, 1167–1170 (1996).
14. Lyles, R.H. *et al.* Natural history of HIV-1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. *J Infect. Dis.* **181**, 872–880 (2000).
15. Goulder, P.J. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat. Med.* **3**, 212–217 (1997).
16. Feeney, M.E. *et al.* Immune escape precedes breakthrough HIV-1 viremia and broadening of the CTL response in a HLA-B27-positive long-term nonprogressing child. *J. Virol.* **78**, 8927–8930 (2004).
17. Wilson, J.D. *et al.* Direct visualization of HIV-1-specific cytotoxic T lymphocytes during primary infection. *AIDS* **14**, 225–233 (2000).
18. Barouch, D.H. *et al.* Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**, 335–339 (2002).
19. Allen, T.M. *et al.* Tat-specific CTL select for SIV escape variants during resolution of primary viremia. *Nature* **407**, 386–390 (2000).
20. Migueles, S.A. *et al.* The differential ability of HLA B*5701+ long-term nonprogressors and progressors to restrict human immunodeficiency virus replication is not caused by loss of recognition of autologous viral gag sequences. *J. Virol.* **77**, 6889–6898 (2003).
21. Riviere, Y. *et al.* Gag-specific cytotoxic responses to HIV type 1 are associated with a decreased risk of progression to AIDS-related complex or AIDS. *AIDS Res. Hum. Retroviruses* **11**, 903–907 (1995).
22. Klein, M.R. *et al.* Kinetics of Gag-specific cytotoxic T lymphocyte responses during the clinical course of HIV-1 infection: a longitudinal analysis of rapid progressors and long-term asymptomatics. *J. Exp. Med.* **181**, 1365–1372 (1995).
23. Ogg, G.S. *et al.* Quantitation of HIV-1-specific cytotoxic T lymphocytes and plasma load of viral RNA. *Science* **279**, 2103–2106 (1998).
24. Edwards, B.H. *et al.* Magnitude of functional $CD8^+$ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *J. Virol.* **76**, 2298–2305 (2002).
25. Novitsky, V. *et al.* Association between virus-specific T-cell responses and plasma viral load in human immunodeficiency virus type 1 subtype C infection. *J Virol.* **77**, 882–890 (2003).
26. Masemola, A. *et al.* Hierarchical targeting of subtype C human immunodeficiency virus type 1 proteins by $CD8^+$ T cells: correlation with viral load. *J. Virol.* **78**, 3233–3243 (2004).
27. Ramduth, D. *et al.* Differential immunogenicity of HIV-1 clade C proteins in eliciting $CD8^+$ and $CD4^+$ cell responses. *J. Infect. Dis.* **192**, 1588–1596 (2005).
28. Zuniga, R. *et al.* Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *J. Virol.* **80**, 3122–3125 (2006).
29. Musey, L. *et al.* Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* **337**, 1267–1274 (1997).
30. Borrow, P. *et al.* Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat. Med.* **3**, 205–211 (1997).
31. Safrit, J.T. *et al.* Characterization of human immunodeficiency virus type 1-specific cytotoxic T lymphocyte clones isolated during acute seroconversion: recognition of autologous virus sequences within a conserved immunodominant epitope. *J. Exp. Med.* **179**, 463–472 (1994).
32. Safrit, J.T. *et al.* A region of the third variable loop of HIV-1 gp120 is recognized by HLA-B7-restricted CTLs from two acute seroconversion patients. *J. Immunol.* **153**, 3822–3830 (1994).
33. Price, D.A. *et al.* Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc. Natl. Acad. Sci. USA* **94**, 1890–1895 (1997).
34. Betts, M.R. *et al.* Analysis of total human immunodeficiency virus (HIV)-specific $CD4^+$ and $CD8^+$ T-cell responses: relationship to viral load in untreated HIV infection. *J. Virol.* **75**, 11983–11991 (2001).
35. Frahm, N. *et al.* Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. *J. Virol.* **78**, 2187–2200 (2004).
36. Briggs, J.A. *et al.* The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **11**, 672–675 (2004).
37. Gallimore, A. *et al.* Early suppression of SIV replication by $CD8^+$ Nef-specific cytotoxic T cells in vaccinated macaques. *Nat. Med.* **1**, 1167–1173 (1995).
38. Hel, Z. *et al.* Improved protection from simian AIDS by the addition of non-structural SIV genes. *J. Immunol.* **176**, 85–96 (2006).
39. Casimiro, D.R. *et al.* Attenuation of SIVmac239 by prophylactic immunization with DNA and recombinant adenoviral vectors expressing Gag. *J. Virol.* **79**, 15547–15555 (2005).
40. Wilson, N.A. *et al.* Vaccine-induced cellular immune responses reduce plasma viral concentrations after repeated low-dose challenge with pathogenic SIVmac239. *J. Virol.* **80**, 5875–5885 (2006).
41. Staprans, S.I. *et al.* Enhanced SIV replication and accelerated progression to AIDS in macaques primed to mount a CD4 T cell response to SIV envelope protein. *Proc. Natl. Acad. Sci. USA* **101**, 13026–13031 (2004).

42. Reimann, K.A. *et al.* Pathogenicity of simian-human immunodeficiency virus SHIV-89.6P and SIVmac is attenuated in cynomolgus macaques and associated with early T-lymphocyte responses. *J. Virol.* **79**, 8878–8888 (2005).
43. Goulder, P.J. *et al.* Rapid definition of five novel HLA-A*3002-restricted human immunodeficiency virus-specific cytotoxic T-lymphocyte epitopes by elispot and intracellular cytokine staining assays. *J. Virol.* **75**, 1339–1347 (2001).
44. Rousseau, C.M. *et al.* Method for high-throughput amplification, cloning and sequencing of full-length HIV-1 subtype C genomes. *J. Virol. Methods* **136**, 118–125 (2006).
45. Moore, C.B. *et al.* Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**, 1439–1443 (2002).
46. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
47. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
48. Maddison, W.P. & Maddison, D.R. Interactive analysis of phylogeny and character evolution using the computer program MacClade. *Folia Primatol. (Basel)* **53**, 190–202 (1989).