Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection

Zabrina L. Brumme[a], Iris Tao[b,*], Sharon Szeto[b,*], Chanson J. Brumme[a], Jonathan M. Carlson[c,d], Dennison Chan[b], Carl Kadie[c], Nicole Frahm[a], Christian Brander[a], Bruce Walker[a,e], David Heckerman[c] and P. Richard Harrigan[b,f]

Objective: Selection of specific human leukocyte antigen (HLA)-restricted cytotoxic T-lymphocyte (CTL) escape mutations in key Gag epitopes has been associated with loss of HIV immune control on an individual basis. Here we undertake a population-based identification of HLA-associated polymorphisms in Gag and investigate their relationship with plasma viral load.

Design: Cross-sectional analysis of 567 chronically HIV subtype B-infected, treatment-naive individuals.

Methods: HLA class I-associated Gag substitutions were identified using phylogenetically corrected analysis methods featuring a multivariate adjustment for HLA linkage disequilibrium and a *q*-value correction for multiple tests. Presence of HLA-associated substitutions and markers of HIV disease status were correlated using Spearman's rank test.

Results: We have created a gene-wide map of HLA class I-associated substitutions in HIV-1 subtype B Gag. This features 111 HLA-associated substitutions occurring at 51 of 500 Gag codons, more than 50% of which occur within published and/or putative HLA-restricted CTL epitopes. A modest inverse correlation was observed between the total number of HLA-associated Gag polymorphic sites within each individual and plasma viral load in chronic untreated infection ($R = -0.17$, $P < 0.0001$), supporting the hypothesis that a broad ability to target Gag *in vivo* contributes to viral control. A modest positive correlation was observed between the proportion of these sites exhibiting HLA-associated substitutions and plasma viral load ($R = 0.09$, $P = 0.03$), consistent with a loss of viremia control with the accumulation of CTL escape mutations.

Conclusion: Results contribute to our understanding of immune-driven viral adaptation and suggest that the accumulation of CTL escape mutations in Gag results in clinically detectable consequences at the population level. These data have implications for HIV vaccines.

AIDS 2008, **22**:1277–1286

[a]Partners AIDS Research Center, Massachusetts General Hospital, Boston, Harvard Medical School, Massachusetts, USA, [b]BC Centre for Excellence in HIV/AIDS, Vancouver, Canada, [c]Microsoft Research, Redmond, Washington, [d]Department of Computer Science, University of Washington, Seattle, Washington, [e]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA, and [f]Division of AIDS, University of British Columbia, Vancouver, British Columbia, Canada.

Correspondence to Zabrina L. Brumme, PhD, Partners AIDS Research Center, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA.

[*] These authors contributed equally to this work.

Received: 24 January 2008; revised: 12 March 2008; accepted: 18 March 2008.

Keywords: cytotoxic T-lymphocyte escape, HIV-1, human leukocyte antigen class I, plasma viral load, viral evolution

Introduction

The human leukocyte antigen (HLA)-restricted cytotoxic T-lymphocyte (CTL) response is a key determinant of immune control of HIV infection [1,2], and recent data suggest that CTL targeting of Gag contributes strongly to this control [3–6]. However, the selection of HLA-restricted CTL escape mutations likely compromises the ability of CTL to suppress HIV replication in the long-term [7–10]. Over the last decade, a number of Gag CTL escape mutations have been identified in cross-sectional and observational studies [7–11]. These, however, have been somewhat biased toward HLA alleles associated with long-term control of HIV infection [11–17], and a population-based, gene-wide identification of HLA-associated polymorphisms across Gag remains to be undertaken. Similarly, although loss of viremia control following CTL escape has been reported for single epitopes on a case-by-case basis [17–19], results appear to be poorly generalizable across all epitopes and HLA restrictions [20,21]. A clear relationship between CTL escape in Gag and HIV disease progression has yet to be demonstrated on a more general level.

Here we identify HLA class I-associated polymorphisms across Gag in a large ($N = 567$), population-based cohort of chronically infected, treatment-naive individuals harboring predominantly subtype B infections [22]. We use novel statistical approaches correcting for the underlying phylogenetic structure of HIV sequence data [23] to create an 'immune escape map' of HLA-associated polymorphisms in Gag. We then use these defined polymorphisms to explore the relationship between plasma viral load (pVL), CD4 cell count, and mutations selected under active in-vivo immune pressure in Gag.

Methods

Cohort description

HLA-associated polymorphisms in Gag were identified in a cross-sectional analysis of 567 antiretroviral naive, chronically infected individuals selected from the British Columbia Highly Active Antiretroviral Therapy (HAART) Observational Medical Evaluation and Research (HOMER) cohort [24] based on the concomitant availability of a peripheral blood sample for HLA typing and HIV Gag sequence data. HLA-associated polymorphisms in protease, reverse transcriptase, Vpr, and Nef were previously defined in this cohort using consistent methods [22]. CD4 count, pVL, and viral sequences correspond to the latest pretherapy measurement collected within 180 days prior to HAART initiation. Ethical approval was granted by the University of British Columbia-Providence Healthcare Research Ethics Board (UBC-PHC REB). For scientists wishing to investigate HLA-associated polymorphisms in this dataset, HLA–HIV data are available upon request pending application to and approval by the UBC-PHC REB.

Laboratory methods

Gag and a portion of protease were amplified from extracted plasma HIV RNA by nested reverse transcriptase–PCR using gene-specific primers, and 'bulk' sequenced in 5′ and 3′ directions on an ABI 3730 automated sequencer. Data were analyzed using Sequencher (GeneCodes, Ann Arbor, Michigan, USA). Nucleotide mixtures were called if the secondary peak exceeded 25% of the dominant peak height. Sequences were aligned to HIV-1 subtype B reference strain HXB2 (GenBank accession no. K03455) using a modified NAP algorithm [25]. Sequences have been deposited in GenBank (EU241938–242504). HLA class I typing was performed as described [22].

Identification of human leukocyte antigen-associated polymorphisms

Previous studies [23] have demonstrated the importance of correcting for the underlying phylogenetic structure of HIV sequence data when identifying sites of HLA-mediated immune selection. To identify HLA-associated polymorphisms in Gag, we therefore used phylogenetically corrected methods described in detail in [22,26], modified to incorporate a correction for linkage disequilibrium between HLA alleles. Briefly, a maximum likelihood phylogenetic tree was constructed from Gag sequences. For each observed amino acid, a likelihood ratio test was used to evaluate whether a model incorporating both phylogenetic structure and HLA-mediated selection pressure (conditional evolution model) explained the data significantly better than a model assuming neutral evolution according to the tree (independent evolution model). The presence or absence of a given HLA allele was correlated with the presence or absence of a given amino acid on a codon-by-codon basis (restricted to HLA–HIV amino acid pairs with a minimum of three actual or expected counts in every cell of the two-by-two contingency table). For every amino acid at each codon, the allele with the strongest association was added to the list of identified associations. Then, individuals expressing this allele were removed and the analysis repeated. This forward-selection procedure was iterated until no HLA allele yielded an association with uncorrected P less than 0.05, thus tending to eliminate spurious associations due to HLA linkage

disequilibrium. A q-value statistic, which estimates the false-discovery rate among identified associations, was used to adjust for multiple comparisons [27,28]. Associations with q-value 0.2 or less (indicating a 20% false-discovery rate) are presented; this corresponded to P-values $6.5 \times 10^{-37} < P < 0.0015$.

Associations were grouped into two categories based on the direction of the selection pressure. We refer to amino acids enriched in the presence of a particular HLA allele (presumably representing the residue most likely to emerge under CTL-selection pressure) as the 'escape' form, and amino acids enriched in the absence of a particular HLA allele (presumably reflecting the 'wild-type' residue most likely to reemerge following transmission to an HLA-unmatched individual) as the 'reversion' form.

Immune escape maps

HLA-associated polymorphisms in Gag were organized into immune escape maps. If an association fell within three amino acids of a published CTL epitope (defined as all HLA class I-restricted 15-mer or less epitopes listed in the Los Alamos HIV Immunology database as of 7 June 2007 [29]) or predicted CTL epitope (as identified by Epipred, a published epitope prediction tool available at http://atom.research.microsoft.com/bio/epipred.aspx [30]), the sequence and HLA restriction of this epitope was indicated. The +/− three amino acid window was chosen to account for putative proteasomal processing escape mutations [31–34]. To provide in-vitro functional support to identified associations, we drew upon an independent dataset of 372 HIV-1 infected individuals screened for CTL-mediated interferon (IFN)-γ responses against a set of 410 overlapping subtype B consensus peptides spanning the whole expressed HIV-1 proteome using an IFN-γ enzyme-linked immunosorbent spot (ELISpot) assay [35]. Fisher's exact test was used to identify HLA alleles significantly associated with IFN-γ production in response to particular peptides, thus supporting the presence of an uncharacterized HLA-restricted CTL epitope within the peptide boundaries.

A 20% false-discovery rate is expected with q-value 0.2 or less [26,27]. Although we attempted to assign as many associations as possible to known or putative CTL epitopes, epitope proximity alone does not guarantee a true-positive result. Furthermore, as associations occurring distant to epitopes may represent compensatory mutations, distant processing mutations, or other factors [22]. Therefore, all identified associations are included on the map.

Results

Human leukocyte antigen-associated substitutions in Gag

A substantial level of amino acid variation was observed in p17 (matrix) and C-terminal proteins (p2/p7/p1/p6), with non-consensus variation at single codons ranging from 0 to 59.6% (interquartile range [IQR] 0.4–11.7%) in p17 and 0 to 57.7% (IQR 0.1–14.3%) in p2/p7/p1/p6 using the cohort (subtype B) consensus as a reference. In contrast, p24 (capsid) was more conserved, with single codon variation ranging from 0 to 42.7% (IQR 0–1.1%). A total of 111 unique HLA-associated polymorphisms were identified with q-value 0.2 or less: these included 36, 42, and 33 associations occurring at 16 (of 132; 12%) codons in p17, 16 (of 231; 7%) codons in p24, and 19 (of 137; 14%) of codons in p2/p7/p1/p6 (Fig. 1). As expected, HLA associations were observed at relatively variable residues, although not exclusively (Fig. 1). Of the 111 observed associations, 50 (45%) represented escape polymorphisms (amino acids enriched in the presence of a specific HLA allele) whereas 61 (55%) represented reversion polymorphisms (those enriched in the absence of a specific HLA allele). Half (55 of 111) of the observed

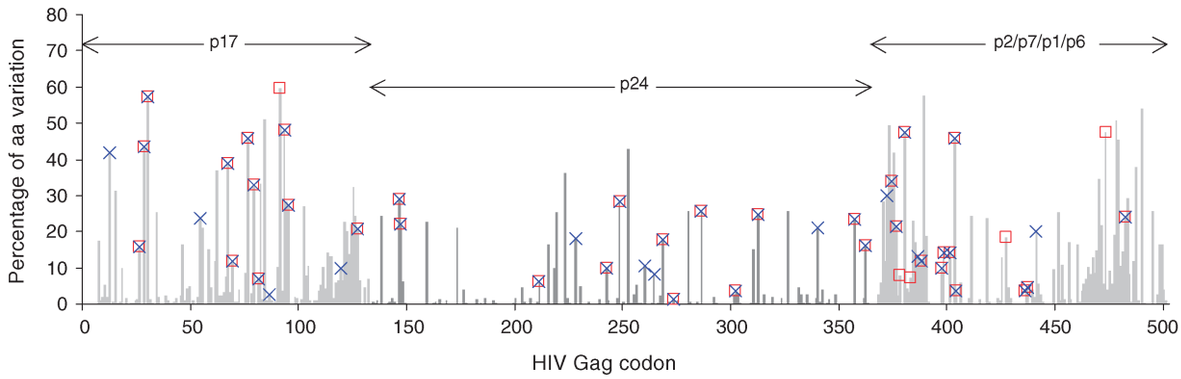

Fig. 1. Codon variability plot and location of HLA-associated viral polymorphisms across HIV Gag. Polymorphism maps display the percentage of amino acid variation observed at each codon in p17, p24, and p2/p7/p1/p6 (C-terminal) HIV-1 Gag. Red square and blue cross symbols indicate codons at which HLA-associated escape and reversion substitutions, respectively, were observed with a q-value of is 0.2 or less.

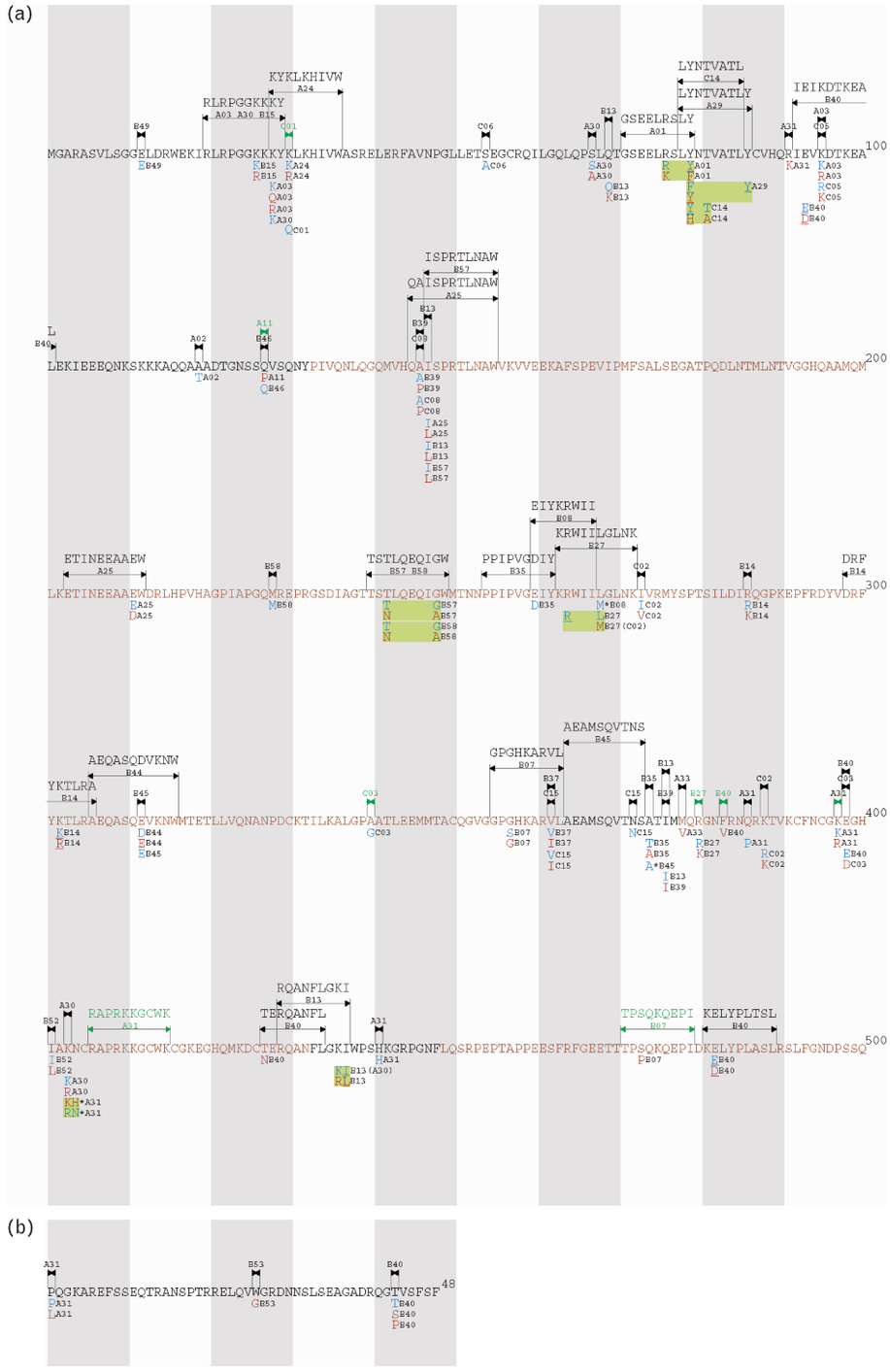

Fig. 2. Immune escape map of HLA-associated polymorphisms identified in Gag (panel a) and p6^{Pol} linker peptide (panel b). The HIV-1 Gag consensus B amino acid sequence is used as a reference. Sequences of the individual Gag proteins (p17, p24, p2, p7, p1, p6) are indicated by alternating black and brown lettering (panel a). The sequence of the p6^{Pol} linker peptide is shown in black (panel b). Shaded bars represent blocks of 10 amino acids. Escape associations (amino acids enriched in the presence of the indicated HLA allele) are red, whereas reversion associations (amino acids enriched in the absence of the indicated allele) are blue. If an HLA-associated polymorphism mapped within or near (+/−3aa) a published or predicted epitope, the epitope sequence and its HLA restriction is indicated in black (published) or green (predicted). In cases where multiple overlapping epitopes were present near the association, a representative epitope was chosen. Associations mapping within Gag regions significantly associated with HLA-restricted in-vitro IFN-γ (ELISpot) responses [35] (but for whom a particular epitope sequence was not predicted) are also shown in green.

associations were attributable to HLA-B alleles, a fraction consistent with the higher number of unique alleles at the B locus compared to A and C loci.

Gag immune escape map

Of the 111 identified HLA-associated polymorphisms, nearly half (51 of 111; 46%) mapped inside or within three amino acids of a published CTL epitope, whereas an additional 13 (12%) mapped within or near a putative epitope as supported by functional IFN-γ ELISpot data [35] and/or the presence of an HLA-binding motif [30] (Fig. 2a). Results confirmed a number of previously identified escape mutations, including the B*57-associated substitution at codon 147 within the B*57-restricted p24 IW9 epitope [32,36], B*57 and B*58-associated substitutions at codons 242 and 248 within the B*57/B*58-restricted p24 TW10 epitope [10,11], B*27-associated substitutions at codons 264 and 268 within the B*27-restricted p24 KK10 epitope [16,17], and B*13-associated substitutions at codons 147, 436, and 437 occurring within or near recently defined B*13-restricted epitopes [12]. We confirmed a number of associations identified in the only other lineage-corrected analysis of Gag to date [23], including B*07-associated substitutions at codon 357, A*31-associated substitutions at codons 397 and 403, and B*40-associated substitutions at codons 398 and 482. Of interest, no HLA-A*02-associated substitutions were observed in the p17 SL9 (SLYNTVATL) epitope, despite previous reports describing such substitutions at positions 3, 6, and 8 (HXB2 codons 79, 82, and 84) [21]. Substitutions were, how-ever, identified within A*01, A*29, and C*14 epitopes overlapping SL9.

A number of epitopes (including A*01-GY9, C*14-LY8, B*57/B*58-TW10, B*27 KK10, and B*13-RI9) harbored HLA-associations at multiple residues, highlighting cases in which escape involves multiple substitutions. Also of interest are six Gag codons (79, 95, 268, 312, 374, and 403) under diametrically opposed HLA-associated selection pressures at the population level (in which the escape form for one allele represents the reversion form for another, and vice versa). Finally, no significant enrichment for HLA-associated polymorphisms was observed at HLA anchor residues (generally residues 2 and C-terminal with some exceptions [29,37]) compared to other positions within CTL epitopes ($P = 0.15$). This indicates that abrogation of peptide–HLA binding is not the preferred route of CTL escape in most cases [22], providing support for the inclusion of HLA-associated polymorphisms into immunogen design [38].

HLA-associated 'insertion' and 'frameshift' polymorphisms

Amino acid insertions are commonly observed at the 3′ end of p17, as well as in p6, and may represent a novel mechanism of CTL escape [39]. In a posthoc analysis, six HLA-associated 'insertion polymorphisms' were identified. Possession of B*44, A*32, or C*05 was associated with having insertions near the C-terminus of p17 (at codons 119 [B*44] or 125 [A*32 or C*05]), whereas C*07 was associated with insertions at codon 471 in p6. In contrast, possession of A*26 and A*02 were associated with the absence of insertions at codons 119 (p17) and 471 (p6), respectively. None of these associations mapped within three amino acids of a known epitope. In addition, the Gag–Pol ribosomal frameshift occurs within p1, and peptides derived from either $p6^{Gag}$ or $p6^{Pol}$ frames could represent CTL targets [29,39]. Six additional HLA-associated polymorphisms were identified at three $p6^{Pol}$ codons (Fig. 2b). Only the B*53 association at codon 34 was, however, unique to the $p6^{Pol}$ reading frame, and none mapped within or near a known or predicted epitope.

The potential breadth of HLA-mediated anti-Gag selection pressure correlates inversely with plasma viral load

To investigate the extent to which an individual's total potential in-vivo HLA-restricted anti-Gag CTL targeting capacity may contribute to HIV immune control [6], we investigated the relationship between the total number of HLA-associated polymorphic Gag sites in each subject and clinical markers of HIV disease (pVL and CD4 count) in a cross-sectional analysis. Each subject was assigned a score reflecting the total number of Gag codons at which an HLA association was observed at the population level. For example, Fig. 2 indicates that the A*01, A*02, B*07, B*57, C*05, and C*06 alleles were associated with substitutions at 2, 1, 2, 3, 1, and 1 Gag codons, respectively; a subject with this HLA profile would thus be assigned a score of 10.

A modest yet significant inverse correlation was observed between the total number of HLA-associated sites in Gag and pVL in chronic untreated infection ($R = -0.17$, $P < 0.0001$; Fig. 3a), supporting the hypothesis that a broad ability to target Gag contributes to HIV immune control [3–7]. This remained significant after removal of individuals expressing known 'protective' alleles B*57, B*27, B*51, or B*13 [12,40]. These trends remained generally consistent when class I loci were examined individually: significant associations were observed

Fig. 2. (*Continued*).
Polymorphisms associated with the same HLA allele(s) occurring within a single published or predicted epitope are boxed in yellow. Underlined associations indicate HLA anchor residue sites. Proximal associations (within +/− 3aa of epitope boundaries) are marked with a star (*). The escape map does not list the locations of all published CTL epitopes. This is available at http://www.hiv.lanl.gov/content/immunology.

between the number of HLA-A ($R=-0.10$, $P=0.01$) and HLA-B ($R=-0.11$, $P=0.01$), but not HLA-C-associated Gag polymorphic sites and pVL. The association between pVL and the total number of polymorphic Gag sites remained significant when individual Gag proteins were examined individually: p17 ($R=-0.11$, $P=0.008$), p24 ($R=-0.1$, $P=0.03$), and p2/p7/p1/p6 ($R=-0.13$, $P=0.007$). Despite statistical significance, the overall magnitude of this effect was small: interpreting the coefficient of determination statistic (R^2) indicates that less than 3% of observed population pVL variability is explained by differences in the number of Gag sites under in-vivo HLA-associated selection pressure. Note that no correlation was observed between the number of HLA-associated Gag sites and CD4 count ($R=0.04$, $P=0.3$; Fig. 3b). Analysis of HLA loci and Gag proteins separately also failed to show a significant association between HLA-associated Gag sites and CD4 cell count (not shown).

The proportion of escaped positions within Gag correlates positively with plasma viral load and inversely with CD4 cell count

Although the clinical consequences of Gag escape have been documented on a case-by-case basis [17–19], the relationship between HLA-associated Gag polymorphisms and pVL has not been investigated on a general level. For each individual, we calculated the 'proportion of escaped sites' (i.e. the proportion of sites displaying the specific HLA-associated substitution identified in Fig. 2) and correlated this statistic with pVL (Fig. 4a) and CD4 cell count (Fig. 4b). To use the previous example, an individual with the HLA profile A*01/*02, B*07/*57, C*05/*06, who had escaped at five of their 10 possible sites, was assigned a value of 0.5. Note that values below 0.2 were relatively rare, consistent with a cohort in relatively advanced disease.

A modest yet significant positive correlation was observed between the proportion of escaped sites and higher pVL

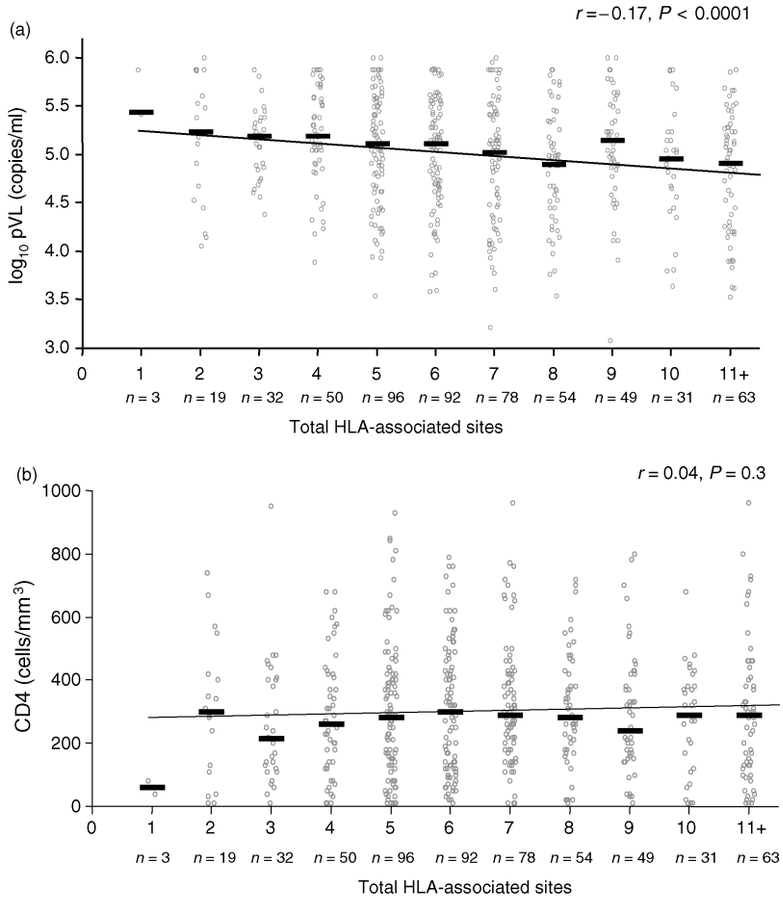

Fig. 3. Relationship between the number of HLA-associated polymorphic sites in Gag and clinical markers of untreated HIV infection. Log_{10} plasma viral loads (panel a) and CD4 counts (panel b) of cohort participants, in terms of the total number of HLA-associated polymorphic sites in Gag, as dictated by their combined HLA class I allele profile. Horizontal bars indicate medians for each group. A regression line is drawn to highlight trends. Numbers for each group are indicated below each graph.

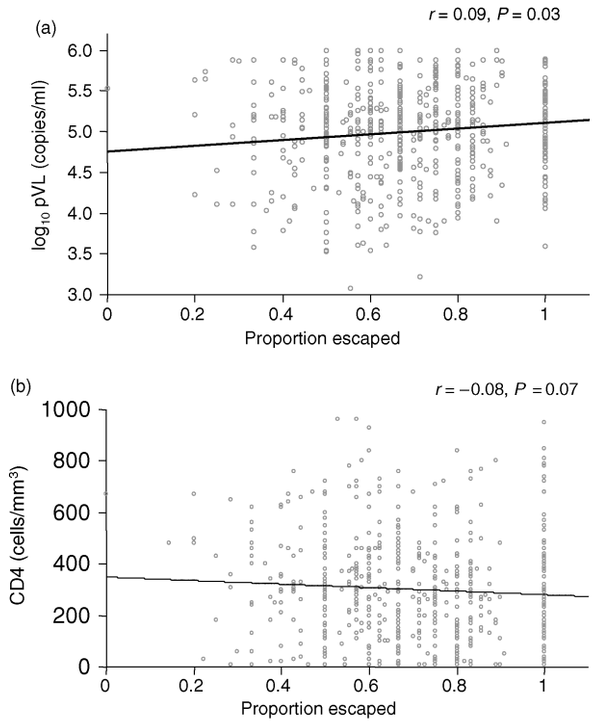

Fig. 4. Relationship between the proportion of escaped sites in Gag and clinical markers of untreated HIV infection. Log_{10} plasma viral loads (panel a) and CD4 counts (panel b) of cohort participants, in terms of the proportion of HLA-specific escaped sites in Gag. A regression line is drawn to highlight trends.

($R = 0.09$, $P = 0.03$). Analyzing each HLA locus and each Gag protein separately revealed no HLA locus-specific or Gag protein-specific correlations, with the exception of p17, which exhibited a significant positive correlation between proportion of escaped sites and higher pVL ($R = 0.09$, $P = 0.03$). A trend toward an inverse correlation was observed between the proportion of escaped Gag codons and lower CD4 cell counts ($R = -0.08$, $P = 0.07$). Of interest, this trend was driven by p24, which exhibited a significant inverse correlation between proportion of escaped codons and lower CD4 cell counts ($R = -0.11$, $P = 0.01$). Analyzing other Gag proteins as well as HLA loci separately revealed no additional significant correlations between proportion of escaped Gag codons and CD4 count.

Discussion

The present study represents the largest population-based assessment of HLA-associated imprinting on HIV-1 Gag to date. The systematic determination and mapping of more than 100 HLA-associated Gag polymorphisms *in vivo* (Fig. 2) provides a complementary analysis with CTL/CD8+ epitope maps available at the Los Alamos HIV database (http://www.hiv.lanl.gov/content/immunology). Results confirm a number of Gag escape mutations previously identified in clade B [10,11,16,17], clade C [12], and mixed-clade [23] studies, indicating that despite extreme global sequence diversity, HIV subtypes share at least some common escape pathways.

The cohort was previously used to map sites of HLA-mediated selection in protease, reverse transcriptase, Vpr, and Nef [22]. Despite the fact that Gag (notably p24) and Nef feature high epitope density [29] and are vigorously targeted by CTL in early infection [41], the overall proportion of Gag codons exhibiting HLA-associated polymorphisms is substantially lower than Nef (~10 vs. 40%, respectively) [22]. Instead, the density of HLA-associated polymorphisms in Gag is comparable to Pol [22] a somewhat less frequently targeted protein with lower epitope density, supporting substantial mutational constraints on Gag despite frequent targeting by CTL [42].

The lack of identification of A*02 SL9 escape mutations in this study is consistent with low in-vivo selection pressure on this epitope [41,43–45], as well as the observation that SL9 variants are found at comparable frequencies among A*02-positive and negative individuals [21]. Although these observations appear to support the 'HLA imprinting hypothesis' (which states that the contemporary HIV consensus reflects adaptation to the most frequently expressed HLA alleles) [46–49], it is important to note that current phylogenetically corrected methods are designed to identify HLA-mediated selection events occurring in the terminal branches of the tree only and thus cannot be used to directly address primordial HIV adaptation to host HLA. The observation that more than 50% of identified associations in Gag (and other proteins [22]) represent reversion associations, combined with the observation that reversion occurs in many cases after transmission to an HLA-unmatched individual [11,50–52] highlights the ongoing, dynamic viral evolution in response to host selection pressures. Analysis of additional contemporary and historical HIV cohorts will further our understanding of how host-related selection forces have shaped HIV evolution at the population level throughout the course of the epidemic, and how they may do so in the future.

Although assessment of in-vitro CTL responses was not possible due to lack of cryopreserved peripheral blood mononuclear cells (PBMC), we observed a significant inverse correlation between potential in-vivo breadth of the anti-Gag CTL response (conservatively estimated by calculating the total number of polymorphic Gag sites specific to each individual's HLA type) and lower pVL, thus supporting the hypothesis that a broad ability to target Gag contributes to HIV control [3,6]. Alleles at the B locus contributed strongly to this trend, consistent with a dominant influence of HLA-B in mediating HIV control [53]. Estimated breadth of CTL responses against p17 and C-terminal Gag proteins correlated most

strongly with pVL, suggesting that targeting of all Gag proteins (not only p24) contribute to immune control.

Although the clinical consequences of single Gag CTL escape mutations have been documented on a case-by-case basis [17–20], the relationship between pVL and the presence of HLA-associated Gag substitutions has not been examined at the population level. We observed a modest yet significant positive correlation between the proportion of escaped positions in Gag and higher pVL, suggesting that the accumulation of escape mutations has measurable consequences for containment of viremia by CTL. Note, however, that the absolute magnitude of the effect was minor: using linear regression, each 10% increase in escaped sites was associated with a 7% increase in pVL. Consistent with an accumulation of CTL escape mutations (and concomitant decline in immune function) over the natural course of infection [22], we observed an inverse correlation, although not statistically significant, between the proportion of escaped codons in Gag and lower CD4 cell count.

Given the potential use of results such as these to guide the selection of CTL vaccine immunogens, it is important to put the results in context of other HIV proteins. In a previous publication, we identified HLA-associated polymorphisms in protease/reverse transcriptase and Nef and investigated the relationship between these polymorphisms and clinical markers of HIV disease using similar methods [22]. To allow an unbiased comparison of results, we reanalyzed the protease/reverse transcriptase and Nef datasets using the methods described in this study. We observed a weak, marginally significant inverse correlation between the total number of HLA-associated sites in Nef and pVL in chronic untreated infection ($R = -0.07$, $P = 0.04$), however, no correlation between pVL and HLA-associated sites in protease/reverse transcriptase and pVL was observed (not shown). These results support the hypothesis that CTL targeting of Gag contributes more to immune control of HIV than CTL targeting of other proteins. With respect to the immune consequences of escape, we observed no significant association between the proportion of escaped sites and pVL in Nef or protease/reverse transcriptase (not shown). Again, this result supports a more substantial consequence of escape in Gag compared to other HIV proteins.

Complex conflicting forces mediate the relationship between escape and clinical progression. In many cases the selected escape variant maintains some HLA binding and/or CTL recognition capacity [54], whereas in others, de-novo CTL responses may arise against escaped variants [55], such that the selection of individual mutations may result in only incremental, if any, consequences to viremia control. In addition, Gag escape mutations may come at a substantial cost to viral fitness (as recently demonstrated for the immunodominant B*57 and B*27-restricted TW10 [13,14] and KK10 [15] epitopes, respectively), thus rendering the potential pVL consequences of escape smaller than anticipated. Note, however, that in the present study, the association between Gag escape and lower pVL remained after controlling for B*57 and/or B*27 expression. Furthermore, the fact that the cohort comprises chronically infected individuals may also underestimate the correlation between anti-Gag CTL targeting (and escape) on clinical disease markers, as the residual effects of anti-HIV CTL activity in earlier infection may be difficult to detect in later stages. Other limitations include the use of a single pVL and CD4 measurement to represent the clinical status of each individual, as well as the lack of longitudinal HIV sequence data and comprehensive untreated clinical histories (including seroconversion dates). Finally, the cross-sectional study design prevents us from drawing conclusions regarding whether higher pVL is a cause or consequence of escape. Even acknowledging these limitations, the relatively weak correlation between estimated anti-Gag CTL targeting potential and untreated pVL implies that a successful HIV vaccine would likely have at least an order of magnitude greater effective immune responses than those generally observed in the context of natural infection [56–58]. The recent failure of a phase-II CTL-based vaccine trial [59,60] further underscores the need to better define the correlates of HIV immune protection, and the need to develop strategies for enhancing the appropriate responses through vaccination [56–58].

Nevertheless, the current study provides evidence that the clinical correlates of CTL targeting and escape are more pronounced for Gag when compared to other HIV proteins [22]. These observations further support an important role for CTL targeting of Gag in containment of HIV viremia [3–6,12,61], and provide support for CTL-based vaccine strategies focusing on immunogenic yet mutationally refractive viral regions such as Gag.

Acknowledgements

The authors wish to thank Theresa Mo and Celia Chui for technical assistance, Brian Wynhoven and Conan Woods for database assistance, Philippa Matthews and Philip Goulder for helping to develop the phylogenetically corrected, HLA linkage-disequilibrium-corrected analysis.

Funding Sources: Z.L.B. is supported by a postdoctoral fellowship from the Canadian Institutes for Health Research (CIHR). D.H., J.C., and C.K. are funded by Microsoft Research. This work has been funded, in part, with Federal funds from the NIH under contract N01-Al-15422 on HLA typing and T cell epitope mapping relative to HIV vaccine design. Sources of support were neither involved in the design and conduct of the study

nor were they involved in the collection, analysis, and interpretation of the data, or in the preparation, review, or approval of the article.

Role of authors: This study was conceived and designed by Z.L.B. and P.R.H. Data acquisition was performed by I.T., S.S., and Z.LB. (HLA and HIV data) as well as C.B., N.F., and B.D.W. (functional ELISpot data). C.J.B., J.M.C., C.K., and D.H. developed analytical methods and analyzed the data. D.C. performed additional data analysis. Initial drafting of the article was performed by Z.L.B. All authors contributed to interpretation of data and critical revision of the manuscript.

References

1. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB. **Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection.** *J Virol* 1994; **68**:6103–6110.
2. Koup RA, Safrit JT, Cao Y, Andrews CA, McLeod G, Borkowsky W, *et al*. **Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome.** *J Virol* 1994; **68**:4650–4655.
3. Zuñiga R, Lucchetti A, Galvan P, Sanchez S, Sanchez C, Hernandez A, *et al*. **Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control.** *J Virol* 2006; **80**:3122–3125.
4. Geldmacher C, Currier JR, Herrmann E, Haule A, Kuta E, McCutchan F, *et al*. **CD8 T-cell recognition of multiple epitopes within specific Gag regions is associated with maintenance of a low steady-state viremia in human immunodeficiency virus type 1-seropositive patients.** *J Virol* 2007; **81**:2440–2448.
5. Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. **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* 2002; **76**:2298–2305.
6. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, *et al*. **CD8+ T-cell responses to different HIV proteins have discordant associations with viral load.** *Nat Med* 2007; **13**:46–53.
7. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, *et al*. **Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition.** *Nature* 1991; **354**:453–459.
8. Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, *et al*. **Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus.** *Nat Med* 1997; **3**:205–211.
9. Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, *et al*. **Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection.** *Proc Natl Acad Sci U S A* 1997; **94**:1890–1895.
10. Goulder PJ, Watkins DI. **HIV and SIV CTL escape: implications for vaccine design.** *Nat Rev Immunol* 2004; **4**:630–640.
11. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, *et al*. **HIV evolution: CTL escape mutation and reversion after transmission.** *Nat Med* 2004; **10**:282–289.
12. Honeyborne I, Prendergast A, Pereyra F, Leslie A, Crawford H, Payne R, *et al*. **Control of human immunodeficiency virus type 1 is associated with HLA-B*13 and targeting of multiple gag-specific CD8+ T-cell epitopes.** *J Virol* 2007; **81**:3667–3672.
13. Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, *et al*. **Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1.** *J Virol* 2006; **80**:3617–3623.
14. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, Desouza I, *et al*. **Escape and compensation from early HLA-B57-mediated CTL pressure on HIV-1 Gag alters capsid interactions with cyclophilin A.** *J Virol* 2007; **81**:12608–12618.
15. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, *et al*. **Escape from the dominant HLA-B27 restricted CTL response in gag is associated with a dramatic reduction in HIV-1 replication.** *J Virol* 2007; **81**:12382–12393.
16. Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, Conlon C, *et al*. **Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses.** *J Exp Med* 2001; **193**:375–386.
17. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, *et al*. **Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS.** *Nat Med* 1997; **3**:212–217.
18. Feeney ME, Tang Y, Roosevent KA, Leslie AJ, McIntosh K, Karthas N, *et al*. **Immune escape precedes breakthrough human immunodeficiency virus type 1 viremia and broadening of the cytotoxic T-lymphocyte response in an HLA-B27-positive long-term-nonprogressing child.** *J Virol* 2004; **78**:8927–8930.
19. Oxenius A, Price DA, Trkola A, Edwards C, Gostick E, Zhang HT, *et al*. **Loss of viral control in early HIV-1 infection is temporally associated with sequential escape from CD8+ T cell responses and decrease in HIV-1-specific CD4+ and CD8+ T cell frequencies.** *J Infect Dis* 2004; **190**:713–721.
20. Bailey JR, Williams TM, Siciliano RF, Blankson JN. **Maintenance of viral suppression in HIV-1-infected HLA-B*57+ elite suppressors despite CTL escape mutations.** *J Exp Med* 2006; **203**:1357–1369.
21. Iversen AK, Stewart-Jones G, Learn GH, Christie N, Sylvester-Hviid C, Armitage AE, *et al*. **Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope.** *Nat Immunol* 2006; **7**:179–189.
22. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, *et al*. **Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1.** *PLoS Pathol* 2007; **3**:e94.
23. Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, *et al*. **Founder effects in the assessment of HIV polymorphisms and HLA allele associations.** *Science* 2007; **315**:1583–1586.
24. Hogg RS, Yip B, Chan KJ, Wood E, Craib KJ, O'Shaughnessy MV, *et al*. **Rates of disease progression by baseline CD4 cell count and viral load after initiating triple-drug therapy.** *JAMA* 2001; **286**:2568–2577.
25. Huang X, Zhang J. **Methods for comparing a DNA sequence with a protein sequence.** *Comput Appl Biosci* 1996; **12**:497–506.
26. Carlson J, Kadie C, Mallal S, Heckerman D. **Leveraging hierarchical population structure in discrete association studies.** *PLoS ONE* 2007; **2**:e591.
27. Storey JD, Tibshirani R. **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003; **100**:9440–9445.
28. Listgarten J, Heckerman D. Determining the number of non-spurious arcs in a learned DAG model: investigation of a Bayesian and a frequentist approach. In: *Proceedings of Twenty Third Conference on Uncertainty in Artificial Intelligence*. Vancouver, Canada: UAI Press; July 2007.
29. Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, et *al*., editors. *HIV Molecular Immunology 2006*. Los Alamos: Los Alamos National Laboratory, Theoretical Biology and Biophysics; 2007.
30. Heckerman D, Kadie C, Listgarten J. **Leveraging information across HLA alleles/supertypes improves epitope prediction**. *RECOMB* 2006; **LNBI 3939**: 296–308.
31. Yokomaku Y, Miura H, Tomiyama H, Kawana-Tachikawa A, Takiguchi M, Kojima A, *et al*. **Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection.** *J Virol* 2004; **78**:1324–1332.
32. Draenert R, Le Gall S, Pfafferott KJ, Leslie AJ, Chetty P, Brander C, *et al*. **Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection.** *J Exp Med* 2004; **199**:905–915.
33. Milicic A, Price DA, Zimbwa P, Booth BL, Brown HL, Easterbrook PJ, *et al*. **CD8+ T cell epitope-flanking mutations disrupt proteasomal processing of HIV-1 Nef.** *J Immunol* 2005; **175**:4618–4626.
34. Nielsen M, Lundegaard C, Lund O, Kesmir C. **The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage.** *Immunogenetics* 2005; **57**:33–41.

35. Frahm N, Korber BT, Adams CM, Szinger JJ, Draenert R, Addo MM, *et al*. **Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities.** *J Virol* 2004; **78**:2187–2200.
36. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan KM, *et al*. **Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution.** *J Virol* 2005; **79**:13239–13249.
37. Rammensee H, Bachmann J, Emmerich NP, Bachor OA, Stevanovic S. **SYFPEITHI: database for MHC ligands and peptide motifs.** *Immunogenetics* 1999; **50**:213–219.
38. Brander C, Frahm N, Walker BD. **The challenges of host and viral diversity in HIV vaccine design.** *Curr Opin Immunol* 2006; **18**:430–437.
39. Cao J, McNevin J, McSweyn M, Liu Y, Mullins JI, McElrath MJ. **Novel CTL escape by a three amino acid insertion in HIV-1 p6Pol and p6Gag late domain associated with drug resistance.** *J Virol* 2008; **82**:495–502.
40. Carrington M, O'Brien SJ. **The influence of HLA genotype on AIDS.** *Annu Rev Med* 2003; **54**:535–551.
41. Altfeld M, Kalife ET, Qi Y, Streeck H, Lichterfeld M, Johnston MN, *et al*. **HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8(+) T cell response against HIV-1.** *PLoS Med* 2006; **3**:e403.
42. Altfeld M, Allen TM. **Hitting HIV where it hurts: an alternative approach to HIV vaccine design.** *Trends Immunol* 2006; **27**:504–510.
43. Brander C, Hartman KE, Trocha AK, Jones NG, Johnson RP, Korber B, *et al*. **Lack of strong immune selection pressure by the immunodominant, HLA-A*0201-restricted cytotoxic T lymphocyte response in chronic human immunodeficiency virus-1 infection.** *J Clin Invest* 1998; **101**:2559–2566.
44. Goulder PJ, Sewell AK, Lalloo DG, Price DA, Whelan JA, Evans J, *et al*. **Patterns of immunodominance in HIV-1-specific cytotoxic T lymphocyte responses in two human histocompatibility leukocyte antigens (HLA)-identical siblings with HLA-A*0201 are influenced by epitope mutation.** *J Exp Med* 1997; **185**:1423–1433.
45. Goulder PJ, Altfeld MA, Rosenberg ES, Nguyen T, Tang Y, Eldridge RL, *et al*. **Substantial differences in specificity of HIV-specific cytotoxic T cells in acute and chronic HIV infection.** *J Exp Med* 2001; **193**:181–194.
46. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA, *et al*. **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level.** *Science* 2002; **296**:1439–1443.
47. Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, *et al*. **Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes.** *Nat Immunol* 2006; **7**:173–178.
48. Brander C, Walker BD. **Gradual adaptation of HIV to human host populations: good or bad news?** *Nat Med* 2003; **9**:1359–1362.
49. Leslie A, Kavanagh D, Honeyborne I, Pfafferott K, Edwards C, Pillay T, *et al*. **Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA.** *J Exp Med* 2005; **201**:891–902.
50. Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, Cullen C, *et al*. **Reversion of CTL escape-variant immunodeficiency viruses in vivo.** *Nat Med* 2004; **10**:275–281.
51. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, LeGall S, *et al*. **Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection.** *J Virol* 2004; **78**:7069–7078.
52. Li B, Gladden AD, Altfeld M, Kaldor JM, Cooper DA, Kelleher AD, *et al*. **Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution.** *J Virol* 2007; **81**:193–201.
53. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, *et al*. **Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA.** *Nature* 2004; **432**:769–775.
54. Turnbull EL, Lopes AR, Jones NA, Cornforth D, Newton P, Aldam D, *et al*. **HIV-1 epitope-specific CD8+ T-cell responses strongly associated with delayed disease progression cross-recognize epitope variants efficiently.** *J Immunol* 2006; **176**:6130–6146.
55. Allen TM, Yu XG, Kalife ET, Reyor LL, Lichterfeld M, John M, *et al*. **De novo generation of escape variant-specific CD8+ T-cell responses following cytotoxic T-lymphocyte escape in chronic human immunodeficiency virus type 1 infection.** *J Virol* 2005; **79**:12952–12960.
56. Letvin NL. **Correlates of immune protection and the development of a human immunodeficiency virus vaccine.** *Immunity* 2007; **27**:366–369.
57. Yang OO. **Aiming for successful vaccine-induced HIV-1-specific cytotoxic T lymphocytes.** *AIDS* 2008; **22**:325–331.
58. Johnston MI, Fauci AS. **An HIV vaccine – evolving concepts.** *N Engl J Med* 2007; **356**:2073–2081.
59. HIV vaccine failure prompts Merck to halt trial. *Nature* 2007; **449:**390.
60. Sekaly RP. **The failed HIV Merck Vaccine Study: a step back or a launching point for future vaccine development?** *J Exp Med* 2008; **205**:7–12.
61. Streeck H, Lichterfeld M, Alter G, Meier A, Teigen N, Yassine-Diab B, *et al*. **Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles.** *J Virol* 2007; **81**:7725–7731.