Payne, R.P., Branch, S., Kloverpris, H., Matthews, P.C., Koofhethile, C.K., Strong, T., Adland, E., Leitman, E., Frater, J., Ndung'u, T., Hunter, E., Haubrich, R., Mothe, B., Edwards, A., Riddell, L., Chen, F., Harrigan, P.R., Brumme, Z.L., Mallal, S., John, M., Jooste, J.P., Shapiro, R., Deeks, S.G., Walker, B.D., Brander, C., Landis, C., Carlson, J.M., Prado, J.G. and Goulder, P.J.R. (2014) Differential escape patterns within the dominant HLA-B*57:03-restricted HIV gag epitope reflect distinct Clade-Specific functional constraints. Journal of Virology, 88 (9). pp. 4668-4678.

JVI Accepts, published online ahead of print on 5 February 2014
J. Virol. doi:10.1128/JVI.03303-13

DIFFERENTIAL ESCAPE PATTERNS WITHIN THE DOMINANT HLA-B*57:03-RESTRICTED HIV GAG EPITOPE REFLECT DISTINCT CLADE-SPECIFIC FUNCTIONAL CONSTRAINTS

Payne RP[1], Branch S[2], Kløverpris H[1,3], Matthews PC[1], Koofhethile CK[1], Strong T[1], Adland E[1], Leitman E[1], Frater J[4,5], Ndung'u T[6], Hunter E[7,8], Haubrich R[9], Mothe B[10], Edwards A[11], Riddell L[12], Chen F[13], Harrigan PR[14], Brumme ZL[14,15], Mallal S[16,17], John M[16,17], Jooste JP[18], Shapiro R[19], Deeks SG[20], Walker BD[6,21], Brander C[10,22], Landis C[2], Carlson JM[23], Prado JG[10], Goulder PJR[1,6].

[1]Department of Paediatrics, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY, UK; [2]Ladymeade Reference Unit, Queen Elizabeth Hospital, Bridgetown, Barbados; [3] K-RITH, University of Kwa-Zulu Natal, Duran, South Africa; [4] Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK; [5] Oxford NIHR Biomedical Research Centre, Oxford, UK; [6] HIV Pathogenesis Program, Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, 4013, South Africa; [7] Emory Vaccine Center, Yerkes National Primate Research Centre, Emory University, Atlanta, Georgia, USA; [8] Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA; [9] Antiviral Research Center, University of California San Diego, San Diego, California, USA; [10] IrsiCaixa AIDS Research Institute - HIVCAT, Hospital Germans Trias i Pujol, Badalona, Spain; [11] Oxford Department of Genitourinary Medicine, The Churchill Hospital, Oxford, UK; [12] Department of Genitourinary Medicine, Northamptonshire Healthcare National Health Service Trust, Northampton General Hospital, Cliftonville, Northampton, UK; [13] Department of Sexual Health, Royal Berkshire Hospital, Reading, UK; [14] British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada; [15] Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada; [16] Centre for Clinical Immunology and Biomedical Statistics, Institute for Immunology and Infectious Diseases, Murdoch University, Murdoch, Western Australia, Australia; [17] Department of Clinical Immunology, Royal Perth Hospital, Perth, Western Australia, Australia; [18] Paediatric Department, Kimberley Hospital, Northern Cape, South Africa; [19] Department of Immunology and Infectious Diseases, Harvard School of Public Health, 110 Francis Street, Boston, MA 02215, USA; [20] Dept of Medicine, University of California, San Francisco, CA, USA; [21] Ragon Institute of MGH, MIT, and Harvard, AIDS Research Center, Rm. 5212D, 149 13th Street, Charlestown, MA 02129, USA; [22] Institució Catalana de Recerca i Estudis Avancats (ICREA), Barcelona, Spain; [23] Microsoft Research, eScience Group, Los Angeles CA, USA

[1] Corresponding author - rebecca.payne@hotmail.co.uk

ABSTRACT

HLA-B*57:01 and HLA-B*57:03, the most prevalent HLA-B*57 subtypes in Caucasian and African populations, respectively, are the HLA alleles most protective against HIV disease progression. Understanding the mechanisms underlying this immune control is of critical importance and yet remains unclear. Unexplained differences are observed in the impact of the dominant CTL response restricted by HLA-B*57:01 and HLA-B*57:03 in chronic infection towards the Gag epitope KAFSPEVIPMF ('KF11',Gag162-172). We previously showed that the HLA-B*57:03-KF11 response is associated with a >1 log lower viral setpoint in C-clade infection and that this response selects escape mutants within the epitope. We first examined the relationship of KF11 responses in B-clade infected subjects with HLA-B*57:01 to immune control and observed that a detectable KF11 response was associated with a >1 log higher viral load (p=0.02). No evidence of HLA-B*57:01-KF11 associated selection pressure was identified in previous comprehensive analyses of >1800 B-clade infected subjects infected. We then studied a B-clade infected cohort in Barbados where HLA-B*57:03 is highly prevalent. In contrast to B-clade infected subjects expressing HLA-B*57:01, we observed strong selection pressure driven by the HLA-B*57:03-KF11 response for the escape mutation S173T. This mutation reduces recognition of virus-infected cells by HLA-B*57:03-KF11 CTL, and is associated with a >1 log increase in viral load in HLA-B*57:03-positive subjects (p=0.009). We demonstrate functional constraints imposed by HIV clade relating to the residue at Gag-173 that explain the differential clade-specific escape patterns in HLA-B*57:03 subjects. Further studies are needed to evaluate the role of the KF11 response in HLA-B*57:01-associated HIV disease protection.

IMPORTANCE SECTION

HLA-B*57 is the HLA class I molecule that affords the greatest protection against disease progression in HIV infection. Understanding the key mechanism(s) underlying immune suppression of HIV is of importance in guiding therapeutic and vaccine-related approaches to improve the levels of HIV control occurring in nature. Numerous mechanisms have been proposed to explain the HLA associations with differential HIV disease outcome but no consensus exists. These studies focus on two subtypes of HLA-B*57, prevalent in Caucasian and African populations, HLA-B*57:01 and HLA-B*57:03, respectively. These alleles appear equally protective against HIV disease progression. The CTL epitopes presented are in many cases identical, and the dominant response in chronic infection in each case is to the Gag epitope KF11. However, there the similarity ends. This paper seeks to better understand the reasons for these differences and what this teaches us about which immune responses are contributing to immune control of HIV infection.

INTRODUCTION

HLA polymorphism has a substantial impact on HIV disease outcome (1-5) and yet the principal mechanisms underlying these effects remain unresolved (5). The most protective HLA class I molecule is HLA-B*57. HLA-B*57:01 is the most prevalent subtype in Caucasian populations, apparently conferring a similar level of protection against HIV disease progression as HLA-B*57:03, the most prevalent subtype in African populations (6). One proposal is that the HLA-B*57-mediated protection is at least in part due to the breadth of the Gag-specific CD8+ T-cell response, and that HLA-associated immune control of HIV is related to the ability of the CTL response to drive selection pressure on the virus, such that escape can only be achieved at significant cost to viral replicative capacity (5, 7-10). However, although HLA-B*57:01 and HLA-B*57:03 appear to present the identical Gag epitopes, previous studies suggest that significant differences exist in the impact of these responses on immune control.

The dominant HIV-specific CD8+ T-cell response in each case is directed towards the Gag epitope KAFSPEVIPMF ('KF11', Gag 162-172). Published studies in HLA-B*57:01-positive subjects have been almost exclusively conducted in B clade infection and these have suggested that a response to KF11 is not associated with immune control (2, 11) and that the magnitude of HLA-B*57:01-KF11 responses may even be higher in progressors. Evaluation of full-length viral sequences in 1888 B clade infected subjects failed to identify any sequence polymorphisms within the KF11 epitope or flanking it that were directly associated with HLA-B*57:01 (12) and that would have suggested strong selection pressure imposed on the virus by this response. By comparison, 20 HIV amino acid polymorphisms were identified elsewhere in the HIV proteome that were directly associated with HLA-B*57:01. In contrast, in studies undertaken in C clade infected African subjects expressing

HLA-B*57:03, a KF11 response is associated with a >1 log lower viral load (13), and there is strong evidence of selection pressure within this epitope, with approximately 70% of subjects carrying mutants at Ala-163 and/or Ser-165, positions 2 and 4 in the epitope (P2 and P4) (9, 10, 14).

The escape mutants that are selected within KF11 in C clade-infected HLA-B*57:03-positive subjects are typically A163G and S165N. The A163G mutation is selected first, reducing CTL recognition, but also significantly lowering viral replicative capacity (14). The S165N mutant is then selected, substantially restoring viral replicative capacity at the same time as entirely abrogating recognition: an ideal result for the virus (14) that has been associated with higher viral loads (15). Thus the impact of the HLA-B*57:03-KF11 response on the virus is consistent with the mechanism of HLA-mediated immune control described above, with CTL activity forcing the selection of viral escape mutants that reduce viral replicative capacity, whereas that of the HLA-B*57:01-KF11 response is not consistent with this mechanistic model.

Initial studies proposed that the HLA-B*57:01-KF11 TCRs are highly conserved, with a dominant or exclusively expressed V□5/V□19 TCR in 60-100% of subjects (15, 16). These HLA-B*57:01-KF11 TCRs were more likely than the HLA-B*57:03-KF11 TCRs to recognise epitopes containing A163G and S165N; hence these mutations would not be selected in individuals expressing HLA-B*57:01 (15, 17). However subsequent studies of HLA-B*57:01-KF11 TCR usage have been contradictory, with the V□5/V□19 TCRs identified as the dominant receptor in 0/6 and 2/10 HLA-B*57:01-positive subjects, respectively (11, 18). Thus it appears unlikely that HLA B*57 subtype-specific effects, namely a public TCR clonotype with high functional avidity for HLA-B*57:01-

KF11 (19) fully explain the observed differences in selection of KF11 escape mutants in HLA-B*57:01 and HLA-B*57:03.

An alternative hypothesis is that escape mutations within KF11 are tolerated in the context of the C-clade, but not in the B-clade Gag sequence, and that this may contribute to the high frequency of escape mutations within the KF11 epitope in HLA-B*57:03-positive individuals infected with C-clade virus. To address this possibility, we here compared a cohort of B-clade infected subjects in Barbados, in which the HLA-B*57:03-subtype predominates, with a cohort of C clade infected subjects in Botswana, where the HLA-B*57:03-subtype also predominates. We identified clade-specific sequence differences that influence the dynamics of viral escape within the HLA-B*57:03-restricted KF11 epitope. These differences were confirmed in a large multi-cohort dataset featuring 3298 subjects (including the Barbados and Botswana cohorts), including 1732 clade C infected Africans and 1566 clade B infected North Americans.

METHODS

Study cohorts and subjects

HIV-1 B clade infected subjects expressing HLA-B*57:01 were studied from ART-naïve cohorts in Oxford, UK (the Thames Valley Cohort, as previously described (20) and in Barcelona, Spain (21). Additional study cohorts for evaluation of HLA-B*57:03 in the context of B clade and C clade infection in Barbados and Botswana, respectively, were: (i) Bridgetown, Barbados (B-clade, n=246 Median age: 38 years IQR 31-47, female:male ratio 60:40, samples collected between 2008-2010), where study subjects were attendees at the Ladymeade Reference Clinic; and (ii) Gaborone, Botswana (C-clade, n=514, Median age; 27 years IQR 23-32, female:male 100:0, samples collected between 2007-2008), where study subjects were antenatal women from the Mma Bana Study, as previously described (8, 13, 22, 23).

Ethics approval was given by the Health Research Development Committee, Botswana Ministry of Health, by the Barbados Ministry of Health, the Hospital Germans Trias i Pujol' Ethics Committee, and by the Oxford Research Ethics Committee. Subjects received voluntary testing and counselling and written informed consent was obtained from all individuals. Viral load in chronic infection was measured using the Roche Amplicor version 1.5 assay; $CD4^+$ T cell counts were measured by flow cytometry. Viral load and absolute CD4 count measurements were obtained at study entry (baseline) for all individuals. All study subjects were ART-naïve.

Four-digit HLA typing of the Class I locus was performed from genomic DNA as previously described (24) by sequence-based typing at the ASHI* accredited HLA typing laboratory, University of Oklahoma Health Sciences Centre, USA. Exons 2 and 3 of HLA Class I were

amplified by locus-specific PCR and then sequenced. Resolution of ambiguities was undertaken according to the ASHI committee recommendations.

Additional viral sequence analyses were performed on two previously described, multi-center cohorts: (1) the International HIV Adaptation Collaborative (IHAC) consisting of 1443 clade B Gag sequences (12); and (2) 1470 African clade C Gag sequences from cohorts based in Durban (8), Bloemfontein (25) and Kimberley (20) South Africa, Zambia and the Thames Valley area of the United Kingdom (20). Where high-resolution HLA typing was unavailable, we employed a published machine learning algorithm trained on a dataset of high resolution HLA class I types from >13,000 individuals with known ethnicity to complete these data to high resolution. (26).

IFN-γ ELISPOT assays

IFN-γ enzyme-linked immunospot (Elispot) assays were performed as previously described (13, 27), using optimally defined epitopes and 18mer overlapping peptides (OLP) with input cells/well ranging from 30,000 to 100,000. The number of specific spot-forming cells (SFC) was calculated by subtracting the mean number of spots in the negative control wells from the number of spots counted in each well. The magnitude of epitope-specific responses was calculated as SFC per million cells.

Site-directed mutagenesis of NL43

The mutation S173T (Serine to Threonine at Gag HXB2 position 173) was introduced by site directed mutagenesis (Quikchange I, Stratagene,UK) into wild-type NL43 plasmid DNA, as well as NL43 containing the mutations A163G and/or S165N (14). Whole plasmid DNA p83-2 (the 5'-half

of HIV-1_{NL43} strain) was PCR-amplified in a mutagenesis reaction with two overlapping primers, containing the target mutation. Primers used for the mutagenesis reaction were: F 5'-CCCAGAAGTAATACCCATGTTTACGGCATTATCAGAAGGAGC-3' and R 5´-GCTCCTCTGATAATGCCGTAAACATGGGTATTACTTCTGGG-3 (the mutagenesis site is underlined) .The presence of mutations were verified by DNA Gag sequencing in newly generated plasmid clones. The DNA fragment ranging from SapI to Apa I restriction sites was then subcloned into a new p83-2 vector to avoid potential carry over of additional mutations during the mutagenesis, and the coding region sequence was verified again as previously described (28)

Virus production and Replication kinetics

Viral stocks were produced by cotransfection of the different site-directed mutant plasmids (5' half of HIV-1_{NL43} strain) with p83-10_{eGFP} (3' half of HIV-1_{NL43} strain) into MT4 cells (29). Viral stocks were harvested and viral RNA extracted (Qiagen, UK). The *gag* p24/p17 coding region was PCR amplified and sequenced to confirm the presence of the mutations in the viral RNA and the absence of any other potential polymorphisms. The 50% tissue culture-infectious dose ($TCID_{50}$) for each viral stock was determined in MT4 using the Reed and Muench method (30). For replication experiments, Jurkat, MT4 and H9 T cells were infected in triplicate with a multiplicity of infection (MOI) of 0.005 in a total volume of 3ml with wild-type or mutant HIV-1 NL43 virus and incubated at 37ºC for 2 hours. Pellets were washed twice with PBS and cultured at 37ºC and 5% of CO_2. After infection around 50,000 cells were harvested daily in order to measure infectivity by percentage of eGFP-positive cells by Fluorescent Activated Cell Sorting (FACS). Replication kinetics were determined by calculating the mean viral slope using the LOGEST function (Microsoft Excel) and

converted to natural logs. Variation in replication slopes was assessed using Student's T test. All statistical calculations were performed in Prism 5.0 (Graphpad).

Amplification and sequencing of proviral DNA

Gag p17/p24 sequences (Cohort;Barbados n=125, Botswana n=322) were generated from genomic DNA extracted from peripheral blood mononuclear cells (PBMC) where available, amplified by nested PCR using previously published primers to obtain population sequences, as previously described (31). Sequencing was undertaken using the Big Dye Ready Reaction Terminator Mix (V3.1) (Applied Biosystems, UK). Sequences were analysed using Sequencher v4.8 (Gene Codes Corporation) and aligned by SeAl to HXB2 B-clade reference strain. Sequences were submitted to Genbank and accession numbers are as follows FJ497801-FJ497875, FJ497885-FJ497899, FJ497901-FJ497905, FJ497907-FJ497916, FJ497918-FJ497950.

Identification of HLA-associated viral polymorphisms from proviral DNA

HLA-associated viral polymorphisms were identified from proviral DNA using a previously described method that corrects for phylogeny, HLA linkage disequilibrium and codon-covariation (8, 32). A q-value statistic, representing the p-value analogue of the false discovery rate (FDR), was computed for each association. The FDR is the expected proportion of false positives among the associations identified at a given p-value threshold; for example, among associations $q \leq 0.2$, we expect 20% to be false positives. The phylogenetically corrected methods rely on an inferred phylogeny. We constructed two phylogenies for this study: (i) a phylogeny consisting of clade B and C sequences from Barbados and Botswana was constructed using Phyml v2.4.5, under the general time reversible (GTR) model (33) (ii) a phylogeny consisting of N=3298 p17/p24

sequences from all cohorts described in this study. This phylogeny was too large for Phyml, so we employed a 3 stage process to infer the phylogeny. (1) A combined alignment was created, then sites with >10% missing data were removed, after which sequences with missing data in >10% of remaining sites were removed (resulting in the above noted N's); (2) a phylogeny was inferred separately for clade B and C alignments, using Phyml v2.4.5 under the GTR model); (3) the resulting phylogenies were joined by adding single common ancestor to the two clade trees, and the branch lengths were optimized using hyphy, under the GTR model (34).

PhyloDOR ratio

Identification of HLA-associated polymorphisms and assessment of differential escape between viral clades and/or closely-related HLA alleles were performed as previously described (12, 32, 35). Briefly, a maximum –likelihood phylogenetic tree was constructed for each gene, and a model of conditional adaptation was inferred for each observed amino acid at each codon (32). In this model, the amino acid is assumed to evolve independently along the phylogeny until it reaches the observed hosts (tree tips). In each host, the HLA-mediated selection pressure is modeled using a weighted logistic regression, in which the individual's HLA repertoire is used as predictors and the bias is determined by the transmitted sequence (35). Because the transmitted sequence is not observed, we average over the possible transmitted sequences, and all possible phylogenetic histories, as inferred from the phylogeny. Similarly, where high resolution HLA types are not available, we perform a weighted average over possible completions (12).

To test for differential escape between HLA-B*57:01 and B*57:03, or to test for clade-specific effects on selection, interaction variables were added to the phylogenetically-corrected logistic

regression model and significance was determined via a likelihood ratio test, as previously described (35).

Effect of S173T mutation on epitope recognition by KF11-specific CD8+ T cells

$CD4^+$ T cells were enriched from PBMCs from healthy donors expressing HLA-B*57:03 using negative selection (Dynabeads) and activated for 3-6 days using IL-2 (50U/ml Roche) and PHA (3ug/ml). KF11-specific $CD8^+$ T cells (<98% specificity) were enriched from PBMCs from HIV-infected donors using tetramers as previously described (36). B*57:03-positive $CD4^+$ T cells were infected with $NL43_{GFP}$ or $NL43_{GFP}$ containing the S173T mutation as described above. To test for epitope recognition, epitope-specific $CD8^+$ T cells (<98% specificity) were cocultured with the HIV-infected $CD4^+$ T cells in the presence of CD107a antibodies (PE-Cy5), 10ug/ml Brefeldin A, Golgi stop (BD), CD49d and CD28 for 6.5 hours at 37^{o}C in a 5% CO2 incubator. Cells were stained for surface and intracellular antibodies against CD4 (APC), CD8 (Alexa Fluor 700), MIP1B (FITC), p24 (PE), IFN-y (PE-Cy7), Live/dead marker (Pacific Blue), and then immediately acquired by FACS (BD LSRII).

RESULTS

B clade HLA-B*57:01-KF11 responders have higher viral setpoints than non-responders

Previous studies of B clade infected subjects using peptide-MHC class I tetramers have suggested that a detectable HLA-B*57:01 response is more frequently observed in progressors (including those with viral loads of >90,000) than in elite controllers/long-term non-progressors (2, 11). These studies however were not sufficiently powered to demonstrate a statistically significant result. We therefore started by comparing responses to KF11 in B clade infected, ART-naïve individuals expressing HLA-B*57:01 whose viral setpoints ranged from undetectable to 500,000 copies/ml (Fig 1). Here the association between KF11 responders and high viral setpoint reaches statistical significance (p=0.02, Mann Whitney test). These findings are consistent with the earlier studies cited of B clade infected subjects expressing HLA-B*57:01, and provide the opposite result to that obtained in HLA-B*57:03-positive individuals infected with C clade virus (13), using the identical approach of measuring IFN-g elispot responses to KF11, where a response was associated with a >10-fold lower viral setpoint. Equivalent studies of KF11 responses in 17 HLA-B*57:03-positive subjects infected with B clade virus similarly showed substantially lower median viral loads in KF11 responders compared to non-responders (median viral load 1,629 versus 6,127 c/ml, respectively), although here this difference did not reach statistical significance (p=0.28, data not shown).

Differential escape in the B*57:03-KF11 epitope in B-clade versus C-clade infection

In order to evaluate further the potential differences between HLA-B*57:01 and HLA-B*57:03, we investigated a B clade infected, ART-naïve study cohort in Barbados where HLA-B*57:03 is highly prevalent. It has been noted in several other studies that HLA-B*57:03 is associated with immune

control of HIV in B clade and C clade infection (1-5). Consistent with these studies, HLA-B*57:03-positive subjects in Barbados exhibited significantly lower median viral loads than HLA-B*57:03-negative subjects (median 3,450 versus 13,350, p=0.015, Mann Whitney test) and significantly higher $CD4^+$ counts (median 565 versus 398, p=0.003 Mann Whitney test) (Fig. 2).

To determine the nature of any selection pressure imposed in the B clade virus through the HLA-B*57:03 KF11 response, we analysed viral sequences in *gag* in the Barbados cohort in order to identify associations between HLA-B*57:03 and viral polymorphisms in the region of the KF11 epitope. This revealed that HLA-B*57:03 expression was associated with the previously described escape mutations T242N, in the epitope TW10 (TSTLQEQIGW; Gag HXB2 240-249) (7, 39), and I147X, in the epitope ISW9 (ISPRTLNAW; Gag HXB2 147-155) (Table I) (40, 41). However, the intra-epitope escape mutations within KF11 (KAFSPEVIPMF; Gag HXB2 162-172), namely A163G and S165N, selected in approximately 70% of C-clade infected HLA-B*57:03-positive subjects (3, 14), were not associated with HLA-B*57:03 in this Barbadian study cohort (Table I-II).

However, in this same Barbados cohort, we identified an HLA-B*57:03-associated viral polymorphism located at Gag HXB2 position 173, which immediately flanks the C-terminus of the KF11 epitope. This mutation has not been observed in association with HLA-B*57:03 in studies of C-clade infected cohorts, which in any case have Threonine as the consensus residue at position 173 (14, 42, 43). The high frequency of selection of S173T by HLA-B*57:03-positive subjects (61% versus 24% in HLA-B*57:03-positive versus HLA-B*57:03-negative subjects) together with the lack of any selection of intra-epitope KF11 mutations led to the hypothesis that selection of S173T

in B-clade virus may mitigate against further selection of KF11 escape mutations A163G and/or S165N (Table II) (see below).

We performed a further analysis using a phylogenetically-corrected method (12) to compare the impact of HLA-B*57:03 on the selection of Gag escape mutants in B-clade versus C-clade HIV, using data from the study cohorts in Barbados (B-clade) and in Gaborone, Botswana (C-clade). We found no statistical difference between odds of HLA-B*57:03-mediated escape in the two cohorts for T242N (p=0.82) or I147L (p=0.29). In contrast, we observed substantial clade differences for all three KF11 escape mutations: the strength of selection for A163G and S165N was significantly greater in the C clade cohort (p=0.006 and p=0.08, respectively), whereas 173T was only selected in the B clade cohort (p=0.0006). In fact, Gag-173T, the consensus in C-clade, arises at significantly *lower* frequency in HLA-B*57:03-positive subjects in Botswana (p=0.0062; discussed further below) (Table III). These data demonstrate clade-specific differences in the impact of HLA-B*57:03 on Gag escape mutant selection, with differential effects at Gag-163, 165 and 173, within or immediately flanking the dominant KF11 epitope.

Impact of S173T on recognition of virus-infected target cells and on viral setpoint

The location of the HLA-B*57:03-associated mutation immediately downstream of the KF11 epitope suggests that the S173T mutant reduces processing of the epitope. To test whether the HLA-B*57:03-associated S173T polymorphism reduces recognition of virus-infected target cells, $CD4^+$ T cells from HLA-B*57:03^+ healthy subjects were infected with NL43 HIV that was either wildtype, expressing Ser-173, or engineered to express the S173T viral polymorphism. Infected cells were incubated with HLA-B*57:03-KF11-specific $CD8^+$ T-cells (>98% specific) and the level

331 of $CD8^+$ T-cell activation monitored by CD107a and MIP1β expression. We observed that the
332 S173T mutant indeed significantly reduced recognition by the KF11-specific $CD8^+$ T cells (Fig 3,
333 p=0.0038, Student's t test). In the same assay, using $CD8^+$ T-cells specific for the HLA-B*57:03-
334 restricted Pol-specific epitope IATESIVIW ('IAW9'), no difference were observed in the level of
335 stimulation by the two viruses on the HLA-B*57:03-restricted IAW9-specific $CD8^+$ T cells (Fig 3).
336 These data support the hypothesis that S173T specifically reduces presentation of the KF11 epitope
337 by HLA-B*57:03. Furthermore, mismatched $CD4^+$ T cells induced consistently low levels of
338 stimulation confirming that activation of the KF11- and IAW9-specific $CD8^+$ T cells was HLA-
339 B*57:03-dependent.
340
341 We next examined the viral setpoints and CD4 counts in HLA-B*57:03-positive subjects with and
342 without the S173T mutation. Viral loads in HLA-B*57:03-positive subjects with the B-clade
343 wildtype, Serine at Gag-173, were more than 10-fold lower than in B*57:03-positive subjects with
344 the S173T polymorphism (median viral load 520 versus 6,905 respectively; p=0.009 Mann Whitney
345 test). Furthermore, 173S was associated with a substantially higher CD4 count in HLA-B*57:03-
346 positive subjects than 173T (median CD4 count 787 versus 375 respectively; p=0.036 Mann
347 Whitney test) (Fig. 4). However no differences in median viral load or CD4 counts were observed
348 in B*57:03-negative subjects with Serine versus Threonine at Gag-173 (median viral load 14,450
349 versus 10,600 respectively; p=0.949 and median CD4 count 358 versus 374 respectively; p=0.522
350 Mann Whitney test). These data together support the conclusion that HLA-B*57:03-KF11
351 responses drive the selection of the S173T mutation in B clade infected individuals expressing
352 HLA-B*57:03; and that this is an escape mutation in that it reduces recognition of virally infected
353 targets. These findings are consistent with the hypothesis that this response contributes to HLA-

B*57:03-associated control of HIV, since viral loads are significantly higher in those with the S173T escape mutation.

S173T with A163G and S165N significantly reduces viral replicative capacity

The observations above prompt the question: if A163G and S165N are escape mutations frequently selected in HLA-B*57:03-positive subjects infected with C clade virus, why are they not selected in HLA-B*57:03-positive subjects with B clade virus infection? To assess the functional significance of the HLA-B*57:03-associated S173T mutation and the possible impact of this polymorphism on the selection of A163G and S165N, the viral polymorphisms S173T, A163G, and S165N were introduced by site-directed mutagenesis into the B-clade backbone of $NL43_{GFP}$. Infectious viral stocks were generated by transfecting MT4 T-cells with the relevant DNA constructs. H9, MT4 or Jurkat T cells were then infected and the rate of viral growth was determined by monitoring the percentage of GFP-infected cells over 14 days.

Analysis of the rate of viral growth in MT4, H9 and Jurkat T-cells showed, first, that the S173T polymorphism had no significant effect on viral fitness in this *in vitro* system in any of these three cells lines used (Fig.5 and data not shown). We previously showed that the introduction of A163G or A163G/S165N into the NL43 backbone significantly reduced viral replicative capacity, with S165N acting as a partial compensatory mutant for A163G that also completely abrogated recognition of KF11 (14, 15). Here we observe that the introduction of either A163G or S165N into the NL43 backbone in combination with S173T also significantly reduces viral spread, but substantially more so than in the absence of S173T. Furthermore, the combination of S173T and both of the KF11 mutations, A163G and S165N, dramatically reduced viral spread even further,

377 indicating a significant cost to viral fitness of this combination of viral mutations in a B clade virus
378 (Fig. 5). These data together suggest that the KF11 escape mutant S173T is more commonly
379 selected in B clade infected subjects expressing HLA-B*57:03 because the cost to replicative
380 capacity is negligible, and less than that resulting from A163G or S165N. Subsequent mutations in
381 addition to S173T result in such a substantial reduction in replicative capacity, without any apparent
382 amelioration from S165N to reduce these fitness costs, that these arise very rarely (Table II).
383
384 As mentioned above, Gag-173T, the consensus in C-clade, arises at a significantly lower frequency
385 in HLA-B*57:03-positive compared to HLA-B*57:03-negative subjects in Botswana (p=0.0062). A
386 larger analysis of the KF11 epitope region of 1899 C-clade sequences confirmed that the presence
387 of A163G, S165N or both in combination, was significantly associated with Serine at position 173
388 and that this was the case both for HLA-B*57:03-positive and HLA-B*57:03-negative individuals
389 (Fig 6). Thus, although 173T is consensus in C-clade, it appears unfavourable in the context of the
390 KF11 intra-epitope escape mutations, supporting the findings in B-clade suggesting that this
391 combination of mutations has a detrimental impact on viral fitness.
392
393

DISCUSSION

HLA-B*57:01 and HLA-B*57:03 are the two most protective HLA molecules against HIV disease progression in both B and C clade infection (5). These molecules differ by only two amino acids (D114N and S116Y respectively), and the peptide binding motifs are almost indistinguishable (44, 45). In chronic infection, the dominant HIV-specific $CD8^+$ T cell response in subjects expressing HLA-B*57:01 or HLA-B*57:03 is to the Gag epitope KAFSPEVIPMF ('KF11', Gag HXB2 162-172) (2, 10, 13). Studies of HLA-B*57:03-positive subjects infected with C clade virus indicate that this KF11 response makes an important contribution to immune control (3, 15), and contributes to the superiority of HLA-B*57:03 as a protective HLA class I molecule over the closely-related HLA-B*57:02 and HLA-B*58:01. (10).

This study sets out to investigate the observations, first, that whilst the HLA-B*57:03-KF11 response is associated with significantly lower viral setpoints in C clade infection (13), studies of HLA-B*57:01-KF11 responses had suggested the opposite in B clade infected indiviuduals (2, 11); and, second, whereas the HLA-B*57:03-KF11 response frequently drives escape mutations within KF11 in C clade infection (A163G and S165N) (8), these are not selected in response to HLA-B*57:01 responses in B clade infection (12).

We first confirmed a statistically significant association between response to the KF11 epitope in B clade infected subjects expressing HLA-B*57:01 and a >1 log higher viral load. This result arose from the identical assays that were used in the studies that showed a KF11 response was a associated with a >1 log lower viral load in C clade infected subjects expressing HLA-B*57:03 (13). We next showed that HLA-B*57:03 has a similar impact in B and C clade infection in terms

417 of the escape mutations selected in the Gag epitopes ISW9 and TW10, but a differential impact in
418 the KF11 epitope. The S173T mutation selected in B clade infection reduces recognition of virus-
419 infected targets and is associated with a >1 log increase in viral setpoint. This is consistent with
420 studies in C clade infection (15) suggesting that this HLA-B*57:03 KF11 response contributes to
421 HLA-B*57:03-associated immune control of HIV infection. The clade specific differences in the
422 selection of KF11-driven escape mutants observed in Barbados (B clade infected cohort) and
423 Botswana (C clade infected cohort) were corroborated in analyses of larger datasets.
424
425 The position of S173T one residue downstream (P1') of the KF11 epitope suggests that it may be a
426 processing mutation since this residue would be involved in the cleavage site of the proteasome
427 (46). Previous studies of peptide cleavage motifs have suggested that the constitutive- and immuno-
428 proteasome have a strong preference for Alanine at P1' but prefer Serine over Threonine; thus the
429 mutation S173T could affect efficient cleavage of the C-terminal end of the KF11 epitope by the
430 proteasome (46).
431
432 We show that the HLA-B*57:03-associated S173T mutation effectively precludes further selection
433 of the KF11 intra-epitope viral mutations, A163G and S165N, since the combination of these three
434 mutations in a clade B backbone results in a virus with severely reduced replicative capacity.
435 Indeed, the close proximity of the amino acid positions 173, 163 and 165 between helix 1 and helix
436 2 of the Gag p24 structure, suggests that structural constraints prevent selection of A163G and
437 S165N if S173T has already been selected. Previous work has shown that, using a B-clade
438 backbone, and in the presence of S173, the mutation A163G reduces replicative capacity, but that
439 the further addition of S165N, as observed *in vivo*, partially restores replicative capacity (14, 47).

This fits with the order of selection of A163G and S165N, with S165N apparently always arising subsequent to A163G (14). However in B clade infection it appears that the selection of S173T prevents the selection of further mutants within the epitope because the fitness cost is too high. S173T appears to be the preferred choice of viral escape from the KF11-specific response since it has minimal effect on viral fitness.

Our inference from the data described above and from previous studies (23) indicating that the HLA-B*57:03-KF11 response contributes to immune control of B and C clade HIV infection. The reduced recognition of S173T-virus-infected cells by KF11-specific CTL together with the lack of cost to viral replicative capacity resulting from S173T is consistent with the observation that viral loads are higher and CD4 counts lower in B clade infected subjects expressing HLA-B*57:03.

In view of the substantial reduction in viral replicative capacity resulting from the A163G/S165N/S173T combination in B clade infection, it is perhaps surprising to observe the selection of A163G/S165N at high frequency in HLA-B*57:03-positive subjects infected with C clade virus in which the vast majority of sequences carry Thr at Gag-173. It may be inferred from this that the presence of consensus 173-Thr in the context of C clade Gag does not have the same prohibitive effect on viral fitness, as it does not prevent the selection of A163G and S165N. Nevertheless, in C clade infection, both in HLA-B*57:03-positive and HLA-B*57:03-negative individuals, A163G/S165N are significantly associated with Ser at Gag-173 (Fig. 6), as opposed to the consensus Thr at this position, suggesting that the combination of A163G/S165N/S173T is not favoured in either B or C clade infection.

462 Gag-173 has been well-studied in relation to HLA-B*27, another protective HLA molecule,
463 because of the S173A mutation that accompanies the R264K escape variant within the dominant
464 HLA-B*27-restricted epitope KRWIILGLNK (KK10) (48-50). It is noteworthy that in C clade
465 infection, R264K escape in HLA-B*27-positive subjects is typically accompanied by compensatory
466 mutations not at Gag-173, but at S165N (Brener *et al*, in preparation). These data underline the tight
467 constraints on amino acid substitutions in the capsid protein, the interdependence of residues at
468 certain key positions in the structure, including Gag-163, Gag-165, and Gag-173, and therefore the
469 impact that clade can have on the escape options for the virus.
470
471 These data help to explain why HLA-B*57:03 is not associated with the 'usual' KF11 intra-epitope
472 mutations A163G/S165N in clade B, but they do not explain why HLA-B*57:01 is not associated
473 with either the S173T flanking mutation, nor with any KF11 intra-epitope mutations. Previous
474 studies have suggested that TCR usage for the HLA-B*57:01-KF11 response allows recognition of
475 the KF11 variants (15), but these initial TCR studies indicating conservation of a 'public' HLA-
476 B*57:01-KF11 TCR have not been borne out by subsequent studies (11, 18). One possible
477 explanation is that the potency of the HLA-B*57:01-KF11 response is so great that a moderate
478 reduction in processed epitope would not affect killing sufficiently to be selected, however
479 preliminary data suggests that HLA-B*57:03 response is, if anything, the more potent. Further
480 studies with a large number of KF11-specific clones would be needed to establish whether clear-cut
481 differences between the responses restricted by HLA-B*57:03 and HLA-B*57:01 exist in terms of
482 potency, and the relevance of this to viral escape patterns. A recent study comparing the impact of
483 individual HLA class I molecules on immune control (viral load <2000 copies/ml) versus non-
484 control (VL>10,000 copies/ml) of B clade infection showed the identical odds ratio for protection

485 via HLA-B*57:01 in a European American cohort as via HLA-B*57:03 in an African American
486 cohort (6).
487
488 These studies therefore provide an explanation for the distinct clade-specific selection of escape
489 mutants by the HLA-B*57:03-KF11 response but do not resolve the question of why the HLA-
490 B*57:01-KF11 response does not select escape mutants. Insufficient studies have been undertaken
491 in C clade infected subjects who express HLA-B*57:01 to be certain of whether this response
492 selects no escape mutants in C clade as well as in B clade infection. The absence of the KF11
493 response in elite controllers with HLA-B*57:01 does not necessarily mean that these responses
494 have not contributed to immune control in these subjects, since it is possible that the period of
495 detectability may be transient. It is clear that many responses that are undetectable in elite
496 controllers can become detectable after peptide stimulation (51). However, if the KF11-specific
497 CTL response contributes to immune control of HIV in HLA-B*57:01-positive subjects in B clade
498 infection, it would be unique in failing to select escape mutants in the process and the mechanism
499 would be invaluable for directing successful vaccine targets.
500
501 ***Funding acknowledgements***
502 PJRG is supported by the Wellcome Trust and the NIH (RO1 AI046995). ZLB is the recipient of a
503 New Investigator Award from the Canadian Institutes of Health Research and a Scholar Award
504 from the Michael Smith Foundation for Health Research.
505
506
507

508

509

510

511

512

513

514

REFERENCES

1. **Kaslow RA, Carrington M, Apple R, Park L, Munoz A, Saah AJ, Goedert JJ, Winkler C, O'Brien SJ, Rinaldo C, Detels R, Blattner W, Phair J, Erlich H, Mann DL.** 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. Nature medicine **2:**405-411.
2. **Migueles SA, Sabbaghian MS, Shupert WL, Bettinotti MP, Marincola FM, Martino L, Hallahan CW, Selig SM, Schwartz D, Sullivan J, Connors M.** 2000. HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. Proceedings of the National Academy of Sciences of the United States of America **97:**2709-2714.
3. **Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klenerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature **432:**769-775.
4. **Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, Walker BD, Ripke S, Brumme CJ, Pulit SL, Carrington M, Kadie CM, Carlson JM, Heckerman D, Graham RR, Plenge RM, Deeks SG, Gianniny L, Crawford G, Sullivan J, Gonzalez E, Davies L, Camargo A, Moore JM, Beattie N, Gupta S, Crenshaw A, Burtt NP, Guiducci C, Gupta N, Gao X, Qi Y, Yuki Y, Piechocka-Trocha A, Cutrell E, Rosenberg R, Moss KL, Lemay P, O'Leary J, Schaefer T, Verma P, Toth I, Block B, Baker B, Rothchild A, Lian J, Proudfoot J, Alvino DM, Vine S, Addo MM, Allen TM, Altfeld M, Henn MR, Le Gall S, Streeck H, Haas DW, Kuritzkes DR, Robbins GK, Shafer RW, Gulick RM, Shikuma CM, Haubrich R, Riddler S, Sax PE, Daar ES, Ribaudo HJ, Agan B, Agarwal S, Ahern RL, Allen BL, Altidor S, Altschuler EL, Ambardar S, Anastos K, Anderson B, Anderson V, Andrady U, Antoniskis D, Bangsberg D, Barbaro D, Barrie W, Bartczak J, Barton S, Basden P, Basgoz N, Bazner S, Bellos NC, Benson AM, Berger J, Bernard NF, Bernard AM, Birch C, Bodner SJ, Bolan RK, Boudreaux ET, Bradley M, Braun JF, Brndjar JE, Brown SJ, Brown K, Brown ST, Burack J, Bush LM, Cafaro V, Campbell O, Campbell J, Carlson RH, Carmichael JK, Casey KK, Cavacuiti C, Celestin G, Chambers ST, Chez N, Chirch LM, Cimoch PJ, Cohen D, Cohn LE, Conway B, Cooper DA, Cornelson B, Cox DT, Cristofano MV, Cuchural G, Jr., Czartoski JL, Dahman JM, Daly JS, Davis BT, Davis K, Davod SM, DeJesus E, Dietz CA, Dunham E, Dunn ME, Ellerin TB, Eron JJ, Fangman JJ, Farel CE, Ferlazzo H, Fidler S, Fleenor-Ford A, Frankel R, Freedberg KA, French NK, Fuchs JD, Fuller JD, Gaberman J, Gallant JE, Gandhi RT, Garcia E, Garmon D, Gathe JC, Jr., Gaultier CR, Gebre W, Gilman FD, Gilson I, Goepfert PA, Gottlieb MS, Goulston C, Groger RK, Gurley TD, Haber S, Hardwicke R, Hardy WD, Harrigan PR, Hawkins TN, Heath S, Hecht FM, Henry WK, Hladek M, Hoffman RP, Horton JM, Hsu RK, Huhn GD, Hunt P, Hupert MJ, Illeman ML, Jaeger H, Jellinger RM, John M, Johnson JA, Johnson KL, Johnson H, Johnson K, Joly J, Jordan WC, Kauffman CA, Khanlou H, Killian RK, Kim AY, Kim DD, Kinder CA, Kirchner JT, Kogelman L, Kojic EM, Korthuis PT, Kurisu W, Kwon DS, LaMar M, Lampiris H, Lanzafame M, Lederman MM, Lee DM, Lee JM, Lee MJ, Lee ET, Lemoine J, Levy JA, Llibre JM, Liguori MA,**

Little SJ, Liu AY, Lopez AJ, Loutfy MR, Loy D, Mohammed DY, Man A, Mansour MK, Marconi VC, Markowitz M, Marques R, Martin JN, Martin HL, Jr., Mayer KH, McElrath MJ, McGhee TA, McGovern BH, McGowan K, McIntyre D, McLeod GX, Menezes P, Mesa G, Metroka CE, Meyer-Olson D, Miller AO, Montgomery K, Mounzer KC, Nagami EH, Nagin I, Nahass RG, Nelson MO, Nielsen C, Norene DL, O'Connor DH, Ojikutu BO, Okulicz J, Oladehin OO, Oldfield EC, 3rd, Olender SA, Ostrowski M, Owen WF, Jr., Pae E, Parsonnet J, Pavlatos AM, Perlmutter AM, Pierce MN, Pincus JM, Pisani L, Price LJ, Proia L, Prokesch RC, Pujet HC, Ramgopal M, Rathod A, Rausch M, Ravishankar J, Rhame FS, Richards CS, Richman DD, Rodes B, Rodriguez M, Rose RC, 3rd, Rosenberg ES, Rosenthal D, Ross PE, Rubin DS, Rumbaugh E, Saenz L, Salvaggio MR, Sanchez WC, Sanjana VM, Santiago S, Schmidt W, Schuitemaker H, Sestak PM, Shalit P, Shay W, Shirvani VN, Silebi VI, Sizemore JM, Jr., Skolnik PR, Sokol-Anderson M, Sosman JM, Stabile P, Stapleton JT, Starrett S, Stein F, Stellbrink HJ, Sterman FL, Stone VE, Stone DR, Tambussi G, Taplitz RA, Tedaldi EM, Telenti A, Theisen W, Torres R, Tosiello L, Tremblay C, Tribble MA, Trinh PD, Tsao A, Ueda P, Vaccaro A, Valadas E, Vanig TJ, Vecino I, Vega VM, Veikley W, Wade BH, Walworth C, Wanidworanun C, Ward DJ, Warner DA, Weber RD, Webster D, Weis S, Wheeler DA, White DJ, Wilkins E, Winston A, Wlodaver CG, van't Wout A, Wright DP, Yang OO, Yurdin DL, Zabukovic BW, Zachary KC, Zeeman B, Zhao M. 2010. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science (New York, N.Y **330:**1551-1557.

5. **Goulder PJ, Walker BD.** 2012. HIV and HLA class I: an evolving relationship. Immunity **37:**426-440.

6. **Apps R, Qi Y, Carlson JM, Chen H, Gao X, Thomas R, Yuki Y, Del Prete GQ, Goulder P, Brumme ZL, Brumme CJ, John M, Mallal S, Nelson G, Bosch R, Heckerman D, Stein JL, Soderberg KA, Moody MA, Denny TN, Zeng X, Fang J, Moffett A, Lifson JD, Goedert JJ, Buchbinder S, Kirk GD, Fellay J, McLaren P, Deeks SG, Pereyra F, Walker B, Michael NL, Weintrob A, Wolinsky S, Liao W, Carrington M.** 2013. Influence of HLA-C expression level on HIV control. Science (New York, N.Y **340:**87-91.

7. **Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, Altfeld M, Brander C, Dixon C, Ramduth D, Jeena P, Thomas SA, St John A, Roach TA, Kupfer B, Luzzi G, Edwards A, Taylor G, Lyall H, Tudor-Williams G, Novelli V, Martinez-Picado J, Kiepiela P, Walker BD, Goulder PJ.** 2004. HIV evolution: CTL escape mutation and reversion after transmission. Nature medicine **10:**282-289.

8. **Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, Rousseau C, Rolland M, Honeyborne I, Carlson J, Kadie C, Brander C, Bishop K, Mlotshwa N, Mullins JI, Coovadia H, Ndung'u T, Walker BD, Heckerman D, Goulder PJ.** 2008. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. Journal of virology **82:**8548-8559.

9. **Crawford H, Lumm W, Leslie A, Schaefer M, Boeras D, Prado JG, Tang J, Farmer P, Ndung'u T, Lakhi S, Gilmour J, Goepfert P, Walker BD, Kaslow R, Mulenga J, Allen S, Goulder PJ, Hunter E.** 2009. Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. The Journal of experimental medicine **206:**909-921.

10. **Kloverpris HN, Stryhn A, Harndahl M, van der Stok M, Payne RP, Matthews PC, Chen F, Riddell L, Walker BD, Ndung'u T, Buus S, Goulder P.** 2012. HLA-B*57 Micropolymorphism shapes HLA allele-specific epitope immunogenicity, selection pressure, and HIV immune control. Journal of virology **86:**919-929.
11. **Mendoza D, Royce C, Ruff LE, Ambrozak DR, Quigley MF, Dang T, Venturi V, Price DA, Douek DC, Migueles SA, Connors M.** 2012. HLA B*5701-positive long-term nonprogressors/elite controllers are not distinguished from progressors by the clonal composition of HIV-specific CD8+ T cells. Journal of virology **86:**4014-4018.
12. **Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, Le AQ, Chui CK, Cotton LA, Knapp DJ, Riddler SA, Haubrich R, Nelson G, Pfeifer N, Deziel CE, Heckerman D, Apps R, Carrington M, Mallal S, Harrigan PR, John M, Brumme ZL.** 2012. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. Journal of virology **86:**13202-13216.
13. **Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, Bishop K, van der Stok M, Nair K, Khan N, Crawford H, Payne R, Leslie A, Prado J, Prendergast A, Frater J, McCarthy N, Brander C, Learn GH, Nickle D, Rousseau C, Coovadia H, Mullins JI, Heckerman D, Walker BD, Goulder P.** 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nature medicine **13:**46-53.
14. **Crawford H, Prado JG, Leslie A, Hue S, Honeyborne I, Reddy S, van der Stok M, Mncube Z, Brander C, Rousseau C, Mullins JI, Kaslow R, Goepfert P, Allen S, Hunter E, Mulenga J, Kiepiela P, Walker BD, Goulder PJ.** 2007. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. Journal of virology **81:**8346-8351.
15. **Yu XG, Lichterfeld M, Chetty S, Williams KL, Mui SK, Miura T, Frahm N, Feeney ME, Tang Y, Pereyra F, Labute MX, Pfafferott K, Leslie A, Crawford H, Allgaier R, Hildebrand W, Kaslow R, Brander C, Allen TM, Rosenberg ES, Kiepiela P, Vajpayee M, Goepfert PA, Altfeld M, Goulder PJ, Walker BD.** 2007. Mutually exclusive T-cell receptor induction and differential susceptibility to human immunodeficiency virus type 1 mutational escape associated with a two-amino-acid difference between HLA class I subtypes. Journal of virology **81:**1619-1631.
16. **Gillespie GM, Stewart-Jones G, Rengasamy J, Beattie T, Bwayo JJ, Plummer FA, Kaul R, McMichael AJ, Easterbrook P, Dong T, Jones EY, Rowland-Jones SL.** 2006. Strong TCR conservation and altered T cell cross-reactivity characterize a B*57-restricted immune response in HIV-1 infection. J Immunol **177:**3893-3902.
17. **Stewart-Jones GB, Simpson P, van der Merwe PA, Easterbrook P, McMichael AJ, Rowland-Jones SL, Jones EY, Gillespie GM.** 2012. Structural features underlying T-cell receptor sensitivity to concealed MHC class I micropolymorphisms. Proceedings of the National Academy of Sciences of the United States of America **109:**E3483-3492.
18. **Simons BC, Vancompernolle SE, Smith RM, Wei J, Barnett L, Lorey SL, Meyer-Olson D, Kalams SA.** 2008. Despite biased TRBV gene usage against a dominant HLA B57-restricted epitope, TCR diversity can provide recognition of circulating epitope variants. J Immunol **181:**5137-5146.
19. **Berger CT, Frahm N, Price DA, Mothe B, Ghebremichael M, Hartman KL, Henry LM, Brenchley JM, Ruff LE, Venturi V, Pereyra F, Sidney J, Sette A, Douek DC,**

652 **Walker BD, Kaufmann DE, Brander C.** 2011. High-functional-avidity cytotoxic T
653 lymphocyte responses to HLA-B-restricted Gag-derived epitopes associated with relative
654 HIV control. Journal of virology **85:**9334-9345.
655 20. **Matthews PC, Adland E, Listgarten J, Leslie A, Mkhwanazi N, Carlson JM, Harndahl**
656 **M, Stryhn A, Payne RP, Ogwu A, Huang KH, Frater J, Paioni P, Kloverpris H, Jooste**
657 **P, Goedhals D, van Vuuren C, Steyn D, Riddell L, Chen F, Luzzi G, Balachandran T,**
658 **Ndung'u T, Buus S, Carrington M, Shapiro R, Heckerman D, Goulder PJ.** 2011. HLA-
659 A*7401-mediated control of HIV viremia is independent of its linkage disequilibrium with
660 HLA-B*5703. J Immunol **186:**5675-5686.
661 21. **Mothe B, Llano A, Ibarrondo J, Zamarreno J, Schiaulini M, Miranda C, Ruiz-Riol M,**
662 **Berger CT, Herrero MJ, Palou E, Plana M, Rolland M, Khatri A, Heckerman D,**
663 **Pereyra F, Walker BD, Weiner D, Paredes R, Clotet B, Felber BK, Pavlakis GN,**
664 **Mullins JI, Brander C.** 2012. CTL responses of high functional avidity and broad variant
665 cross-reactivity are associated with HIV control. PloS one **7:**e29717.
666 22. **Shapiro RL, Hughes MD, Ogwu A, Kitch D, Lockman S, Moffat C, Makhema J, Moyo**
667 **S, Thior I, McIntosh K, van Widenfelt E, Leidner J, Powis K, Asmelash A, Tumbare**
668 **E, Zwerski S, Sharma U, Handelsman E, Mburu K, Jayeoba O, Moko E, Souda S,**
669 **Lubega E, Akhtar M, Wester C, Tuomola R, Snowden W, Martinez-Tristani M,**
670 **Mazhani L, Essex M.** 2010. Antiretroviral regimens in pregnancy and breast-feeding in
671 Botswana. The New England journal of medicine **362:**2282-2294.
672 23. **Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, Brumme CJ, Wang**
673 **B, Losina E, Miura T, Chonco F, van der Stok M, Mncube Z, Bishop K, Goulder PJ,**
674 **Walker BD, Brockman MA, Ndung'u T.** 2010. Gag-protease-mediated replication
675 capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical
676 parameters. Journal of virology **84:**10820-10831.
677 24. **Bunce M, O'Neill CM, Barnardo MC, Krausa P, Browning MJ, Morris PJ, Welsh KI.**
678 1995. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4,
679 DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-
680 SSP). Tissue antigens **46:**355-367.
681 25. **Huang KH, Goedhals D, Fryer H, van Vuuren C, Katzourakis A, De Oliveira T,**
682 **Brown H, Cassol S, Seebregts C, McLean A, Klenerman P, Phillips R, Frater J.** 2009.
683 Prevalence of HIV type-1 drug-associated mutations in pre-therapy patients in the Free
684 State, South Africa. Antiviral therapy **14:**975-984.
685 26. **Listgarten J, Brumme Z, Kadie C, Xiaojiang G, Walker B, Carrington M, Goulder P,**
686 **Heckerman D.** 2008. Statistical resolution of ambiguous HLA typing data. PLoS
687 computational biology **4:**e1000016.
688 27. **Altfeld MA, Trocha A, Eldridge RL, Rosenberg ES, Phillips MN, Addo MM, Sekaly**
689 **RP, Kalams SA, Burchett SA, McIntosh K, Walker BD, Goulder PJ.** 2000.
690 Identification of dominant optimal HLA-B60- and HLA-B61-restricted cytotoxic T-
691 lymphocyte (CTL) epitopes: rapid characterization of CTL responses by enzyme-linked
692 immunospot assay. Journal of virology **74:**8541-8549.
693 28. **Prado JG, Honeyborne I, Brierley I, Puertas MC, Martinez-Picado J, Goulder PJ.**
694 2009. Functional consequences of human immunodeficiency virus escape from an HLA-
695 B*13-restricted CD8+ T-cell epitope in p1 Gag protein. Journal of virology **83:**1018-1025.
696 29. **Weber J, Weberova J, Carobene M, Mirza M, Martinez-Picado J, Kazanjian P,**
697 **Quinones-Mateu ME.** 2006. Use of a novel assay based on intact recombinant viruses

expressing green (EGFP) or red (DsRed2) fluorescent proteins to examine the contribution of pol and env genes to overall HIV-1 replicative fitness. Journal of virological methods **136:**102-117.

30. **Muench L.J. aHR.** 1938. A simple method of estimating fifty percent endpoints. Am. J. Hyg.**:**493-497.
31. **Leslie A, Kavanagh D, Honeyborne I, Pfafferott K, Edwards C, Pillay T, Hilton L, Thobakgale C, Ramduth D, Draenert R, Le Gall S, Luzzi G, Edwards A, Brander C, Sewell AK, Moore S, Mullins J, Moore C, Mallal S, Bhardwaj N, Yusim K, Phillips R, Klenerman P, Korber B, Kiepiela P, Walker B, Goulder P.** 2005. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. The Journal of experimental medicine **201:**891-902.
32. **Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JI, Walker BD, Harrigan PR, Goulder PJ, Heckerman D.** 2008. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS computational biology **4:**e1000225.
33. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic biology **52:**696-704.
34. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676-679.
35. **Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, Walker BD, Ndung'u T, Shapiro R, Frater J, Brumme ZL, Goulder PJ, Heckerman D.** 2012. Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. Journal of virology **86:**5230-5243.
36. **Payne RP, Kloverpris H, Sacha JB, Brumme Z, Brumme C, Buus S, Sims S, Hickling S, Riddell L, Chen F, Luzzi G, Edwards A, Phillips R, Prado JG, Goulder PJ.** Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. Journal of virology **84:**10543-10557.
37. **Sacha JB, Chung C, Reed J, Jonas AK, Bean AT, Spencer SP, Lee W, Vojnov L, Rudersdorf R, Friedrich TC, Wilson NA, Lifson JD, Watkins DI.** 2007. Pol-specific CD8+ T cells recognize simian immunodeficiency virus-infected cells prior to Nef-mediated major histocompatibility complex class I downregulation. Journal of virology **81:**11703-11712.
38. **Payne RP, Kloverpris H, Sacha JB, Brumme Z, Brumme C, Buus S, Sims S, Hickling S, Riddell L, Chen F, Luzzi G, Edwards A, Phillips R, Prado JG, Goulder PJ.** 2010. Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. Journal of virology **84:**10543-10557.
39. **Feeney ME, Tang Y, Roosevelt KA, Leslie AJ, McIntosh K, Karthas N, Walker BD, Goulder PJ.** 2004. 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. Journal of virology **78:**8927-8930.
40. **Draenert R, Le Gall S, Pfafferott KJ, Leslie AJ, Chetty P, Brander C, Holmes EC, Chang SC, Feeney ME, Addo MM, Ruiz L, Ramduth D, Jeena P, Altfeld M, Thomas S, Tang Y, Verrill CL, Dixon C, Prado JG, Kiepiela P, Martinez-Picado J, Walker BD, Goulder PJ.** 2004. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. The Journal of experimental medicine **199:**905-915.

41. **Boutwell CL, Rowley CF, Essex M.** 2009. Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein. Journal of virology **83:**2460-2468.
42. **Chopera DR, Woodman Z, Mlisana K, Mlotshwa M, Martin DP, Seoighe C, Treurnicht F, de Rosa DA, Hide W, Karim SA, Gray CM, Williamson C.** 2008. Transmission of HIV-1 CTL escape variants provides HLA-mismatched recipients with a survival advantage. PLoS pathogens **4:**e1000033.
43. **Huang KH, Goedhals D, Carlson JM, Brockman MA, Mishra S, Brumme ZL, Hickling S, Tang CS, Miura T, Seebregts C, Heckerman D, Ndung'u T, Walker B, Klenerman P, Steyn D, Goulder P, Phillips R, van Vuuren C, Frater J.** 2011. Progression to AIDS in South Africa Is Associated with both Reverting and Compensatory Viral Mutations. PloS one **6:**e19018.
44. **Barber LD, Percival L, Arnett KL, Gumperz JE, Chen L, Parham P.** 1997. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. J Immunol **158:**1660-1669.
45. **Ostrov DA, Grant BJ, Pompeu YA, Sidney J, Harndahl M, Southwood S, Oseroff C, Lu S, Jakoncic J, de Oliveira CA, Yang L, Mei H, Shi L, Shabanowitz J, English AM, Wriston A, Lucas A, Phillips E, Mallal S, Grey HM, Sette A, Hunt DF, Buus S, Peters B.** 2012. Drug hypersensitivity caused by alteration of the MHC-presented self-peptide repertoire. Proceedings of the National Academy of Sciences of the United States of America **109:**9959-9964.
46. **Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, Laplace C, Zwinderman A, Dick TP, Muller J, Schonfisch B, Schmid C, Fehling HJ, Stevanovic S, Rammensee HG, Schild H.** 2001. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. The Journal of experimental medicine **194:**1-12.
47. **Boutwell CL, Carlson JM, Lin TH, Seese A, Power KA, Peng J, Tang Y, Brumme ZL, Heckerman D, Schneidewind A, Allen TM.** 2013. Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins. Journal of virology **87:**3952-3965.
48. **Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, Conlon C, Workman C, Shaunak S, Olson K, Goulder P, Brander C, Ogg G, Sullivan JS, Dyer W, Jones I, McMichael AJ, Rowland-Jones S, Phillips RE.** 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. The Journal of experimental medicine **193:**375-386.
49. **Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, Rinaldo CR, Craggs SL, Allgaier RL, Power KA, Kuntzen T, Tung CS, LaBute MX, Mueller SM, Harrer T, McMichael AJ, Goulder PJ, Aiken C, Brander C, Kelleher AD, Allen TM.** 2007. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. Journal of virology **81:**12382-12393.
50. **Schneidewind A, Brockman MA, Sidney J, Wang YE, Chen H, Suscovich TJ, Li B, Adam RI, Allgaier RL, Mothe BR, Kuntzen T, Oniangue-Ndza C, Trocha A, Yu XG, Brander C, Sette A, Walker BD, Allen TM.** 2008. Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-

restricted epitope in human immunodeficiency virus type 1 capsid. Journal of virology **82:**5594-5605.

51. **Ndhlovu ZM, Proudfoot J, Cesa K, Alvino DM, McMullen A, Vine S, Stampouloglou E, Piechocka-Trocha A, Walker BD, Pereyra F.** 2012. Elite controllers with low to absent effector CD8+ T cell responses maintain highly functional, broadly directed central memory responses. Journal of virology **86:**6959-6969.

798 **FIGURE LEGENDS**

799 **Figure 1. Median Viral of KF11 responders versus nonresponders in B clade-infected**
800 **individuals expressing HLA-B*57:01**

801 PBMCs from B clade-infected, ART-naïve individuals expressing HLA-B*57:01 were analysed by
802 IFN-γ ELISPOT assay for responses to the KF11 epitope. Viral loads of responders and
803 nonresponders were compared. Mann-Whitney U tests were performed.

804

805 **Figure 2. Median Viral Load and CD4 count of B*57:03-positive versus B*57:03-negative**
806 **subjects in Barbados cohort (B-clade).**

807 B*57:03-positive subjects were compared to B*57:03-negative subjects for (A) Viral Load and (B)
808 CD4 count. Median and 5-95 percentiles are shown. Mann-Whitney U tests were performed.

809

810 **Figure 3.Effect of viral mutation S173T on epitope recognition of HIV-infected cells by KF11-**
811 **specific $CD8^+$ T-cells**

812 *Ex vivo* $CD4^+$ T-cells from $B*57:03^+$ and $B*57:03^-$ donors were infected with wildtype NL43 virus
813 or NL43 virus harbouring the S173T viral mutation. Infected $CD4^+$ T-cells were then cultured with
814 KF11-specific $CD8^+$ T-cells (A) or IAW9-specific $CD8^+$ T-cells (B) and the level of $CD8^+$ T-cell
815 activation monitored by expression of CD107 and Mip1β. Data from both experiments were
816 standardized relative to % recognition by wildtype virus (C). Experiments were performed in
817 triplicate, mean and SD are shown. Student's t test were performed. p value summary is as follows;
818 $p<0.01$ *, $p<0.001$ **, $p<0.0001$ ***.

819

Figure 4. Median Viral Load and CD4 count of B*57:03-positive HIV-infected subjects with viral polymorphisms S173 and T173.

Proviral DNA sequences from B*57:03-positive subjects from the Barbados cohort (B-clade) were analysed for the presence of the viral polymorphisms S173 and T173. Viral Loads (A) and CD4 counts (B) were compared. Median and 5-95 percentiles are shown. Mann-Whitney t tests were performed.

Figure 5. Viral replication capacity of $NL43_{GFP}$ virus with multiple B*57:03-associated viral mutations

$NL43_{GFP}$ virus was engineered to contain combinations of the viral mutations 173T, 163G and 165N. MT4 cells were infected and monitored for GFP-positive cells over 14 days (A). The slope of the curve was calculated from the exponential growth phase using the LOGEST function and converted to natural logs (B). Experiments were performed in triplicate and mean and SD are shown. Dunnett's multiple comparison tests were performed. p value summary is as follows; $p<0.01$ *, $p<0.001$ **, $p<0.0001$ ***.

Figure 6. Frequency of KF11 mutations (A163G and S165N) and S173T in C-clade *gag* sequences (n=1899)

HIV-1 p24 *gag* sequences (n=1899) were analysed for the presence of the KF11 mutations, A163G and S165N, in the presence of 173T and 173S. B*57:03-positive subjects (A) and B*57-negative subjects (B) were analysed. Fisher's exact tests were performed.

Table I. Summary of HLA-B*57:03-associated viral mutations in Barbados cohort (n=125).

A phylogenetically-corrected method was used to determine the location of HLA-B*57:03-associated viral mutations in proviral *gag* p17 and p24. Polymorphism location and HXB2 number are shown. A q value cut-off of q<0.2 was used.

Table II. Summary of amino acid polymorphisms in the KF11 epitope region (HXB2 162-173) in Barbados cohort (n=125).

Proviral sequences were grouped into those from HLA-B*57:03-positive and B*57:03-negative donors. A summary of all HIV-1 polymorphisms is shown for HXB2 region 162-173.

Table III Summary of PhyloDOR of B*57:03-associated mutations located in HIV-1 p24 *gag* between Barbados [BB] and Botswana [BW] cohorts.

Proviral sequences from Barbados (n=125) and Botswana (n=322) were analysed from HLA-B*57:03-positive and -B*57:03-negative individuals for the presence of B*57:03-associated viral mutations located in known viral epitopes in p24 *gag*. HXB2 position of mutations is shown. p values were calculated from phylogenetically-corrected Odds ratios (PhyloDOR) using an online tool (35).

866
867
868
869

870 Table I
871
872

Protein	HLA	Polymorphism location	HXB2	p value	q value
p17	B*5703	A S V L S G G E L D R W E K I R L R P G G	15	0.00211	0.115
	B*5703	Q P S L Q T G S E E L K S L Y N T V A T L	75	0.00388	0.14
p24	B*5703	N L Q G Q M V H Q A I S P R T L N A W V K	147	0.00544	0.169
	B*5703	A F S P E V I P M F S A L S E G A T P Q D	173	0.000301	0.0218
	B*5703	R G S D I A G T T S T L Q E Q I G W M T N	242	3.02E-11	6.56E-09

881
882
883
884
885
886
887

888 Table II
889

890

HXB2 position 162-173	B*5703 positive (n=15)		B*5703 negative (n=110)	
	n	%	n	%
K A F S P E V I P M F S	1	7	77	70
- - - - - - - - - - - T	7	47	25	23
- G - N - - - - - - - -	0	0	1	1
- S - - - - I - - - - -	1	7	0	0
- S - - - - - - - - - -	1	7	1	1
- - - K - - - - - - - -	1	7	0	0
- - - N - - - - - - - -	0	0	2	2
- - - - - - I - - - - -	0	0	1	1
- N - - - - - - - - - T	0	0	1	1
- S - - - - - - - - - T	1	7	0	0
- - - - - - I - - - - T	1	7	0	0
- - - N - - - - - - - I	1	7	0	0
- - - - - - - - - - - I	1	7	0	0
- - - - - - - - - - - A	0	0	1	1
- G - N - - - - - - - A	0	0	1	1

908
909
910
911
912
913
914
915
916
917

918 Table III

919

X (HXB2)	Country	**Log(PhyloDOR)		*Difference?	
			p-value	Y/N	p-value
242N	Botswana	4.9	3.26E-07	N	0.815
	Barbados	11.67	6.15E-14		
147L	Botswana	10.5	4.99E-05	N	0.2937
	Barbados	13.26	4.00E-03		
163G	Botswana	13.22	0.0003	Y	0.0067
	Barbados	9.343	0.2006		
165N	Botswana	4.213	0.0025	N	0.0818
	Barbados	11.88	0.0887		
173T	Botswana	-2.798	0.0062	Y	0.0006
	Barbados	13.52	0.0033		

*Is B*5703's effect on X different in Botswana versus Barbados?

**Phylogenetically-corrected Odds ratio

929