LETTERS

Adaptation of HIV-1 to human leukocyte antigen class I

Yuka Kawashima[1], Katja Pfafferott[3,6], John Frater[4,5], Philippa Matthews[3], Rebecca Payne[3], Marylyn Addo[7], Hiroyuki Gatanaga[2,8], Mamoru Fujiwara[1], Atsuko Hachiya[1,8], Hirokazu Koizumi[1], Nozomi Kuse[1], Shinichi Oka[2,8], Anna Duda[4,5], Andrew Prendergast[3], Hayley Crawford[3], Alasdair Leslie[3], Zabrina Brumme[7], Chanson Brumme[7], Todd Allen[7], Christian Brander[7,9], Richard Kaslow[10], James Tang[10], Eric Hunter[11], Susan Allen[12], Joseph Mulenga[12], Songee Branch[13], Tim Roach[13], Mina John[6], Simon Mallal[6], Anthony Ogwu[14], Roger Shapiro[14], Julia G. Prado[3], Sarah Fidler[15], Jonathan Weber[15], Oliver G. Pybus[16], Paul Klenerman[4,5], Thumbi Ndung'u[17], Rodney Phillips[4,5], David Heckerman[19], P. Richard Harrigan[18], Bruce D. Walker[7,17,20], Masafumi Takiguchi[1] & Philip Goulder[3,6,17]

The rapid and extensive spread of the human immunodeficiency virus (HIV) epidemic provides a rare opportunity to witness host–pathogen co-evolution involving humans. A focal point is the interaction between genes encoding human leukocyte antigen (HLA) and those encoding HIV proteins. HLA molecules present fragments (epitopes) of HIV proteins on the surface of infected cells to enable immune recognition and killing by $CD8^+$ T cells; particular HLA molecules, such as HLA-B*57, HLA-B*27 and HLA-B*51, are more likely to mediate successful control of HIV infection[1]. Mutation within these epitopes can allow viral escape from $CD8^+$ T-cell recognition. Here we analysed viral sequences and HLA alleles from >2,800 subjects, drawn from 9 distinct study cohorts spanning 5 continents. Initial analysis of the HLA-B*51-restricted epitope, TAFTIPSI (reverse transcriptase residues 128–135), showed a strong correlation between the frequency of the escape mutation I135X and HLA-B*51 prevalence in the 9 study cohorts ($P = 0.0001$). Extending these analyses to incorporate other well-defined $CD8^+$ T-cell epitopes, including those restricted by HLA-B*57 and HLA-B*27, showed that the frequency of these epitope variants ($n = 14$) was consistently correlated with the prevalence of the restricting HLA allele in the different cohorts (together, $P < 0.0001$), demonstrating strong evidence of HIV adaptation to HLA at a population level. This process of viral adaptation may dismantle the well-established HLA associations with control of HIV infection that are linked to the availability of key epitopes, and highlights the challenge for a vaccine to keep pace with the changing immunological landscape presented by HIV.

The extent to which HIV is evolving at the population level in response to immune selection pressure is under debate[2–6]. Resolving the impact of HLA class I alleles on viral evolution is problematic because it can be obscured by other influences, such as founder effect[6] (polymorphisms present within the early strains establishing the epidemic in a group). In addition, most HLA alleles do not drive significant selection pressure on HIV, a proportion of escape mutations revert to wild type after transmission, and different HLA alleles may drive the identical escape mutation[7].

To test the hypothesis that the frequency of escape mutations in a given population is correlated with the prevalence of the relevant HLA allele in that population, we studied nine distinct cohorts from North America, the Caribbean, Europe, sub-Saharan Africa, Australia and Japan, in which we performed HLA typing, and defined the viral mutations arising within $CD8^+$ T-cell epitopes. We focused initially on a well-characterized mutation, I135X, within the HLA-B*51-restricted epitope, TAFTIPSI (RT 128–135)[8], because it arises in acute infection, non-HLA-B*51 alleles do not also select this mutation[7,9], and it does not revert to Ile 135 after transmission to HLA-B*51-negative subjects[9]. Thus, if highly prevalent HLA alleles drive a high frequency of escape mutations in the population, this would be most obvious in relation to HLA-B*51 and the escape mutant I135X. We then considered an additional 13 well-defined escape mutations, including those known to reduce viral fitness and therefore liable to revert after transmission.

I135X was selected in 205 of 213 (96%) HLA-B*51-positive individuals analysed (Figs 1 and 2, and Supplementary Fig. 1). The I135X variants do not significantly affect viral replicative capacity *in vitro*, other than the rare I135V mutation. This was the only variant observed to revert to wild-type *in vivo* during a 3-year follow-up of 38 HLA-B*51-negative subjects identified during acute HIV infection who carried I135X mutant viruses at transmission (Fig. 1e). The I135X mutants substantially affect HLA binding, and therefore also recognition by $CD8^+$ T cells (Fig. 1f–h). Thus, HIV transmission from HLA-B*51-positive subjects would probably involve transmission of I135X, which would persist in the new host. Newly infected HLA-B*51-positive subjects receiving an I135X mutant would be unable to generate an HLA-B*51-TAFTIPSI-specific response.

To test the hypothesis that the population frequency of I135X is correlated with HLA-B*51 prevalence, HIV sequence and HLA data were collated from the nine study cohorts. One cohort comprised subjects with acute/early HIV infection; the remaining cohorts comprised chronically infected subjects. In all cohorts the odds ratio strongly favoured I135X in the HLA-B*51-positive subjects, even in the acute cohort where I135X was selected sufficiently early to be already over-represented in HLA-B*51-positive subjects (odds ratio 1.65, $P = 0.07$, Fig. 2a). In Japan, where HLA-B*51 is highly

[1]Divisions of Viral Immunology and [2]Infectious Disease, Center for AIDS Research, Kumamoto University, 2-2-1 Honjo, Kumamoto 860-0811, Japan. [3]Department of Paediatrics, [4]Nuffield Department of Clinical Medicine and [5]The James Martin 21st Century School, Peter Medawar Building for Pathogen Research, South Parks Road, Oxford OX1 3SY, UK. [6]Centre for Clinical Immunology and Biomedical Statistics, Royal Perth Hospital and Murdoch University, Western Australia 6000, Australia. [7]Partners AIDS Research Center, Massachusetts General Hospital, 13th Street, Building 149, Charlestown, Boston, Massachusetts 02129, USA. [8]AIDS Clinical Center, International Medical Center of Japan, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. [9]Fundació IrsiCaixa-HIVACAT, Hospital Germans Trias i Pujol, Badalona and Institucio Catalana de Recerca i Estudis Avancats (ICREA), Barcelona 08916, Spain. [10]University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [11]Emory University Vaccine Center and Yerkes National Primate Research Center, Atlanta, Georgia 30329, USA. [12]Zambia Emory HIV Research Project, and the Zambia Blood Transfusion Service, Lusaka, Zambia. [13]Ladymeade Reference Unit, University of West Indies, Bridgetown BB11156, Barbados. [14]Botswana-Harvard School of Public Health AIDS Initiative Partnership, Gaborone, Botswana. [15]Division of Medicine, Wright Fleming Institute, Imperial College, St Mary's Hospital, Norfolk Place, Paddington, London W2 1PG, UK. [16]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3SY, UK. [17]HIV Pathogenesis Programme, The Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban 4013, South Africa. [18]Miscrosoft Research, One Microsoft Way, Redmond, Washington 9805, USA. [19]BC Centre for Excellence in HIV/AIDS, Vancouver, British Columbia V6Z 1Y6, Canada. [20]Howard Hughes Medical Institute, Chevy Chase, Maryland 20185, USA.

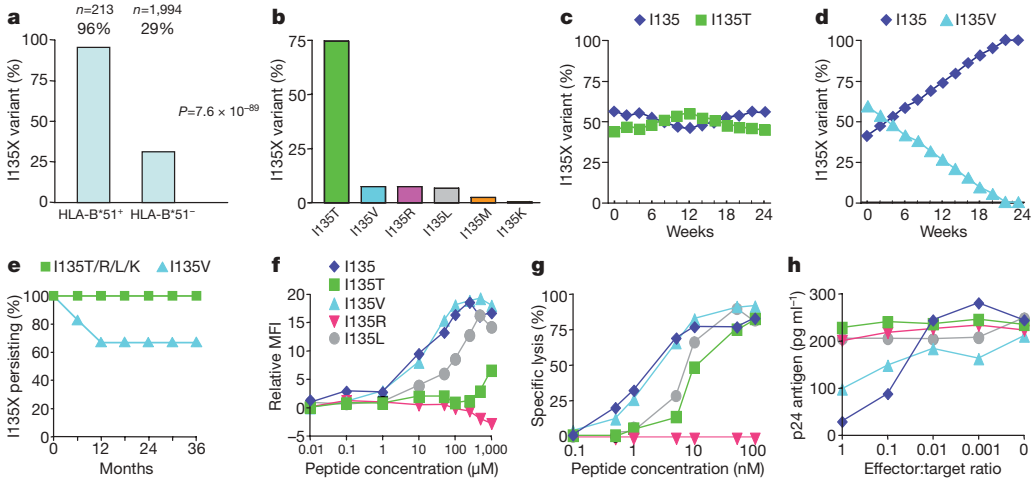

Figure 1 | Selection and fitness cost of I135X escape variants and recognition by the HLA-B*51–TAFTIPSI (RT 128–135)-specific CD8+ T cells. **a**, Association between I135X and HLA-B*51 in all study cohorts. **b**, Ile 135 variation in HLA-B*51-positive subjects. **c**, **d**, *In vitro* competition assays between NL4-3 wild-type virus and I135X viral variants (I135T (**c**) and I135V (**d**)). I135R and I135L showed no fitness cost (not shown). **e**, Persistence of I135X mutants in 38 HLA-B*51-negative subjects followed from acute infection. **f**, TAFTIPSI variant binding to HLA-B*51 (see Methods). MFI, mean fluorescence intensity. **g**, **h**, Recognition of peptide-pulsed HLA-B*51-matched targets and viral variants by representative TAFTIPSI-specific CD8+ T-cell clones.

prevalent[10] (21.9% of the study cohort), the frequency of I135X was >50%, and overall across all cohorts the I135X frequency was strongly correlated with HLA-B*51 prevalence ($P = 0.0001$, Fig. 2b). To control for the possibility that disproportionately more virus sequences from HLA-B*51-positive subjects were analysed, the same analysis comparing I135X frequency in HLA-B*51-negative subjects only was undertaken, with similar findings (Fig. 2c, $P = 0.0006$). These data suggest that HIV may be adapting to HLA-B*51 with respect to the HLA-B*51–TAFTIPSI response in localities where HLA-B*51 is at high prevalence.

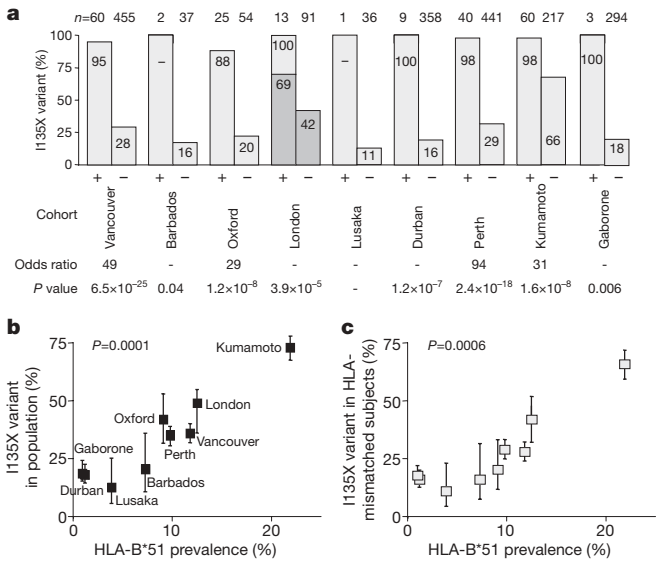

Figure 2 | Correlation between frequency of HLA-B*51-associated escape mutations and HLA-B*51 prevalence in study cohorts. **a**, Frequency of I135X mutations within TAFTIPSI (RT 128–135) in HLA-B*51-positive (+) and -negative (−) subjects within nine study cohorts. In the acute cohort (London) 69% of HLA-B*51-positive subjects expressed I135X mutant at enrolment, 100% within 2 years of baseline (Supplementary Fig. 1). **b**, Correlation between frequency of I135X mutation and HLA-B*51 prevalence in the nine study populations. Logistic regression $P = 0.0001$ (Supplementary Table 1). **c**, Correlation between I135X frequency in HLA-B*51-negative subjects and HLA-B*51 prevalence in nine study populations. Error bars represent 95% confidence limits, obtained using a binomial error distribution.

Additional evidence that I135X is accumulating in Japan comes from the observation that only 3 of 14 (21%) HLA-B*51-negative Japanese haemophiliacs infected in 1983 carried I135X, compared with 30 of 43 (70%) HLA-B*51-negative subjects infected between 1997 and 2008 ($P = 0.002$). Furthermore, HLA-B*51 does not protect against disease progression in Japanese subjects infected between 1997 and 2008, whereas HLA-B*51-positive haemophiliacs infected in 1983 had lower viraemia levels and higher CD4 counts than HLA-B*51-negative haemophiliacs (Supplementary Fig. 2). These data are consistent with fewer HLA-B*51-positive subjects targeting TAFTIPSI during 1997–2008, owing to a population-level increase in the HLA-B*51 I135X escape mutation over this 14–25-year period.

To investigate HIV adaptation to other HLA alleles, we initially examined other escape mutations shown previously to persist stably after transmission[5,7]. We selected the three non-reverting Gag polymorphisms that, from analysis of 673 study subjects in Durban, South Africa[7], were most strongly associated with the relevant restricting allele ($P < 10^{-6}$ after phylogenetic correction), namely, S357X, D260X and D312X within epitopes restricted, respectively, by HLA-B*07 (GPSHKARVL, Gag 355–363), HLA-B*35 (PPIPVGDIY, Gag 254–262) and HLA-B*44 (AEQATQDVKNW, Gag, 306–316). In addition, we analysed a non-reverting I31V variant (LPPIVAKEI, Int 28–36) previously hypothesized to increase in relation to population HLA-B*51 prevalence[5]. These additional polymorphisms show a similar relationship to that between I135X and HLA-B*51, overall showing a strongly significant correlation between variant frequency and prevalence of the restricting HLA allele (Figs 3 and 4a, and Supplementary Fig. 3).

The spectrum of HLA-associated polymorphisms also includes mutations reducing viral fitness[1]. These either revert to wild type after transmission, or persist in the presence of compensatory mutations. We extended these analyses to include epitopes restricted by HLA-B*27 and HLA-B*57, alleles strongly associated with successful immune control of HIV[11,12]. The mutations analysed themselves are associated with precipitating loss of immune control[13–16] and all inflict a documented viral fitness cost, either demonstrated by *in vitro* fitness studies and/or *in vivo* reversion[7,14,17–21] (data not shown for V168I).

Again, a strong correlation between escape mutant frequency and prevalence of the restricting HLA allele was observed (Figs 3c–f and 4b, and Supplementary Fig. 3; overall, for these nine variants affecting viral fitness, $r = 0.69$, $P < 0.0001$). Unexpectedly, this correlation

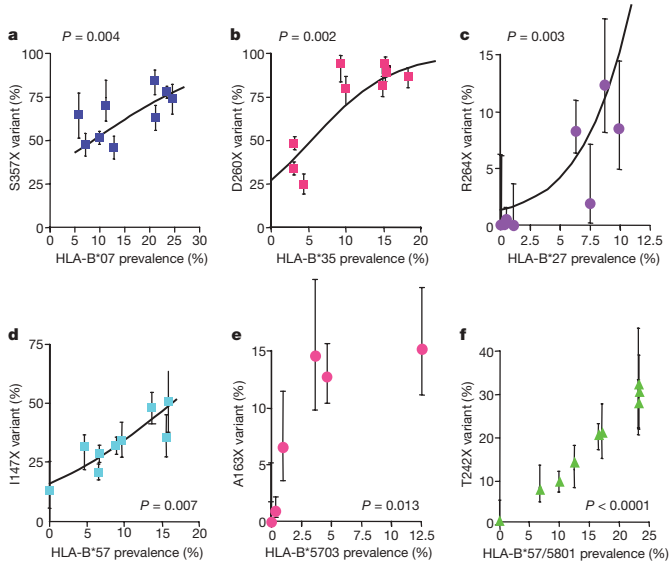

Figure 3 | Correlation between frequency of HIV sequence variant and HLA prevalence for six additional well-characterized epitopes. *P* values calculated after logistic regression analysis as shown (calculations after linear regression analysis are shown in Supplementary Table 1). **a**, Frequency of the S357X mutation within the HLA-B*07-restricted epitope GPSHKARVL (Gag 355–363). **b**, Frequency of the D260X mutation within the HLA-B*35-restricted epitope PPIPVGDIY (Gag 254–262). **c**, Frequency of the R264X mutation within the HLA-B*27-restricted epitope KRWIILGLNK (Gag 263–272). **d**, Frequency of the I147X mutation within the HLA-B*57-restricted epitope ISPRTLNAW (Gag 147–155). **e**, Frequency of the A163X mutation associated with the HLA-B*5703-restricted epitope KAFSPEVIPMF (Gag 162–172). **f**, Frequency of the T242X mutation within the B*57/5801-restricted epitope TSTLQEQIAW (Gag 240–249). Error bars represent 95% confidence limits, obtained using a binomial error distribution.

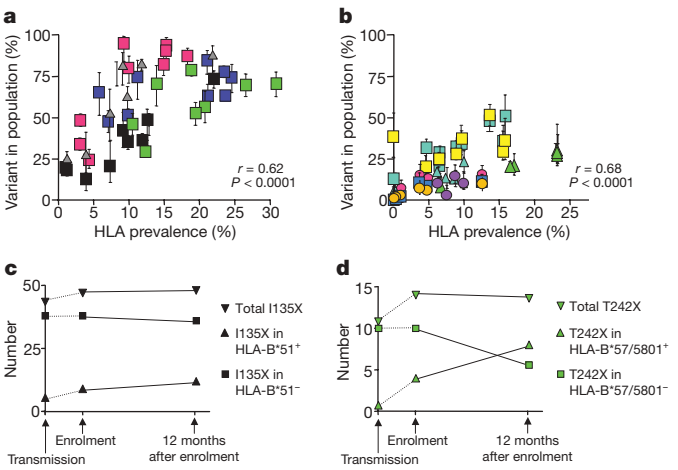

Figure 4 | Correlation between HIV variant frequency and HLA prevalence for all epitopes studied. **a**, Correlation between HLA prevalence and the five stable, non-reverting variants (symbols in Figs 2 and 3, and Supplementary Fig. 3; grey triangles, I31V; green squares, D312X). **b**, Eight variants demonstrated to reduce viral fitness (see text, Fig. 3 and Supplementary Fig. 3; turquoise triangles, L268X; yellow squares. A146X; sky-blue squares, V168I; yellow circles, I247X). **c**, **d**, Data from acute London cohort. **c**, Number of HLA-B*51-positive and HLA-B*51-negative subjects carrying the non-reverting I135X variant. The percentage of I135X in HLA-B*51-negative subjects at enrolment (42%) assumed the percentage of I135X in all subjects at transmission (I135X frequency in HLA-B*51-positive subjects at enrolment was 69%, $P = 0.07$). **d**, The reverting HLA-B*57/5801-restricted T242X mutation. T242X frequency in HLA-B*57/5801-negative subjects at enrolment was 7%, versus 33% in HLA-B*57/5801-positive subjects ($P = 0.01$). Error bars represent 95% confidence limits, obtained using a binomial error distribution.

remained significant even when comparing HLA prevalence with variant frequency in the HLA-mismatched population ($r = 0.40$, $P = 0.0004$). As anticipated, non-reverting variants such as I135X accumulate at the population level, but even rapidly reverting[18,20] mutations such as T242N can accumulate, if the selection rate exceeds the reversion rate (Fig. 4c, d).

Although frequency of the analysed HIV polymorphisms and HLA prevalence were strongly correlated overall, some anomalies were observed. For example, despite a 0% prevalence of HLA-B*57 in Japan[10], 38% of the Japanese cohort had the HLA-B*57-associated A146X variant. One potential explanation might be A146X selection by non-HLA-B*57 Japanese alleles. Analysing Gag sequences from Japanese study subjects, we observed a strong association between A146P and HLA-B*4801 ($P = 0.00035$), and then that A146P is indeed selected in HLA-B*4801-positive subjects (Supplementary Fig. 4a, b). We defined a novel HLA-B*4801-restricted epitope (Gag 138–147), showing also that A146P is an escape mutant (Supplementary Fig. 4c–f). These data illustrate that more than one HLA allele can drive the selection of a particular escape mutant (Supplementary Fig. 5). Also, in populations where HIV-specific $CD8^+$ T-cell responses are incompletely characterized, the influences of locally prevalent HLA alleles on HIV sequence variation are unknown.

These data show a strong correlation between HLA-associated HIV sequence variation and HLA prevalence in the population ($r = 0.69$, $P < 0.0001$, Supplementary Fig. 6), suggesting that the frequency of the studied variants is substantially driven by the HLA-restricted $CD8^+$ T-cell responses. Non-reverting variants[5,7], as well as those previously shown to arise at a fitness cost[7,14,16–21], were studied. The latter constitute approximately 55–65% of HLA-associated polymorphisms[7,20]. This current analysis included epitopes whose role in HIV immune control is unknown, as well as those believed to contribute significantly to containment of HIV[4,7,13–19]. Analysis of well-characterized epitopes only also served to limit potential confounding influences of epitope clustering (selection of the same variant by different HLA alleles) and of founder effect. Either would be capable of obscuring a true HLA effect on population variant frequency.

The HLA-B*57-associated A146X mutation illustrates the complexity that may result from epitope clustering. A146X is selected by at least six distinct HLA alleles (Supplementary Fig. 5). A true correlation existing between mutation frequency and individual HLA allele prevalence might thus be obscured by selection of the same mutation by other alleles.

Founder effect also has an undoubted influence on population frequencies of particular polymorphisms[6]. Phylogenetic correction of sequence data excludes founder effect as a confounder[6,7,9], and the highly significant associations between the presence of particular HLA alleles and all 14 HIV polymorphisms studied, persisting after phylogenetic correction (Supplementary Table 3), provide compelling evidence that the effects observed here are substantially HLA-driven. The large numbers of study subjects in these current studies reduce the likelihood of genuine HLA associations with HIV amino acid polymorphisms being obscured by founder effects. The relative impact of HLA and founder effect on variant frequency is harder to quantify, and is likely to differ substantially between particular populations.

The consequence of HIV adapting to certain $CD8^+$ T-cell responses is unknown. For non-reverting polymorphisms such as HLA-B*35-associated D260E, the variant approaches fixation, because even at population frequencies of 90%, D260E is still significantly selected in HLA-B*35-positive subjects (Supplementary Fig. 7b). Important questions relevant to vaccine design include the extent and rate of sequence change in populations. Relevant factors include the selection rate in subjects expressing the HLA allele, the reversion rate in HLA-mismatched subjects, the population HIV

transmission rate, and HLA allele prevalence. Models would need to include factors such as the selection of compensatory mutations to slow reversion rates, and antiretroviral therapy access that would slow transmission rates.

HLA adaptation to certain $CD8^+$ T-cell responses may also alter currently established HLA associations with slow disease progression. Data here suggest that, whereas 25 years ago HLA-B*51 was protective in Japan[11,12], this is no longer the case (Supplementary Fig. 2). The apparent increase in I135X frequency in Japan over this time supports the notion that HLA-B*51 protection against HIV disease progression hinges on availability of the HLA-B*51-restricted TAFTIPSI response. However, whether this is the case remains unknown.

For HLA-B*27 and HLA-B*57, there is more clear-cut evidence that their association with HIV control depends on the Gag-specific epitopes presented and analysed here[4,7,13–15,18,19]. For each of the HLA-B*27- and HLA-B*57-associated Gag mutations studied, an *in vitro* fitness cost or *in vivo* reversion has been observed. A strong correlation between variant frequency and HLA prevalence even for rapidly reverting variants can be explained, either by mutant acquisition exceeding reversion rate (Fig. 4D), or by selection of compensatory mutations slowing or halting reversion altogether. The clearest example of the latter is the HLA-B*27-associated R264K mutation, 'corrected' by S173A[19]. Compensatory mutations are also well described for the HLA-B*57-associated Gag mutations[14,18]. These data suggest that the escape mutations in these HLA-B*27- and HLA-B*57-restricted epitopes are accumulating over time. Several studies have now demonstrated that transmission of viruses encoding escape mutants in the critical Gag epitopes to individuals expressing the relevant MHC class results in failure to control viraemia[2,21,22]. The accumulation at the population level of these escape mutations in HLA-B*27 and HLA-B*57 Gag epitopes is therefore likely to reduce the facility of these alleles to slow HIV disease progression.

The longer-term consequences of this process for immune control of HIV are unknown. Loss of currently immunodominant epitopes would promote subdominant $CD8^+$ T-cell responses, which can be more effective[23,24]. Also, the adapted virus provides new epitopes that can be presented, potentially with beneficial effects. In hepatitis C virus, for example, HLA-A*0301 holds a particular advantage, but only against the specific strain of virus responsible for the Irish outbreak[25]. In HIV, HLA-B*1801 is associated with high viraemia in C clade but not in B clade infection[10,11,26]; the opposite applies to HLA-B*5301.

Thus, the data presented here, showing evidence that the virus is adapting to $CD8^+$ T-cell responses, some of which may mediate the well-established associations (HLA-B*57, HLA-B*27 and HLA-B*51) with immune control of HIV, highlight the dynamic nature of the challenge for an HIV vaccine. Important questions to be addressed include the speed and extent of sequence change, particularly in Gag, the most effective target for $CD8^+$ T-cell responses[1,7,13,21]. The induction of broad Gag-specific $CD8^+$ T-cell responses may be a successful vaccine strategy, but such a vaccine will be most effective if tailored to the viral sequences prevailing, and thus may need to be modified periodically to keep pace with the evolving virus. Moreover, the strong associations between certain HLA class molecules, such as HLA-B*57, HLA-B*27 and HLA-B*51, and slow disease progression may decline as the epidemic continues, particularly where these HLA alleles are highly prevalent, and where HIV transmission rates are high.

METHODS SUMMARY

Overall 2,875 subjects were studied, from 9 previously established study cohorts. These cohorts comprised subjects from North America, the Caribbean, Europe, sub-Saharan Africa, Australasia and Asia. All subjects were antiretroviral-therapy-naive. Apart from the London acute cohort ($n = 142$), all cohorts comprised chronically infected subjects. The 14 variants studied are well-defined escape mutations within well-characterized $CD8^+$ T-cell epitopes, and included those persisting after transmission and likely to have little effect on viral fitness ($n = 5$), as well as those shown previously to reduce viral fitness ($n = 9$). Autologous HIV-1 sequences, and HLA class I types, were determined for all study subjects. The replicative capacity of I135X variants selected within the HLA-B*51-restricted epitope TAFTIPSI (RT 128–135) was assessed via *in vitro* competition assays and also via longitudinal follow-up of HLA-B*51-negative subjects infected acutely with I135X variants. Polymorphism frequency in the study cohorts was compared with prevalence of the relevant HLA molecule in the study cohort using a logistic regression model taking into account the different numbers of study subjects in each cohort. Demonstration of an HLA allele driving escape at Gag 146 in the Japanese cohort was undertaken first by identification of an association between HLA-B*4801 and A146P, subsequent definition of an HLA-B*4801-restricted $CD8^+$ T-cell response to a novel epitope Gag 138–147 (LI10), and finally demonstration that A146P reduced viral recognition by LI10-specific $CD8^+$ T cells.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 October; accepted 22 December 2008.
Published online 25 February 2009.

1. Goulder, P. J. R. & Watkins, D. I. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Rev. Immunol.* **8,** 619–630 (2008).
2. Goulder, P. J. R. *et al.* Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* **412,** 334–338 (2001).
3. Moore, C. B. *et al.* Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296,** 1439–1443 (2002).
4. Draenert, R. *et al.* Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J. Exp. Med.* **199,** 905–915 (2004).
5. Leslie, A. J. *et al.* Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. *J. Exp. Med.* **201,** 891–902 (2005).
6. Bhattacharya, T. *et al.* Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* **315,** 1583–1586 (2007).
7. Matthews, P. *et al.* Central role of reverting mutations in HLA associations with viral setpoint. *J. Virol.* **82,** 8548–8559 (2008).
8. Tomiyama, H. *et al.* Identification of multiple HIV-1 CTL epitopes presented by HLA-B*5101 molecules. *Hum. Immunol.* **60,** 177–186 (1999).
9. Brumme, Z. *et al.* Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS* **22,** 1277–1286 (2008).
10. Itoh, Y. *et al.* High throughput DNA tying of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics* **57,** 717–729 (2005).
11. Kaslow, R. A. *et al.* Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature Med.* **2,** 405–411 (1996).
12. O'Brien, S. J., Gao, X. & Carrington, M. HLA and AIDS: a cautionary tale. *Trends Mol. Med.* **7,** 379–381 (2001).
13. Kiepiela, P. *et al.* $CD8^+$ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature Med.* **13,** 46–53 (2007).
14. Leslie, A. J. *et al.* HIV evolution: CTL escape mutation and reversion after transmission. *Nature Med.* **10,** 282–289 (2004).
15. Goulder, P. J. R. *et al.* Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Med.* **3,** 212–217 (1997).
16. Feeney, M. E. *et al.* Immune escape precedes breakthrough HIV-1 viremia and broadening of the CTL response in a HLA-B27-positive long-term nonprogressing child. *J. Virol.* **78,** 8927–8930 (2004).
17. Martinez-Picado, J. *et al.* Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J. Virol.* **80,** 3617–3623 (2006).
18. Crawford, H. *et al.* 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. *J. Virol.* **81,** 8346–8351 (2007).
19. Schneidewind, A. *et al.* Escape from a dominant Gag-specific CTL response in HLA-$B27^+$ subjects is associated with a dramatic reduction in HIV-1 replication. *J. Virol.* **81,** 12382–12393 (2007).
20. Brumme, Z. *et al.* Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J. Virol.* **82,** 9216–9227 (2008).
21. Goepfert, P. *et al.* Transmission of Gag immune escape mutations is associated with reduced viral load in linked recipients. *J. Exp. Med.* **205,** 1009–1017 (2008).
22. Seki, S. *et al.* Transmission of SIV carrying multiple cytotoxic T lymphocyte escape mutations with diminished replicative capacity can result in AIDS progression in Rhesus macaques. *J. Virol.* **82,** 5093–5098 (2008).

23. Gallimore, A., Dumrese, T., Hengartner, H., Zinkernagel, R. M. & Rammensee, H. G. Protective immunity does not correlate with the hierarchy of virus-specific cytotoxic T cell responses to naturally processed peptides. *J. Exp. Med.* **187,** 1647–1657 (1998).
24. Holtappels, R. *et al.* Subdominant CD8 T-cell epitopes account for protection against cytomegalovirus independent of immunodomination. *J. Virol.* **82,** 5781–5796 (2008).
25. McKiernan, S. M. *et al.* Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* **40,** 108–114 (2004).
26. Kiepiela, P. *et al.* Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432,** 769–775 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is funded by grants from the National Institutes of Health (RO1AI46995 (P.G.), 1 R01 AI067073 (B.D.W.), R01AI64060 (E.H.)), the Wellcome Trust (P.G., P.K.), the UK Medical Research Council (J.F., A.P. and P.M.), and the Mark and Lisa Schwartz Foundation, the Ministry of Health, Labour and Welfare (Health and Labour HIV/AIDS Research Grants 012), the NIHR Biomedical Research Centre Programme and the Ministry of Education, Science, Sports and Culture (number 18390141), Japan (M.T.). P.G. is an Elizabeth Glaser Pediatric AIDS Foundation Scientist; J.G.P. is a Marie Curie Fellow (contract number IEF-041811). The authors are also grateful to A. McLean and H. Fryer for discussions of the manuscript.

Author Contributions Y.K., K.P., J.F. and P. M. undertook much of the experimental work and data analysis, and contributed equally. M.T. and P.G. undertook much of the project conception, planning, supervision, analysis and writing of the manuscript, and contributed equally.

Author Information Accession numbers for newly determined viral sequences are included in Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.G. (philip.goulder@paediatrics.ox.ac.uk).

METHODS

Study subjects. The study cohorts have been described more fully elsewhere[3,7,9,13,14,18,20,21,27]. All comprise chronically infected and highly active antiretroviral therapy (HAART)-naive study subjects, with the exception of the London acute cohort ($n = 142$), who were enrolled immediately after seroconversion between 1999 and 2004, and 54 subjects enrolled during acute infection in Japan between 1997 and 2008. Viral sequences in all 2,679 chronically infected study subjects (all of whom were HAART-naive) were determined from time points after 2000, with the exception of 9 study subjects in the Japanese chronic cohort (1998–99) and all of the British Columbia cohort (1996–99). Sequencing data were obtained from 566 study subjects in the British Columbia cohort, 53 study subjects in the Barbados cohort, 106 in the Oxford cohort, 673 in the Durban cohort, 226 in the Lusaka cohort (chronically infected subjects enrolled between 2005–08), 481 study subjects in the Perth cohort, 277 chronically infected subjects in the Kumamoto cohort, 297 in the Gaborone cohort, and 142 subjects in the acute London cohort. An additional cohort in Japan comprised 117 haemophiliacs who were infected before 1985, the majority of which were believed to have been infected in 1983, and who were enrolled and followed up in out-patient clinics since 1997. These haemophiliacs are all now on HAART except for 4 HAART-naive subjects.

HLA-associated HIV amino acid polymorphisms studied. Variants studied that were shown to reduce viral fitness comprised polymorphisms within the HLA-B*27-restricted Gag epitope KRWIILGLNK (Gag 263–272; R264X and L268X) and mutations in three HLA-B*57-restricted Gag epitopes: ISPRTLNAW (ISW9, Gag 147–155), KAFSPEVIPMF (KF11, Gag 162–172) and TSTLQEQIAW (TW10, Gag 240–249). T242X is strongly selected by HLA-B*5801 in addition to HLA-B*57 subtypes[7,14,17]. The HLA-B*57-associated polymorphisms at residues Gag 146, 147 and 248 are selected by all HLA-B*57 subtypes, whereas Gag 163, 165, 166 and 247 are only selected by the HLA-B*5703 subtype (refs 7, 18 and H.C., unpublished data).

Statistics. Polymorphism frequency in the study cohorts was compared with prevalence of the relevant HLA molecule in the study cohort using a logistic regression model. To take account of the different numbers of study subjects in each cohort, appropriate confidence limits for the mutation frequencies were calculated, using the Adjusted–Wald method for binomial variables[28]. Logistic regression was calculated by GLMStat (http://www.glmstat.com) using a binomial error distribution and a logit link function. In addition, the Spearman's rank correlation coefficient was calculated in the context of a linear regression model (data shown in Supplementary Tables 1 and 2).

HLA class I typing. Because HLA typing was not undertaken consistently to four-digit resolution in all cohorts, two-digit HLA types only were used for these analyses, with the exception of the HLA-B*5703-associated polymorphisms (the Barbados and Oxford cohorts being excluded from these latter analyses as HLA-B*57 subtyping data were not available). Genomic DNA samples were initially typed to an oligo-allelic (two-digit) level using Dynal RELITM reverse SSO kits for the HLA-A, HLA-B and HLA-C loci (Dynal Biotech). Refining the genotype to the allele level was performed using Dynal Biotech sequence-specific priming (SSP) kits in conjunction with the previous SSO type. HLA phenotypic frequencies were determined from the HIV-infected study cohorts themselves.

Sequencing of viral RNA and proviral DNA. Viral sequencing of *gag* and *pol* from plasma RNA and proviral DNA was undertaken, using primers as previously described[7,9]. PCR products were sequenced directly or they were cloned by using a TOPO TA cloning kit (Invitrogen) and then sequenced. Sequencing was done with a Big Dye terminator v1.1. cycle sequencing kit (Applied Biosystems) and analysed by an ABI PRISM 310 genetic Analyser.

Competitive HIV-1 replication assay. Freshly prepared H9 cells (3×10^5) were exposed to the mixtures of paired virus preparations (300 blue cell-forming units) each of NL-432 versus mutant virus (I135T, I135V, I135R and I135L)), to be examined for their replication ability for 2 h, washed twice with PBS, and cultured as described previously[29]. On day 1, one-third of infected H9 cells were harvested and washed twice with PBS, and the proviral HIV-1 reverse transcriptase gene was sequenced (0 week). Every 7 days, the supernatant of the virus culture was transmitted to new uninfected H9 cells. The cells harvested at the end of every other passage (that is, at 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 and 24 weeks) were subjected to direct DNA sequencing of the HIV-1 reverse transcriptase gene, and the viral population change was determined by the relative peak height on the sequencing electrogram. The persistence of the original amino acid substitution was confirmed for all infectious clones used in this assay.

HLA-B*5101 stabilization assay. Binding of HIV-1-derived peptides to HLA-B*5101 was measured as previously described by using RMA-S-B*5101 cells[8].

Assays to determine recognition of peptide-pulsed or virus-infected targets. C1R and .221 cells expressing HLA-B*5101 or HLA-B*4801 were generated as previously described[30]. All cells were maintained in RPMI 1640 medium supplemented with 10% FCS and 0.15 mg ml^{-1} hygromycin B. Cytotoxicity of CD8$^+$ T cells for C1R-B*5101 cells pre-pulsed with peptide measured by the standard ^{51}Cr release assay was as previously described[8]. .221-B*4801 and .221 cells infected with NL4-3 or NL4-3 A146P mutant virus were used as target cells for intracellular cytokine staining assay.

Generation of the NL4-3 A146P mutant virus. The p82-2 plasmid containing the A146P mutation[4] was digested with BssHII and ApaI. The BssHII–ApaI 1.3-kb fragment was purified and then ligated into the same site of BssHII–ApaI-digested pNL-432 plasmid. To obtain pNL-432 including the A146P mutant (pNL-432 A146P), 293T cells were transfected with pNL-432 A146P using Lipofectamine 2000 (Invitrogen). Supernatants from transfected 293T cell cultures were stored at −80 °C.

Generation of CD8$^+$ T-cell clones and peptide-specific CD8$^+$ T-cell lines. Cytotoxic T lymphocyte (CTL) clones were generated from HIV-1-specific bulk-cultured T cells by limiting dilution as previously described[8]. Peptide-specific CD8$^+$ T-cell lines were generated by stimulating peripheral blood mononuclear cells (PBMCs) from the HLA-B*4801-positive HIV-1-seropositive individual KI-092 with the NI11 (NLQGQMVHQAI) peptide and then culturing them for 2 weeks[8]. Cytotoxicity of CD8$^+$ T cells for target cells pre-pulsed with peptide measured by the standard ^{51}Cr release assay was as previously described[8].

Suppression assay of HIV-1 replication by HIV-1-specific CTLs. The ability of HIV-1-specific CTLs to suppress HIV-1 replication was examined as previously described[30].

Intracellular cytokine staining assays. PBMCs from HIV-1-infected individuals were stimulated with the desired peptide (1 μM) and cultured for 12–14 days. These cultured PBMCs were assessed for IFN-γ-producing activity as previously described[30].

27. Tang, J. *et al.* Favorable and unfavorable HLA class I alleles and haplotypes in Zambians predominantly infected with clade C human immunodeficiency virus type 1. *J. Virol.* **76,** 8276–8284 (2002).
28. Agresti, A. & Coull, B. Approximate is better than 'exact' for interval estimation of binomial proportions. *Am. Stat.* **52,** 119–126 (1998).
29. Gatanaga, H., Hachiya, A., Kimura, S. & Oka, S. Mutations other than 103N in human immunodeficiency virus type 1 reverse transcriptase (RT) emerge from K103R polymorphism under non-nucleoside RT inhibitor pressure. *Virology* **344,** 354–362 (2006).
30. Tomiyama, H., Akari, H., Adachi, A. & Takiguchi, M. Different effects of Nef-mediated HLA class I down-regulation on HIV-1-specific CD8$^+$ T cell cytokine activity and cytokine production. *J. Virol.* **76,** 7535–7543 (2002).