

1 Widespread Impact of HLA Restriction 2 on Immune Control and Escape Pathways 3 in HIV-1

4 Jonathan M. Carlson^{1*}, Jennifer Listgarten¹, Nico Pfeifer¹, Vincent Tan¹, Carl Kadie² Bruce D. Walker^{3,4,5},
5 Thumbi Ndung'u^{3,5}, Roger Shapiro⁶, John Frater⁷, Zabrina L. Brumme^{8,9}, Philip J. R. Goulder^{5,10}, David
6 Heckerman¹

7 ¹Microsoft Research, eScience Group, Los Angeles, CA, USA; ²Microsoft Research, eScience Group,
8 Redmond, WA, USA; ³Ragon Institute of Massachusetts Institute of Technology and Harvard University,
9 Boston, MA, USA; ⁴Howard Hughes Medical Institute, Chevy Chase, Maryland, USA; ⁵HIV Pathogenesis
10 Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of
11 KwaZulu-Natal, Durban, South Africa; ⁶Division of Infectious Diseases, Beth Israel Deaconess Medical
12 Center, Boston, MA, USA; ⁷Nuffield Department of Clinical Medicine, Oxford University, Oxford, United
13 Kingdom; ⁸Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada; ⁹British Columbia
14 Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada; ¹⁰Department of Paediatrics, University of
15 Oxford, Oxford, United Kingdom

16

17

18 * Corresponding author:
19 Jonathan Carlson, Ph.D.
20 Microsoft Research
21 1100 Glendon Ave, PH-1
22 Los Angeles, CA 90024
23 carlson@microsoft.com
24 Tel:(415) 704-8906 Fax: (425) 936-7392

25

26 **Running title:** Differential escape mediated by HLA restriction

27

28 **Key words:** HIV-1, HLA class I, Immune control and escape, differential escape, CTL epitopes

29 **Word Count:** 9,159

30 **Abstract Word Count:** 248

31 **Abstract**

32 The promiscuous presentation of epitopes by similar HLA class I alleles holds promise for a universal T-
33 cell based HIV-1 vaccine. However, in some instances CTL restricted by HLA alleles with similar or
34 identical binding motifs are known to target epitopes at different frequencies, with different functional
35 avidities and with different apparent clinical outcomes. Such differences may be illuminated by the
36 association of similar HLA alleles with distinctive escape pathways. Using a novel computational method
37 featuring phylogenetically-corrected odds ratios, we systematically analyzed differential patterns of
38 immune escape across all optimally defined epitopes in Gag, Pol and Nef in 2,126 HIV-1 clade C infected
39 adults. Overall, we identified 301 polymorphisms in 90 epitopes associated with HLA alleles belonging to
40 shared supertypes. We detected differential escape in 37 of 38 epitopes restricted by more than one
41 allele, which included 278 instances of differential escape at the polymorphism level. The majority (66-
42 97%) of these resulted from the selection of unique HLA-specific polymorphisms rather than differential
43 epitope targeting rates, as confirmed by IFN- γ Elispot data. Discordant associations between HLA alleles
44 and viral load were frequently observed between allele pairs that selected for differential escape.
45 Furthermore, the total number of associated polymorphisms strongly correlated with average viral load.
46 These studies confirm that differential escape is a widespread phenomenon and may be the norm when
47 two alleles present the same epitope. Given the clinical correlates of immune escape, such
48 heterogeneity suggests that certain epitopes will lead to discordant outcomes if applied universally in a
49 vaccine.

50 **Introduction**

51 Variation within the highly polymorphic MHC region is the primary genetic component linked to immune
52 control of HIV-1 (28, 76). This effect is due almost entirely to specific HLA-I alleles, many of which have
53 been previously linked with rates of HIV disease progression in molecular epidemiology studies (22, 24,

54 34, 42, 44). HLA-I associated immune control of HIV is mediated by CD8+ T cells, which recognize viral
55 epitopes presented by HLA-I proteins on the surface of infected cells. HIV-1, however, is able to evade
56 recognition by HLA-restricted CD8+ T-cells through the selection of immune escape mutations (33, 63).

57 Recently, HLA-restricted immune escape pathways have been systematically identified through
58 population-level analyses of linked HLA class I and HIV sequence datasets, yielding detailed “immune
59 escape maps” of the HIV-1 proteome (14–16, 19, 60, 65). The discovery that immune escape pathways
60 are generally predictable based on the host HLA repertoire represents a major step forward in HIV
61 vaccine research (1, 18); however, substantial differences in the probability of escape have been
62 observed between populations (4, 19, 41), between individuals (58, 67, 81), and even between members
63 of the same HLA allelic family (41, 50). Achieving a deeper understanding of the host correlates of
64 immune escape is therefore of utmost importance to T-cell-based HIV-1 vaccine design.

65 HLA class I peptide binding specificities are largely defined by polymorphisms in the peptide-
66 binding groove of the HLA molecule (6, 7, 70). HLA alleles with similar sequences in the binding groove
67 therefore tend to bind similar or even identical peptides, which allows HLA alleles to be grouped into
68 families, or “supertypes”, based on shared peptide presentation (17, 71, 72). A large number of HLA-
69 restricted CD8+ T-cell epitopes have been optimally defined in HIV-1 (52), and vaccine strategies based
70 on the design of universal “supertope” immunogens have been proposed as a method to elicit broad
71 immune responses using a limited number of epitopes (70, 72). However, despite these common
72 patterns, substantial caveats remain. Although some epitopes display promiscuity of HLA binding (30),
73 meaning they can be presented by a variety of HLA-I alleles, both within (17, 30, 66, 77) and between
74 (30, 54, 66) HLA supertypes, the frequency of epitope targeting and/or mutational escape may vary
75 depending on which HLA allele is presenting the epitope. For example, members of the B7 supertype
76 exhibit vastly different escape and functional characteristics despite similar epitope targeting

77 frequencies (50), while members of the B58 supertype exhibit very different targeting frequencies (2,
78 46, 53, 57). Perhaps as a result of restricting different epitopes (25, 45, 46), or differential escape within
79 commonly restricted epitopes (40), members of both B7 and B58 superotypes have discordant
80 associations with viral control (44, 49). More broadly, comparative studies of immune escape across
81 cohorts, ethnicities and geographic regions have revealed that alleles of the same supertype or type
82 (formerly referred to as 2-digit allele group) are not always associated with the same immune escape
83 patterns (41), and identical alleles may select different escape patterns in different ethnic groups (4).
84 Taken together, these studies suggest that CD8+ targeting frequency and risk of immune escape are
85 highly dependent on the genetic context in which the epitope is presented, a result that may have
86 profound consequences for subsequent viral control. In this study, we explore in detail the relationship
87 between HLA allele carriage (at the subtype level) and risk of immune escape in HIV-1 and ability to
88 control viral replication.

89 Systematic analysis of context-dependent immune escape has been limited by a lack of
90 appropriate statistical tools. Studies to date have relied on comparative analyses of HLA-associated
91 polymorphisms identified in different HIV-1 cohorts worldwide (4, 41), an approach that is error prone
92 due to high false negative rates and statistical power that varies based on HLA allele frequency and
93 cohort sample size. We therefore developed a statistical approach to compare the magnitude of
94 immune selection pressure (and thus by extension the risk of immune escape) on a given HIV codon, in
95 different host genetic contexts. We then applied this method to a population-based dataset of linked
96 CD8 T-cell responses, HLA class I types and HIV sequences from Southern Africa, to investigate the
97 patterns and genetic correlates of immune escape within all optimally defined CD8+ T-cell epitopes in
98 HIV-1 Gag, Pol and Nef. Using this method, we identified members of the same HLA supertype that
99 restrict the same optimally defined epitopes, as evidenced by the presence of HLA-associated
100 polymorphisms at the population level (5, 9). We then systematically tested for differential selection

101 among members of the same HLA supertype that restrict the same epitope. Finally, we explored the
102 potential effects of differential selection on plasma viral load.

103 **Materials and Methods**

104 **Study subjects**

105 We studied 2126 chronically HIV-1 subtype C-infected, antiretroviral naïve adults from five established
106 African cohorts: (i) Durban, South Africa (N=1218) (49, 56); (ii) Bloemfontein, South Africa (N=261) (39);
107 (iii) Kimberley, South Africa (N=31) (55); (iv) Gaborone, Botswana (N=514) (69); and (v) from Southern
108 African subjects attending outpatient HIV clinics in the Thames Valley area of the U.K. (N=102),
109 originating from Botswana, Malawi, South Africa and Zimbabwe (55). Ethics approval was granted by the
110 University of KwaZulu-Natal Biomedical Research Ethics Committee and the Massachusetts General
111 Hospital Review Board (Durban cohort); the University of the Free State Ethics Committee (Kimberley
112 and Bloemfontein cohorts); the Office of Human Research Administration, Harvard School of Public
113 Health and the Health Research Development Committee, Botswana Ministry of Health (Gaborone
114 cohort); and the Oxford Research Ethics Committee (Durban, Kimberley, and Thames Valley cohorts).
115 Study subjects from all cohorts gave written informed consent for their participation.

116 High resolution sequence-based HLA typing was performed as previously described (55). For the
117 present study, all HLA alleles that could not be resolved to the subtype level were considered as missing
118 (2,919 of 14,486; 20.2%). HLA supertype, type and subtype frequencies are shown in **Table S1**.
119 Population sequences of HIV-1 proviral DNA-derived *gag* (p17+p24, N=1,327), *pol* (*protease* N = 865,
120 *reverse transcriptase* N = 905, *integrase* N=344), and *nef* (N=738) were obtained (**Table S2**), as previously
121 described (55).

122 Viral load in chronic infection was measured using the Roche Amplicor version 1.5 assay and
123 CD4+ T cell counts were measured by flow cytometry, as previously described (55). Individuals with

124 <2,000 viral copies/ml plasma and >250 CD4+ T cells/mm³ were defined to be “viremic controllers”. Due
125 to the geographic heterogeneity of the Thames Valley cohort, this cohort was excluded from viral load
126 analyses. Viral load and high resolution HLA typing were available for 1,870 individuals from the
127 remaining cohorts.

128 **A phylogenetically-corrected odds ratio**

129 To allow us to quantify and compare the strength of selection pressure exerted by a particular HLA allele
130 on a given HIV-1 codon, we adapted standard logistic regression techniques to take into consideration
131 underlying evolutionary relationships between the HIV-1 sequences in the dataset, yielding a statistic we
132 call the “phylogenetically-corrected odds ratio” of escape, which measures the strength of selection
133 exerted by an HLA allele on a given polymorphism.

134 Logistic regression is a model used for predicting the probability of occurrence of a binary event,
135 making it useful for modeling the probability of observing particular viral amino acids as a function
136 various predictors (such as HLA alleles or viral load). For this reason, logistic regression was used in the
137 first population-level immune escape study (60). The model can be described as follows. Suppose we are
138 interested in the probability of seeing a particular amino acid at a particular position, say 242N in HIV-1
139 Gag. If p is the probability of observing 242N, then the *odds* of observing 242N is $p/(1 - p)$. Logistic
140 regression models the log of the odds (“log-odds”) as a linear function of predefined predictors. For
141 example, if we assume the odds of seeing 242N depends on whether an individual expresses HLA alleles
142 X or Y , then $\ln\left(\frac{p}{1-p}\right) = aX + bY + c$, where X and Y are taken to be 0/1 binary variables, and a , b and c
143 are scalar parameters whose values are chosen so as to maximize the likelihood of the data.
144 Conveniently, the maximum likelihood parameters have intuitive interpretations: c represents the log-
145 odds of observing 242N among individuals who express neither X nor Y ; and a is the log-odds ratio of
146 242N among individuals who express HLA X compared to individuals who do not express X (and

147 similarly for b and Y). A positive log-odds ratio ($a > 0$) indicates that 242N is more likely to be observed
148 among individuals expressing the allele than among those not expressing the allele, while a negative log-
149 odds ratio ($a < 0$) indicates the opposite. Thus, if a typical escape is T242N mediated by X=B*57:03,
150 then we would expect to see a negative weight when computing the odds of T and a positive weight
151 when computing the odds of N.

152 Although logistic regression is broadly applied in biomedical research, it can yield surprisingly
153 high false positive and false negative rates when applied to viral sequences, which share an evolutionary
154 relationship (11, 21). This problem can be circumvented in the special case where the transmitted virus
155 sequence is known; however, in the vast majority of cases the transmitted viral sequence is unknown.
156 To get around this issue, we perform maximum-likelihood phylogenetic reconstructions of the HIV-1
157 sequences observed in the dataset (one maximum likelihood tree for *gag-pol* and another for *nef*,
158 estimated using PhyML 3.0 (36)) in order to estimate the transmitted viral sequence for each subject. A
159 statistical model can then be made “phylogenetically-corrected” by designing that model to make use of
160 both the estimated transmitted and the observed current viral sequences, then averaging over the
161 possible phylogenetic histories, as previously described (19, 21).

162 To create a phylogenetically-corrected logistic regression test, we therefore first need to define
163 a logistic regression model for cases in which both the transmitted and current viral sequences are
164 known for each individual. To this end we modify the above definition to be $\ln\left(\frac{p}{1-p}\right) = aX + bY + cT$,
165 where T represents a binary variable indicating whether or not the transmitted sequence contained
166 242N. We model T as a -1/1 binary variable whereas the HLA variables X and Y are modeled as 0/1
167 binary variables. Thus, if an individual expresses neither X nor Y , then the log-odds of observing 242N
168 will be c if the transmitted sequence contained 242N, and $-c$ if it did not. After picking maximum-
169 likelihood values for a , b and c , we can then interpret a as the log odds ratio comparing the odds of

170 observing 242N among individuals expressing HLA-*X* compared to those not expressing HLA-*X*,
171 conditioned on the transmitting sequence.

172 The distinction between an odds ratio conditioned on the transmitted sequence and a more
173 traditional odds ratio is important. In the traditional case, we model the odds of carriage of a specific
174 polymorphism (regardless of whether it was acquired at transmission or subsequently selected in vivo)
175 in individuals expressing the relevant HLA allele compared to those not expressing it. The magnitude of
176 the traditional odds ratio is therefore influenced by the frequency of the polymorphism in persons
177 expressing the relevant HLA allele, as well as its prevalence in the overall population. Thus a high odds
178 ratio may result either from a high probability of escape in individuals expressing the HLA allele, or a
179 high level of conservation among individuals not expressing the allele. In contrast, when we condition on
180 the transmitting sequence, we effectively model the odds of observing the selection of this mutation in
181 vivo (because both the observed and transmitted variants are included in the model). In the context of
182 HLA-mediated escape, the magnitude of an odds ratio that is conditioned on the transmitted virus can
183 therefore be viewed as a measure of the strength of selection in vivo.

184 In practice, we cannot observe the infecting sequence in large cross sectional cohorts.
185 Therefore, we perform a weighted average over all possible infecting sequences, where the weights are
186 informed by the phylogeny (19).

187 **Hypothesis testing with the phylogenetically-corrected logistic regression model**

188 Various hypotheses can be tested using the likelihood ratio test, which compares the likelihood of the
189 null model against the likelihood of a richer model. To test if an HLA allele is associated with a given
190 polymorphism, we compare the null model $\ln\left(\frac{p}{1-p}\right) = cT$ (meaning the odds of observing a
191 polymorphism is completely determined by the inferred transmitted sequence) to the alternative model
192 $\ln\left(\frac{p}{1-p}\right) = aX + cT$ (meaning the odds of observing the polymorphism is mediated by selection

193 pressure imposed by HLA allele X). We can also test whether HLA alleles X and Y exert differential
194 selection pressure on 242N. First, we construct a new variable $\max(X, Y)$, which is 1 only if an
195 individual expresses either X or Y . We then compare the null model $\ln\left(\frac{p}{1-p}\right) = a \max(X, Y) + cT$ to
196 the alternative model $\ln\left(\frac{p}{1-p}\right) = a \max(X, Y) + bX + cT$ to test if there is sufficient evidence that X
197 and Y should be treated as separate variables. To test the hypothesis that HLA allele X exerts differential
198 selection pressure on 242N when co-expressed with HLA allele Y , we construct an interaction term XY ,
199 which is 1 only if an individual expresses both X and Y . We then compare the null model $\ln\left(\frac{p}{1-p}\right) =$
200 $aX + cT$, to the alternative interaction model $\ln\left(\frac{p}{1-p}\right) = aX + bXY + cT$. The parameter b can then be
201 interpreted as the log-odds ratio of escape in individuals co-expressing both X and Y compared to
202 individuals expressing only X . This interaction model is also used when Y is a continuous variable (e.g.,
203 log viral load).

204 **Multiple hypothesis testing**

205 In the present study, we perform thousands of statistical tests. In such scenarios, the standard
206 interpretation of the p-value has relatively little meaning. We therefore primarily report false discovery
207 rates, which addresses multiple hypothesis testing (8). The false discovery rate (FDR) is a property of a p-
208 value (p_0) in the context of a specific set of tests, and is defined as the expected proportion of tests for
209 which $p \leq p_0$ that are false positive. The false discovery rate can be estimated using $FDR(p_0) =$
210 $\pi_0 p_0 N / R$, where N is the total number of tests performed, R is the number of tests with $p \leq p_0$, and π_0
211 is the (unknown) proportion of all tests that are truly null (74). A straightforward, robust estimate of π_0
212 is $\hat{\pi}_0 = 2 \cdot \text{avg}(p)$, where $\text{avg}(p)$ is the average p-value of all the tests (64). To ensure monotonicity
213 with respect to p-values, the FDR is reported as a q-value, which is the minimum false discovery rate for
214 all $p \geq p_0$ (73).

215 The appropriate choice of q-value threshold is context-specific and depends on how the results
216 will be interpreted. In the present study, we typically report all tests with $q < 0.2$ (implying that we expect
217 20% of reported tests to be false positives), but will sometimes report lower q-values when more
218 conservative interpretations are appropriate.

219 **Definition of expanded optimal epitopes**

220 Optimally-defined (52), HLA-restricted CTL epitopes in HIV-1 Gag, Pol and Nef proteins were retrieved
221 from the Los Alamos Database
222 (http://www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html, last updated August
223 31, 2009) and hand edited to reflect recent published corrections. These optimal epitope definitions are
224 derived from in vitro epitope finemapping and HLA restriction experiments reported in the literature.
225 Therefore, published epitopes have not necessarily been tested in the context of all possible HLA alleles
226 that could present them, nor have the restricting HLA alleles been defined at the same level of
227 resolution throughout. Indeed, many epitopes have only been restricted to one or two alleles whereas
228 others have been attributed to broad serotypes. In recognition of the fact that alleles with shared similar
229 binding grooves are likely to present similar peptides, we expanded the optimal epitope list to include all
230 HLA subtypes belonging to the published HLA type, supertype, or serotype, as follows. For each optimal
231 epitope, we expanded the list of restricting HLA alleles to include all members of the HLA supertype to
232 which the original restricting allele belonged (71). For optimal epitopes restricted by HLA alleles defined
233 by their serotype only, we expanded the list to include all HLA alleles belonging to that serotype (37).
234 For HLA-C alleles, which do not have supertype definitions, we expanded the list to include all HLA
235 subtypes belonging to the HLA type of the published restricting allele.

236 We next sought to identify putative HLA escape mutations for each optimal epitope by
237 identifying polymorphisms at sites within or flanking each epitope that were positively or negatively
238 associated with particular HLA alleles. Specifically, for each observed amino acid at each position within

239 3 amino acids of the optimal epitope boundary, we ran a forward selection procedure to identify all HLA
240 alleles that were independently associated with the amino acid. Only HLA alleles that were expressed by
241 at least three individuals in the present study were analyzed; likewise, only polymorphisms that were
242 observed in at least three individuals, and at most N-3 individuals, were considered. For each round of
243 forward selection, we tested each HLA allele using a likelihood ratio test that compared an alternative
244 phylogenetically-corrected logistic regression model that included the new allele to a null model that
245 included all alleles that had been added in previous iterations. After each iteration, the most significant
246 HLA allele was added to the model. The p-value reported for each HLA allele was that computed when
247 the allele was added to the model. As a post-processing step, we filtered the final output to include only
248 those HLA alleles that are in the expanded list of potential restricting HLA alleles and computed q-values
249 based on the resulting subset. In some cases, one escape association could be ascribed to multiple
250 overlapping optimal epitopes, each of which is putatively restricted by the same HLA allele or HLA alleles
251 in the same supertype (e.g., the overlapping Gag epitopes KIRLRPGGK, RLRPGGKKK, and RLRPGGKKKY
252 are all published as A*03:01 optimal epitopes, while the overlapping B7-restricted epitopes VLRPMTY
253 and RPMTYKAAL are published as B*35:01 and B*07:02 restricted, respectively). In these cases,
254 overlapping optimal epitopes were grouped by published restricting supertype so that each such
255 polymorphism was only analyzed once. We only tested for differential escape between HLA alleles that
256 restricted the same optimal epitope (as determined by the supertype/serotype expansion described
257 above).

258 **IFN- γ ELISPOT assays**

259 In vitro HIV-specific CD8+ T-cell responses were determined in a cohort of 1010 subtype-C infected
260 individuals using IFN- γ ELISPOT assays using a set of 410 overlapping 18mer peptides (OLPs) spanning
261 the whole HIV-1 subtype C proteome (2001 consensus sequence). Overlapping peptides were arranged
262 in a matrix system with 11-12 peptides in each pool. Responses to matrix pools were deconvoluted by

263 subsequent testing with the individual 18mer peptides within each pool, and the identity of the
264 individual 18mers recognized were thus confirmed, as previously described (44). Each optimal epitope
265 was mapped to the OLP(s) that completely contained the optimal. The CTL targeting frequency of each
266 optimal epitope was defined as the targeting frequency of the OLP containing it (or, in the case where it
267 was contained in two OLPs, the maximum targeting frequency between them). Associations between
268 HLA alleles and OLP responses were assessed using a stepwise Fisher's exact procedure. For each OLP,
269 we identified the most significantly associated HLA allele using Fisher's exact test. We then removed all
270 individuals who expressed that allele, and repeated until all HLA alleles had been added to the model.
271 We then computed false discovery rates for each HLA-allele:OLP pair using the method described in
272 (20), using a web server provided by the authors ([http://research.microsoft.com/en-](http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/)
273 [us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/](http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/)).

274 **Results**

275 **Systematic identification of escape mutations in optimally-defined epitopes**

276 This study focuses primarily on differential escape within epitopes presented by similar HLA alleles. To
277 this end, we developed a phylogenetically-corrected logistic regression model, which estimates the
278 relative odds of escape among individuals who express a given HLA allele compared to those who do
279 not. As described in Methods, our model conditions on the transmitted sequence (as estimated from the
280 phylogeny), thereby removing any confounding that may arise from evolutionary relatedness among the
281 HIV sequences (11, 19, 21). By building on the logistic regression model, our model allows us to estimate
282 the relative odds of escape, as well as to explicitly test for differential escape (difference of odds of
283 escape between two alleles) or escape that is dependent on external factors (interaction effects).

284 We first applied this phylogenetically-corrected model to a large population-based dataset to
285 identify associations between individual HLA alleles and HIV-1 polymorphisms occurring within 3 amino

286 acids of all optimal epitopes potentially restricted by those alleles. Potential HLA-optimal epitope
287 restriction was defined by expanding the published list of optimally-defined epitopes (52) to include all
288 HLA alleles in the same supertype family as the published restricting alleles (see Methods). A forward
289 selection algorithm was used to reduce the risk of false positives arising from linkage disequilibrium
290 among HLA alleles (19). We identified 301 significant ($q < 0.2$, $p < 0.004$) HLA-HIV associations in Gag
291 ($n=147$), Pol ($n=110$), and Nef ($n=44$), covering 90 of 157 (57%) optimal epitopes (**Table S3**). In what
292 follows, we say that an HLA allele “restricts” an epitope if that allele is in the expanded optimal list and
293 is associated with at least one escape polymorphism. There was an average of 1.9 HLA alleles that
294 restricted each of those 90 optimal epitopes. Thirty-eight epitopes were restricted by more than one
295 HLA allele (**Table 1**) and 67 epitopes were restricted by an allele other than its published restricting one.
296 Thus, in addition to identifying putative HLA-specific escape mutations, this analysis expands the
297 number of closely related HLA alleles capable of presenting each optimal epitope by using escape
298 mutations as indicators of active immune selection pressure in vivo.

299 **Widespread differential escape among HLA alleles restricting the same** 300 **epitope**

301 Examination of HLA-associated polymorphisms in Table 1 gives the impression that different HLA alleles
302 restricting the same epitope will select for the same escape mutation only rarely. However it would be
303 premature to draw this conclusion from the association lists alone without undertaking rigorous
304 statistical tests. For example, the absence of any particular association may be due to lack of statistical
305 power. Furthermore, two apparently identical associations may actually occur at substantially different
306 frequencies among individuals expressing two different HLA alleles despite achieving statistical
307 significance in both cases. We therefore created a statistical test for differential escape based on the
308 phylogenetically-corrected logistic regression that allows us to explicitly test whether the odds of escape
309 mediated by two different HLA alleles are different.

310 For each HLA-associated polymorphism in Table 1 we tested for differential selection between
 311 the reported allele and every other HLA allele that restricted the same epitope. In so doing, we confirm
 312 that HLA alleles restricting the same epitope exhibit differential escape in the vast majority of cases.
 313 Using the estimation method of Pounds and Cheng (64), which compares the observed distribution of p-
 314 values for a large number of statistical tests against the expected distribution of p-values under the null
 315 hypothesis, we estimate that roughly 70% of the 499 comparisons represent truly differential selection.
 316 Thus, differential selection appears to be the norm among HLA alleles that restrict the same epitope.
 317 Indeed, of the 38 epitopes that are restricted by multiple members of the same supertype, 37 (97%)
 318 exhibited differential escape in at least one position within or flanking the epitope. The only exception
 319 was RT-IL9 (IEELRQHLL), which was restricted by B44 supertype members B*18:01 and B*44:03. Tests
 320 for differential escape did not achieve statistical significance despite the observation that the two alleles
 321 were associated with different polymorphisms (Table 1). Overall, a total of 278 instances of differential
 322 escape within the same epitope were observed at $p < 0.05$ ($q < 0.025$); these are listed in **Table S4**. **Figure**
 323 **1** displays the subset of these instances for which $p < 0.005$ ($q < 0.006$).

324 **Three broad categories of differential immune escape**

325 Differential escape (Table S4) can be classified into three patterns. Firstly, we observe cases where two
 326 alleles select for the same escape mutation, but to differing degrees. Secondly, we observe cases where
 327 one allele selects for escape whereas the other allele shows no association whatsoever. Finally, we
 328 observe cases where one allele is significantly positively associated with a polymorphism, and the other
 329 allele is significantly negatively associated with the same polymorphism, a phenomenon termed “push-
 330 pull” escape (14).

331 The B7 supertype alleles B*42:01, B*81:01, B*39:10 and B*67:01, all of which are associated
 332 with escape in Gag-TL9 (TPQDLNTML), illustrate all three categories of differential escape. The first type
 333 (identical escape patterns that differ in statistical strength) is illustrated by the selection of T186X by

334 both B*81:01 and B*39:10, but with a significantly higher absolute odds ratio for B*81:01 compared to
335 B*39:10 at this residue (ln odds ratios of -12 vs. -10, $q=0.016$; negative ln odds ratio indicate selection
336 against a polymorphism, in this case the T variant). The second type (selection of escape by one but not
337 other related alleles) is illustrated by the lack of significant association between T186 and B*42:01. The
338 third type, “push-pull” escape, is illustrated by the selection of X182T (wild type is Q) by B*42:01, but
339 the specific selection against 182T by B*81:01 (which instead selects for Q182E/G/S). In this epitope, we
340 also observed examples in which two alleles selected for the same escape patterns with the same
341 frequencies: both B*39:10 and B*81:01 were associated with selection of E177D 3 amino acids
342 upstream of TL9 with a ln odds ratio of 4 ($p=0.5$ for differential escape between the two alleles).

343 Remarkably, there were only nine clear cases of differential escape in which two HLA alleles
344 selected for the same polymorphisms but to a varying degree. These included B*57:03/B*58:01
345 mediated selection of T242N in Gag-TW10, A146P in Gag-IW9, and X116N in Nef-HW9 (where B*57:03
346 exhibited a higher odds of escape compared to B*58:01 in all three cases); B*81:01/B*39:10 mediated
347 selection of T186X in Gag-TL9 (where B*81:01 exhibited higher odds of escape compared to B*39:10);
348 B*35:01/B*53:01 mediated selection of V133X in Nef-TL10 (where B*35:01 exhibited higher odds of
349 escape compared to B*53:01); and finally A*24:02/A*23:01 mediated selection of R28X (where A*24:02
350 exhibited higher odds of escape compared to A*23:01). Similarly, there were only two cases of
351 significant push-pull: in addition to the B*81:01/B*42:01 example cited above, B*58:01 selected for
352 S309A in Gag-QW9 (QASQEVKNW), while B*53:01 selected for A309X.

353 The remaining 267 (96%) examples of differential HLA-associated escape within the same
354 epitope represented cases where one allele was significantly associated with a polymorphism at a given
355 position and the other was not. Although some of these could represent cases of escape varying by
356 degree where statistical power was insufficient to detect it, the observation that 182 (65% of total) of

357 these instances represent cases where the log odds ratios of the two alleles are in opposite directions
358 argues against this interpretation in most cases. Similarly, although some of these could represent cases
359 of “push-pull” escape where statistical power was insufficient to detect it, this is also not likely to be the
360 explanation in most cases. Specifically, because odds ratios simply reflect the odds of selection among
361 individuals who express the allele versus individuals who do not, observation of a statistically
362 insignificant negative odds ratio by one allele alongside a significant positive odds ratio by another does
363 not necessarily imply active selection against the polymorphism by the former allele. More likely, these
364 insignificant negative odds ratios indicate a complete lack of selection on the part of the former
365 restricting allele. What can thus be clearly concluded from the data is that at least 184 of 278 (66%)
366 cases of observed differential selection represents instances in which the two HLA alleles drive distinct
367 escape pathways within the epitope, as evidenced by opposing odds ratios.

368 **Differential escape among protective B58 supertype alleles**

369 We next used this approach to study in detail the escape pathways selected by the clinically important
370 B58 supertype alleles B*57:02, B*57:03 and B*58:01 (note that B*57:01 frequency is negligible in
371 African populations). We systematically compared the odds ratio of escape among the three alleles for
372 every significant association reported in Table S3 (**Figure 2**; q-values computed separately for this
373 analysis). The results highlight widespread variation in the selection patterns of these alleles, with an
374 estimated 49% of comparisons representing true differences. For example, B*58:01, but not B*57:02 or
375 B*57:03, selects for escape in Gag-QW9, with escape occurring most strongly at positions 309 (S309A)
376 and 310 (T310S). These differences are statistically significant for T310S ($q < 0.05$) but not for S309X, for
377 which B*58:01-mediated escape is comparably weaker. Gag-KF11 represents another striking example,
378 with B*57:03 (but not B*57:02 or B*58:01) selecting for escape in positions -1, 2 and 4, and relatively
379 weak B*58:01-mediated selection at position 5 of the epitope. Gag-TW10 is the only epitope for which
380 all three alleles select for escape at the same position (T242N). At this position, we find that the odds of

381 escape are significantly higher for B*57:03 than for B*58:01 ($q=0.05$) and possibly B*57:02 ($q=0.2$); no
 382 differences were observed between B*57:02 and B*58:01 ($q>0.4$). B*57:03 selects for I247V whereas
 383 B*57:02 selects for I247M and B*58:01 does not appear to select for escape at this position. Rather,
 384 B*58:01 selects for 248A (which is the HIV-1 subtype C consensus residue), whereas there is no selection
 385 mediated at this position by B*57:02 or B*57:03. In the Gag-IW9 epitope, B*57:02 and B*57:03 both
 386 exhibit stronger selection pressure than B*58:01 at both positions 146 and 147 ($q<0.001$). No significant
 387 differences between B*57:02 and B*57:03 were detected in this epitope, likely due to the relatively
 388 small number of individuals expressing B*57:02 ($q>0.2$ for all comparisons).

389 **Differential targeting frequency does not explain differential escape**

390 Selection of escape indicates that at least some individuals expressing the restricting allele have CTL that
 391 target the epitope in question. However, absence of escape patterns at the population level does not
 392 necessarily imply a lack of targeting, nor do differential odds of escape necessarily imply differential
 393 odds of targeting. These observations are particularly evident for the B58-supertype epitopes, for which
 394 targeting was recently studied in detail (46). Comparing published B58 supertype-associated epitope
 395 targeting frequencies (46) with corresponding log odds ratios of escape (Figure 2) reveals several
 396 notable observations. First, the observation that Gag-KF11 is under strong B*57:03-mediated selection
 397 at multiple positions, whereas it is under only weak B*58:01-mediated selection and no B*57:02-
 398 mediated selection, is consistent with the observation that CTL frequently targeted KF11 when the
 399 epitope was presented by B*57:03, but rarely targeted KF11 when presented by B*58:01 and never
 400 targeted KF11 when presented by B*57:02 (46). In contrast, despite frequent targeting of RT-IW9 by
 401 both B*58:01- and B*57:03- (but not B*57:02-) restricted CTL (46), B*58:01 exhibits significantly higher
 402 odds of escape than either of the B*57 alleles at multiple positions within the epitope. Moreover, odds
 403 of B*57:03-mediated T242N escape within Gag-TW10 are significantly higher compared to B*58:01,
 404 despite the observation that B*58:01+ individuals target this epitope more frequently than do B*57+

405 individuals (46) (although decline of CTL responses following rapid escape in acute/early infection could
406 provide an alternative explanation (1), as could the selection of the alternative 248A escape
407 polymorphism in B*58:01+ individuals).

408 To test if odds of escape are correlated with odds of epitope targeting across all alleles in our
409 study, we analyzed a dataset of 1,010 adults with chronic C clade infection screened for responses to a
410 panel of 18mer peptides overlapping by 10 amino acids using IFN- γ ELISPOT assays. Defining odds of
411 escape for a given HLA allele in a given epitope as the maximum absolute log-odds ratio over all
412 significant HLA-associated polymorphisms in the epitope, we observed no correlation between odds of
413 escape and odds of ELISPOT response ($R^2 < 0.01$). When we compare the odds of observing an ELISPOT
414 response between two alleles exerting selection pressure on the same codon but to potentially varying
415 degrees (all allele pairs from Figure 1 for which the sign of the log-odds ratios is the same for both
416 alleles), we observed a weak negative trend between ELISPOT response frequency and odds of escape
417 ($p = 0.02$, binomial test, data not shown). Although OLP data are inherently noisy, owing to the presence
418 of multiple optimal epitopes per 18mer, these data support the observation that differential escape is
419 primarily the result of the selection of different escape pathways rather than differential frequencies of
420 epitope targeting during chronic infection.

421 **Risk of escape is not affected by HLA co-expression**

422 We hypothesized that the risk of escape could be modulated by the co-expression of other alleles. For
423 example, a subdominant epitope may be less likely to be targeted (and thus escape) if the individual co-
424 expresses an HLA allele that restricts one or more strongly immunodominant epitopes. Alternatively the
425 risk of escape may change if two overlapping epitopes are targeted at the same time. To test this
426 hypothesis, we devised a statistical test that utilized a multiplicative interaction term between two
427 alleles. Although several tests had $p < 0.001$, these were not significant after correcting for multiple tests
428 ($q > 0.9$ over 13 545 tests; data not shown). We next hypothesized that individuals who are homozygous

429 for a restricting allele will be more likely to escape. Once again, we observed no clear trends in the data
430 (7 associations with $0.2 < q < 0.6$, the rest with $q > 0.9$; data not shown). Overall these results indicate
431 that modulation of immune escape by HLA allele homozygosity or co-expression is not a general
432 phenomenon; however, the observation of a number of results with low p-values indicates that such
433 interactions could occur in specific cases, though the present study is underpowered to identify such
434 rare effects (note that the relationship between p- and q- values is a function of the number of tests
435 exceeding the significance of a given p-value relative to the total number of tests).

436 **Risk of escape is independent of cohort**

437 One possible cause of differential escape is within-host T-cell receptor diversity, a factor that could also
438 vary by population studied. Such variations could arise due to population-specific genetic characteristics
439 or variations in antigenic exposure arising from region-specific vaccinations or diseases. Although we
440 cannot explore the impact of TCR diversity on escape at the individual level, it is possible to investigate
441 whether population level differences could confound the present analyses. To test this, we recomputed
442 differential escape p- and q- values while conditioning on the cohort for which each individual was
443 recruited. The resulting q-values were nearly identical to the original analysis ($R^2 = 0.99$, data not
444 shown), indicating that differential escape could not be explained by region specific variations (as
445 approximated by cohort). We next tested if the odds of escape mediated by a specific allele were
446 dependent on either cohort or country of origin (excluding the heterogeneous Thames Valley Cohort).
447 Once again, no significant cohort effects were observed (minimum $q=1$ for both tests). Taken together,
448 we found no evidence for odds of escape being a function of cohort or country of origin, suggesting that
449 the dominant causal mechanism underlying the differential escape observed in the present study is
450 more closely linked to specific HLA alleles than any unmeasured attributes that would be expected to
451 correlate with ethnicity or region.

452 **Population escape patterns predict the majority of intra-epitopic variation**

453 The statistical evaluation of escape across individuals, such as the analyses described here, are
454 inherently biased towards identification of common pathways of escape. Although the large size of our
455 combined cohorts allows us to identify some uncommon escape pathways (over all associations,
456 frequency of escape in individuals with the associated HLA allele ranged from 1.6% to 100%, IQR 11%-
457 73%), very rare escapes, or rare escapes to uncommon HLA alleles, will go undetected (the statistical
458 power falls precipitously for HLA alleles occurring in fewer than 1% of the population; data not shown).

459 To investigate the ability of population-based approaches to detect evidence of rare escape, we
460 sought to identify whether optimal epitopes inherently display more sequence variation in individuals
461 expressing the restricting allele compared to those who do not. For each optimal epitope we tested for
462 association between expression of any of the restricting HLA alleles and the presence of at least one
463 non-consensus residue within the epitope, excluding at defined escape sites. This analysis will therefore
464 identify epitopes in which variation commonly or occasionally occurs at any epitope position not
465 identified in our previous analyses. Only 32 of 90 (36%) epitopes exhibited signs of increased general
466 variation among individuals expressing the relevant HLA allele ($q < 0.2$). The majority of these ($N=24$)
467 were in Pol, for which the present study had the least statistical power due to low sequence coverage
468 (e.g. integrase sequences were only available for 344 individuals). Overall, the median proportion of
469 HLA-matched individuals with a non-consensus residue at ≥ 1 non-HLA-associated site was 18%
470 compared to 13% in HLA mismatched individuals. To provide context, the median proportion of HLA-
471 matched individuals with a non-consensus residue at ≥ 1 HLA-associated site was 40%. This analysis
472 suggests that the majority of escape mutations within HLA-optimal epitope pairs analyzed in this study is
473 captured by the list of HLA-associated polymorphisms in Table S3, but also supports the selection of
474 unidentified rare escape pathways in some cases. This conclusion is broadly in line with a previous
475 report on longitudinal acute clade B data, in which 32-58% of observed substitutions (those achieving

476 >25% frequency in a given quasispecies, as limited by “bulk” RT-PCR and sequencing protocols (47, 48))
477 in the first two years of infection exactly matched predicted HLA-associated polymorphisms identified in
478 a chronically infected clade B cohort (13). Restricting that analysis to substitutions occurring inside
479 optimally defined, HLA-matched epitopes shows that 80%, 52% and 43% of intra-epitopic substitutions
480 in Nef, Gag and Pol, respectively, are attributable to HLA-associations used in that study ((13) and
481 unpublished data).

482 Taken together, these data suggest that population studies with statistical power comparable to
483 the present study are able to identify the majority of common escape mutations occurring in optimally
484 defined epitopes, as well as some rarer mutations that smaller studies have missed. There is also,
485 however, evidence of intra-epitopic variation that is not captured by the present study and which may
486 confer immune escape. It is unknown to what extent such rare escape pathways play a role in immune
487 evasion. Furthermore, the current study focused exclusively on well-characterized epitopes, which may
488 be more conserved than uncharacterized epitopes and may therefore display less variability in escape
489 patterns.

490 **Alleles exhibiting differential escape exhibit discordant associations with viral** 491 **load**

492 The B58 supertype alleles B*57:03 and B*58:02 exhibit opposing correlations with plasma viral load (VL)
493 in clade C infection, with B*57:03 strongly correlated with low VL and B*58:02 strongly correlated with
494 high VL (44, 49). These two alleles restrict completely different epitopes in HIV-1, which may account for
495 these differences. Likewise, the B7 epitopes B*81:01, B*42:01 and B*07:01, which select for differential
496 escape patterns within shared epitopes also exhibit discordant associations with VL (44, 49, 50). We thus
497 hypothesized that similar HLA alleles that select differential escape mutations within the same epitope
498 will commonly exhibit discordant associations with VL.

499 We therefore analyzed a dataset of 1,870 chronically C-clade infected, antiretroviral naïve adult
500 Africans to test for associations between HLA alleles and VL. We first sought to identify which HLA alleles
501 are independently and significantly associated with viral load. To this end, we tested all HLA subtypes
502 using forward selection on a linear regression model, conditioned on the cohorts from which each
503 sample was derived, with \log_{10} VL as the dependent variable. From the distribution of p-values, we
504 estimate that 20% of the 98 HLA alleles tested are truly associated with VL. Using $p < 0.05$ ($q < 0.13$) as a
505 threshold, we identified 20 HLA alleles that contribute to VL. These alleles were jointly added to a linear
506 regression model to determine their independent contributions to VL (**Figure 3A**). Eight of these alleles
507 were associated with reduced VL (“protective” alleles), while 12 were associated with increased VL
508 (“hazardous” alleles). Of note, 6 of the 12 (50%) hazardous alleles selected for escape in an epitope that
509 was also restricted by at least one protective allele and 5 of those cases were classified as differential
510 escape.

511 Simply identifying HLA alleles independently and significantly associated with VL, however, may
512 be overly conservative. Indeed, two alleles that are not individually significantly associated with VL may
513 have significantly discordant associations with VL if, for example, one allele tends to increase while the
514 other tends to decrease VL. We therefore tested for discordant associations between HLA alleles and VL
515 using the linear analogue of the differential selection model (with no correction for phylogeny, as none
516 was needed). To reduce the possibility of confounding due to linkage disequilibrium, we conditioned all
517 tests on the set of HLA alleles individually associated with VL (those in Figure 3A). Using this model, an
518 estimated 35% of HLA alleles that restrict the same epitope but select for differential escape also have
519 discordant associations with VL. Twenty-seven pairs were significant at $q < 0.2$ ($p < 0.1$; **Table S5**), and 11
520 were significant at $q < 0.05$ ($p < 0.011$; **Figure 3B**). These differences were dominated by members of the
521 A1, A3, B7 and B58 supertypes. Thus, these results indicate that similar HLA alleles that restrict the same
522 epitope, yet select for different escape pathways, often have discordant associations with viral load.

523 We next looked at whether various features of escape or targeting differentiated protective HLA
524 alleles from hazardous ones. For this analysis, we built a single linear model that included all HLA alleles
525 from Figure 3A and B except the HLA-C alleles (for which there are few published epitopes), and
526 interpreted the β estimates as the relative contribution of each allele to VL. We then correlated various
527 HLA allele features against these β estimates. Over all 32 HLA alleles there was a strong correlation
528 between the total number of Gag-OLPs associated with the allele and VL contribution (Spearman $\rho=-$
529 0.50, $p=0.006$), and a weak association between Pol/Nef-OLPs with VL contribution ($\rho=-0.41$, $p=0.03$;
530 **Figure 4A**). An even stronger correlation was observed between VL contribution and the total number of
531 optimal epitopes with associated escape polymorphisms in both Gag ($\rho=-0.72$, $p=1.7\times 10^{-5}$) and Pol/Nef
532 ($\rho=-0.46$, $p=0.01$; **Figure 4B**). Furthermore, the total number of escape polymorphisms observed per
533 epitope across Gag/Pol/Nef was strongly correlated with VL contribution in HLA-B alleles ($\rho=-0.77$,
534 $p=3.5\times 10^{-4}$) but not HLA-A alleles ($\rho=-0.04$, $p=0.9$; **Figure 4C**), and the overall strength of escape
535 associations was more statistically significant in protective alleles (median $q=0.001$) than in hazardous
536 alleles (median $q=0.03$; $p=0.003$, Mann-Whitney test). Of note, there was no difference in the entropy of
537 epitopes restricted by protective vs. hazardous alleles ($p=0.38$), nor was there any difference in the
538 entropy at the sites of associated escape ($p=0.96$). Taken together, these results indicate that the
539 presence of HLA-associated polymorphisms at the population level is a marker of effective epitope
540 targeting, especially among CTL that target HLA-B restricted Gag epitopes.

541 Although escape at the population level may indicate that CTL restricted by an HLA allele can be
542 quite effective, escape in an individual may indicate that the epitope can no longer be effectively
543 targeted in that individual. We therefore tested each HLA-associated polymorphism for an association
544 with viremic controller status (VL<2000 copies/ml and CD4 counts >250), using the interaction model
545 described in Methods. Although only four associations were significant at $q<0.2$ (data not shown), the
546 overall trends were striking. Consistent with observations of reduced escape in clade B infected elite

547 controllers (59), 201 of 300 (67%) tests indicated that viremic controllers were less likely to have
548 selected for a given escape than were non-controllers ($p=3.9 \times 10^{-9}$); 13 of 15 (88%; $p=0.0002$)
549 associations with $q < 0.5$ indicated that viremic controllers were less likely to have selected for escape.
550 This effect was largely driven by conserved regions: when a site is relatively conserved, viremic
551 controllers were much less likely to escape than were non-controllers, whereas the odds of escape was
552 similar between the two groups in non-conserved regions (Spearman correlation between entropy and
553 relative log-odds of escape between controllers and progressors was $\rho = -0.31$, $p = 0.0002$; data not
554 shown). Of note, protective alleles were not more likely than other alleles to exhibit differential odds of
555 escape between viremic controllers and progressors ($p = 0.77$, Fisher's exact test).

556 Discussion

557 The present study represents the first large scale, systematic analysis of differential immune escape in
558 HIV-1. Starting with optimally defined, published epitopes (52), we identified all related HLA alleles
559 driving immune escape mutations in Gag (p17+p24), Pol and Nef. This list included 38 epitopes
560 restricted by more than one HLA allele, which underscores the promiscuous nature of many CTL
561 epitopes (17, 30, 54, 66, 77). Remarkably, distinct mutational patterns and risk of escape were observed
562 in 37 of 38 of those epitopes, indicating that differential escape within promiscuous epitopes is typical.
563 These numbers are almost certainly underestimates resulting from restricting the study to known,
564 optimally defined epitopes.

565 There are several reasons why the odds of selecting a given escape polymorphism may differ
566 based on the specific HLA allele restricting the epitope. One possibility is that epitope targeting
567 frequency differs based on the restricting HLA allele. If this were the case, then differential selection
568 pressure would tend to be simply a matter of degree, with the more frequently-targeted HLA restricted-
569 epitopes exhibiting higher odds of escape. Although we do observe a small number of distinct escape

570 patterns that that can be explained in this straightforward way (e.g., B*57:03, B*57:02 and B*58:01 all
571 select for T242N escape in Gag-TW10, but to differing degrees), the vast majority cannot. Furthermore,
572 in the relatively uncommon cases where two alleles select for the same amino acid polymorphism, no
573 correlation between odds of escape and odds of OLP targeting in chronic infection was observed
574 (although the abrogation of CTL responses following escape in vivo must be acknowledged as a potential
575 limitation of this analysis). Instead, between 66% and 97% of observed cases of differential escape
576 reflect instances where two alleles select for different polymorphisms at the same site or at different
577 sites within the epitope. Taken together, our observations indicate that differential immune selection by
578 closely-related alleles is a widespread phenomenon, and one that typically manifests itself via distinct
579 escape pathways selected by the restricting HLA alleles, rather than common escape patterns that differ
580 in their relative risk of occurrence. This observation is in line with previous studies, which have reported
581 variations in functional avidity, TCR usage, and selection pressure, even in the absence of differential
582 targeting frequency, for several B7- and B58-restricted epitopes (50, 53, 82).

583 Differential selection and epitope targeting between related HLA alleles suggests that such
584 alleles will have discordant associations with viral load: indeed, this turns out to be true in
585 approximately 35% of cases in which HLA alleles exhibit distinct escape patterns within the same
586 restricted epitope. As such, our results complement previously-described discordant associations with
587 VL among alleles of the B58 supertype (2, 42, 44, 49), at least some of which appear to be due to the
588 specific epitopes restricted by each allele (25). Differential escape mutations within A2- (40), B58- (51,
589 53, 57, 82) and B7- (50, 82) restricted epitopes have also been previously reported, while case studies of
590 individual epitopes have linked differential escape pathways with discordant clinical outcomes (40) and
591 recruitment of distinct TCR repertoires exhibiting differential functional avidities (50). The present study
592 extends these observations by revealing that discordant associations with viral load are common among
593 closely related HLA alleles restricting different epitopes and/or selecting for different escape mutations.

594 Historically, the relationship between immune escape and disease progression has been difficult
595 to elucidate. The complexities of these relationships are illustrated by case studies describing loss of
596 viral control following escape within the immunodominant B*27-restricted Gag-KK10 epitope (27, 35,
597 43), followed by a dramatic broadening of the CTL response (27) (though breadth of targeting appears to
598 wane as many individuals progress to AIDS (38)). Thus, in these instances, KK10 escape appears to be a
599 direct cause of viral breakthrough, whereas any escape in epitopes targeted by the subsequent
600 broadened response would occur only after the VL increase. The complexities are compounded by the
601 observation that escape is typically a marker of an (at least previously) effective in vivo CTL response
602 (40). Indeed, expression of HLA class I alleles associated with a large number of population-level Gag
603 escape associations (16, 31, 65), a large number of reverting associations (56), and/or a large number of
604 associations in conserved regions (79), is predictive of relative viral control. Although escape inherently
605 implies a net improvement of in vivo viral fitness, a number of escape polymorphisms have been linked
606 to decreased in vitro (12, 23, 53, 62, 68, 78), and in vivo (31, 56) fitness in the absence of CTL pressure,
607 suggesting an incomplete recovery of viral replicative capacity upon escape. Epidemiologically, the
608 presence of costly escape positions could thus be a marker for immune control, as they identify cases of
609 partial immune-mediated attenuation of HIV-1 (58, 65). Over all associations in the present study,
610 escape was strongly linked to higher VL, an effect that was primarily driven by escape in conserved
611 regions. However, HLA alleles that were associated with many escape polymorphisms, especially in Gag,
612 were themselves associated with low viral load, a correlation that was much stronger than that
613 observed with OLP-measured targeting of Gag. Taken together, these data suggest that, although the
614 presence of population-level escape associations is a marker of the capacity of CTL restricted by that
615 allele to effectively target the virus, loss of viral control is closely linked to actual immune escape in
616 individuals, as was suggested in a chronically infected clade B cohort (16) and in elite controllers (59).
617 Thus, the study of immune escape in general, and differential escape in particular, may shed light on

618 which epitopes are most effective to target in vivo. From a vaccine design perspective, it is equally
619 important to determine if it is possible to block escape from occurring, either through a polyvalent
620 vaccine that primes the immune system to recognize escape variants (29), or by constraining escape
621 pathways by blocking compensatory mutations through the targeting of other epitopes (80). The
622 prospects of the latter approach may appear dim given that we found no instances in which the odds of
623 escape were reduced in the context of the co-expression of another HLA allele; however, the present
624 study was underpowered to identify such associations due to the large number (>13 000) of required
625 statistical tests and the low frequency of any given pair of HLA alleles. Some of these associations may
626 represent true interactions and the analytical tool developed here may prove useful for future studies
627 that consider a more restricted set of hypotheses.

628 One key assumption of the present study is that similar HLA alleles that restrict an epitope in a
629 given region are likely to restrict the same optimal epitope. Violations of this assumption could lead to
630 spurious identification of differential escape. Although this assumption remains largely untested, there
631 are several lines of evidence supporting its validity in the majority of cases. First, HLA supertype
632 definitions derive from shared binding profiles and epitope repertoires (17, 30, 71, 72, 77). The
633 observation that supertypes tend to restrict the same epitopes has been demonstrated in a number of
634 studies (3, 10, 32, 46, 50, 66, 77) and detailed studies of B7- (50) and B58- (46) supertypes consistently
635 yielded identical optimal epitope definitions when multiple alleles were associated with the same OLP.
636 Furthermore, many of the optimal epitopes used in the present study were previously tested in a cohort
637 of 103 HIV-infected individuals (30). In addition to observing widespread promiscuity, titration
638 experiments using truncated and extended peptides demonstrated that the same optimal epitope was
639 presented in the majority of cases, though several exceptions were noted. Moreover the same epitope
640 was frequently optimal for alleles even of different loci, an effect that may be due to HLA-independent
641 mechanisms such as proteasomal processing, epitope transport or trimming (26, 61, 75), suggesting that

642 our present approach of limiting epitope expansion to supertype members is conservative. Taken
643 together, the identification of an HLA-associated polymorphism within an optimal epitope known to be
644 restricted by a similar HLA allele suggests that the associated HLA restricts the same optimal epitope.
645 Nevertheless, a handful of known counter examples exist in the published optimal list, indicating that
646 some instances of differential escape may be due to related alleles restricting overlapping epitopes.
647 Future work is therefore required to validate proposed novel restrictions and to disentangle the causal
648 mechanisms of apparent differential escape.

649 These studies were facilitated by a novel statistical model that enables quantifying and
650 comparing the odds of immune escape while correcting for statistical confounding that may arise due to
651 phylogenetic relatedness of HIV sequences. This model was first developed to compare the odds of
652 escape between individuals who have progressed to AIDS and those who have not (38) and was here
653 refined and extended to model differential escape. The resulting model is quite versatile, enabling direct
654 tests for differential selection between two HLA alleles or differential selection mediated by one allele in
655 various genetic or environmental contexts. The present studies demonstrate the widespread extent of
656 differential escape in a relatively homogeneous population. Natural extensions will include studies of
657 how escape varies among ethnic populations or viral clades, and studies of differential escape in the
658 context of genetic variation outside the MHC-I locus or in the context of environmental factors,
659 including antiretroviral treatment, which may alter immune function or the virus' ability to tolerate
660 variation. A webserver implementation of the differential escape methods described herein is available
661 at <http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/phyloDOddsRatio/>.

662 Widespread differential immune selection pressure mediated by the specific HLA allele
663 restricting the epitope raises additional challenges for an epitope-based CTL vaccine. Differential escape
664 has been linked to differential CTL functional avidity (50) and in vivo efficacy (40), and the present study

665 indicates that differential escape may be broadly related to differential viral control. These observations
666 raise the possibility that an epitope-based vaccine will have varying results in different individuals,
667 potentially reducing the efficacy of the vaccine or even representing a hazard to certain individuals by
668 focusing their immune system on an ineffective response (50). In cases where differential escape has no
669 direct in vivo consequence, understanding the specifics may help in the design of a polyvalent vaccine,
670 as the escape routes of all common and rare alleles could be included in the vaccine (29). Although the
671 present study confirms and extends our understanding of the nature and impact of differential immune
672 selection by closely related HLA alleles, a number of limitations merit mention. The present study
673 focused only on known optimal epitopes in Gag, Pol and Nef, and was restricted to a cohort of clade C
674 infected individuals. Furthermore, working with high resolution HLA data reduces statistical power for
675 most rare alleles, a problem that is quadratically compounded when co-expression of high resolution
676 types is considered. Finally, although the large number of associations identified in this and other
677 studies suggests that many escape polymorphisms are repeatedly selected in individuals expressing the
678 same allele, the present study also identified a number of novel, rare escapes and suggested the
679 presence of even rarer undetected escapes. It is unknown to what extent such rare escapes occur in
680 vivo, to what extent they contribute to immune evasion, or whether their selection is attributable to
681 specific environmental or genetic contexts. Large data sets that include thousands of ethnically diverse
682 individuals, coupled with expanded high-fidelity epitope data, will be necessary to fully appreciate the
683 extent and specifics of differential immune escape and the implication of alternative escape pathways
684 on vaccine design.

685 **Acknowledgements/Funding:**

686 The authors would like to thank Chanson Brumme and Henrik Kloverpris for helpful discussions. JMC, JL,
687 NP, VT, CK and DH are employed by Microsoft Corp. ZLB is supported by a New Investigator Award from

688 the Canadian Institutes of Health Research (CIHR). The Durban cohort was supported by the NIH
689 (Contract NO1-A1-15422, 2R01AI46995-06, R01AI067073), the South African AIDS Vaccine Initiative, and
690 the Mark and Lisa Schwartz Foundation. TN holds the South African DST/NRF Chair in Systems Biology of
691 HIV/AIDS.

692 **References**

1. **Allen T. M., M. Altfeld, S. C. Geer, E. T. Kalife, C. Moore, K. M. O. Sullivan, I. Desouza, M. E. Feeney, R. L. Eldridge, E. L. Maier, D. E. Kaufmann, M. P. Lahaie, L. Reyor, G. Tanzi, M. N. Johnston, C. Brander, R. Draenert, J. K. Rockstroh, H. Jessen, E. S. Rosenberg, S. A. Mallal, and B. D. Walker.** 2005. Selective Escape from CD8+ T-Cell Responses Represents a Major Driving Force of Human Immunodeficiency Virus Type 1 (HIV-1) Sequence Diversity and Reveals Constraints on HIV-1 Evolution. *Journal of virology* **79**:13239-13249.
2. **Altfeld M., M. M. Addo, E. S. Rosenberg, F. M. Hecht, P. K. Lee, M. Vogel, X. G. Yu, R. Draenert, M. N. Johnston, D. Strick, T. M. Allen, M. E. Feeney, J. O. Kahn, R. P. Sekaly, J. a Levy, J. K. Rockstroh, P. J. Goulder, and B. D. Walker.** 2003. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* **17**:2581-2591.
3. **Altfeld M., A. Trocha, R. L. Eldridge, E. S. Rosenberg, M. N. Phillips, M. M. Addo, R. P. Sekaly, S. A. Kalams, S. A. Burchett, K. McIntosh, B. D. Walker, and P. J. Goulder.** 2000. Identification of dominant optimal HLA-B60- and HLA-B61-restricted cytotoxic T-lymphocyte (CTL) epitopes: rapid characterization of CTL responses by enzyme-linked immunospot assay. *Journal of virology* **74**:8541-9.
4. **Avila-Rios S., C. E. Ormsby, J. M. Carlson, H. Valenzuela-Ponce, J. Blanco-Heredia, D. Garrido-Rodriguez, C. Garcia-Morales, D. Heckerman, Z. L. Brumme, S. Mallal, M. John, E. Espinosa, and G. Reyes-Teran.** 2009. Unique features of HLA-mediated HIV evolution in a Mexican cohort: a comparative study. *Retrovirology* **6**:72.
5. **Bansal A., J. M. Carlson, J. Yan, O. T. Akinsiku, M. Schaefer, S. Sabbaj, A. Bet, D. N. Levy, S. Heath, J. Tang, R. a Kaslow, B. D. Walker, T. Ndung'u, P. J. Goulder, D. Heckerman, E. Hunter, and P. a Goepfert.** 2010. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *The Journal of experimental medicine* **207**:51-9.
6. **Barber L. D., B. G. Castro, L. Percival, X. Li, C. Clayberger, and P. Parham.** 1995. Overlap in the repertoires of peptides bound in vivo by a group of related class I HLA-B allotypes. *Current Biology* **5**:179-190.

7. **Barber L., L. Percival, K. Arnett, J. Gumperz, L. Chen, and P. Parham.** 1997. Polymorphism in the alpha 1 helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *Journal of immunology* **158**:1660-1669.
8. **Benjamini Y., and Y. Hochberg.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **57**:289-300.
9. **Berger C. T., J. M. Carlson, C. J. Brumme, K. L. Hartman, Z. L. Brumme, L. M. Henry, P. C. Rosato, A. Piechocka-Trocha, M. A. Brockman, P. R. Harrigan, D. Heckerman, D. E. Kaufmann, and C. Brander.** 2010. Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *The Journal of experimental medicine* **207**:61-75.
10. **Bertoni R., J. Sidney, P. Fowler, R. W. Chesnut, F. V. Chisari, and A. Sette.** 1997. Human histocompatibility leukocyte antigen-binding supermotifs predict broadly cross-reactive cytotoxic T lymphocyte responses in patients with acute hepatitis. *The Journal of clinical investigation* **100**:503-13.
11. **Bhattacharya T., M. Daniels, D. Heckerman, B. Foley, N. Frahm, C. Kadie, J. M. Carlson, K. Yusim, B. McMahon, B. Gaschen, S. Mallal, J. I. Mullins, D. C. Nickle, J. Herbeck, C. Rousseau, G. H. Learn, T. Miura, C. Brander, B. D. Walker, and B. Korber.** 2007. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* **315**:1583-1586.
12. **Brockman M. A., A. Schneidewind, M. Lahaie, A. Schmidt, T. Miura, I. Desouza, F. Ryvkin, C. a Derdeyn, S. Allen, E. Hunter, J. Mulenga, P. a Goepfert, B. D. Walker, and T. M. Allen.** 2007. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. *Journal of virology* **81**:12608-18.
13. **Brumme Z. L., C. J. Brumme, J. M. Carlson, H. Streeck, M. John, Q. Eichbaum, B. L. Block, B. Baker, C. Kadie, M. Markowitz, H. Jessen, A. D. Kelleher, E. Rosenberg, J. Kaldor, Y. Yuki, M. Carrington, T. M. Allen, S. Mallal, M. Altfeld, D. Heckerman, and B. D. Walker.** 2008. 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. *Journal of virology* **82**:9216-27.
14. **Brumme Z. L., C. J. Brumme, D. Heckerman, B. T. Korber, M. Daniels, J. M. Carlson, C. Kadie, T. Bhattacharya, C. Chui, J. Szinger, T. Mo, R. S. Hogg, J. S. G. Montaner, N. Frahm, C. Brander, B. D. Walker, and P. R. Harrigan.** 2007. Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS pathogens* **3**:e94.
15. **Brumme Z. L., M. John, J. M. Carlson, C. J. Brumme, D. Chan, M. A. Brockman, L. C. Swenson, I. Tao, S. Szeto, P. Rosato, J. Sela, C. M. Kadie, N. Frahm, C. Brander, D. W.**

- Haas, S. a Riddler, R. Haubrich, B. D. Walker, P. R. Harrigan, D. Heckerman, and S. Mallal.** 2009. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PloS one* **4**:e6687.
16. **Brumme Z. L., I. Tao, S. Szeto, C. J. Brumme, J. M. Carlson, D. Chan, C. Kadie, N. Frahm, C. Brander, B. D. Walker, D. Heckerman, and P. R. Harrigan.** 2008. Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. *AIDS* **22**:1277-86.
17. **Burrows S. R., R. A. Elkington, J. J. Miles, K. J. Green, S. Walker, S. M. Haryana, D. J. Moss, H. Dunckley, J. M. Burrows, and R. Khanna.** 2003. Promiscuous CTL recognition of viral epitopes on multiple human leukocyte antigens: biological validation of the proposed HLA A24 supertype. *Journal of immunology* **171**:1407-1412.
18. **Carlson J. M., and Z. L. Brumme.** 2008. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. *Microbes and infection* **10**:455-61.
19. **Carlson J. M., Z. L. Brumme, C. M. Rousseau, C. J. Brumme, P. Matthews, C. Kadie, J. I. Mullins, B. D. Walker, P. R. Harrigan, P. J. R. Goulder, and D. Heckerman.** 2008. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS computational biology* **4**:e1000225.
20. **Carlson J. M., and D. Heckerman.** 2009. Estimating false discovery rates for contingency tables. Microsoft Research MSR-TR-2009-53.
21. **Carlson J. M., C. Kadie, S. Mallal, and D. Heckerman.** 2007. Leveraging hierarchical population structure in discrete association studies. *PloS one* **2**:e591.
22. **Carrington M.** 1999. HLA and HIV-1: Heterozygote Advantage and B*35-Cw*04 Disadvantage. *Science* **283**:1748-1752.
23. **Christie N. M., D. O. Willer, M. a Lobritz, J. K. Chan, E. J. Arts, M. a Ostrowski, A. Cochrane, M. a Luscher, and K. S. MacDonald.** 2009. Viral fitness implications of variation within an immunodominant CD8+ T-cell epitope of HIV-1. *Virology* **388**:137-46.
24. **Deeks S. G., and B. D. Walker.** 2007. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* **27**:406-16.
25. **Dinges W. L., J. Richardt, D. Friedrich, E. Jalbert, Y. Liu, C. E. Stevens, J. Maenza, A. C. Collier, D. E. Geraghty, J. Smith, Z. Moodie, J. I. Mullins, M. J. McElrath, and H. Horton.** 2010. Virus-specific CD8+ T-cell responses better define HIV disease progression than HLA genotype. *Journal of virology* **84**:4461-8.
26. **Endert P. van.** 2011. Post-proteasomal and proteasome-independent generation of MHC class I ligands. *Cellular and molecular life sciences* **68**:1553-67.

27. **Feeny M. E., Y. Tang, K. A. Roosevelt, A. J. Leslie, K. McIntosh, N. Karthas, B. D. Walker, and P. J. R. Goulder.** 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.
28. **Fellay J., K. V. Shianna, D. Ge, S. Colombo, B. Ledergerber, M. Weale, K. Zhang, C. Gumbs, A. Castagna, A. Cossarizza, A. Cozzi-Lepri, A. De Luca, P. Easterbrook, P. Francioli, S. Mallal, J. Martinez-Picado, J. M. Miro, N. Obel, J. P. Smith, J. Wyniger, P. Descombes, S. E. Antonarakis, N. L. Letvin, A. J. McMichael, B. F. Haynes, A. Telenti, and D. B. Goldstein.** 2007. A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**:944-7.
29. **Fischer W., S. Perkins, J. Theiler, T. Bhattacharya, K. Yusim, R. Funkhouser, C. Kuiken, B. Haynes, N. L. Letvin, B. D. Walker, B. H. Hahn, and B. T. Korber.** 2007. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. *Nature medicine* **13**:100-6.
30. **Frahm N., K. Yusim, T. J. Suscovich, S. Adams, J. Sidney, P. Hrabec, H. S. Hewitt, C. H. Linde, D. G. Kavanagh, T. Woodberry, L. M. Henry, K. Faircloth, J. Listgarten, C. Kadie, N. Jovic, K. Sango, N. V. Brown, E. Pae, M. T. Zaman, F. Bihl, A. Khatri, M. John, S. Mallal, F. M. Marincola, B. D. Walker, A. Sette, D. Heckerman, B. T. Korber, and C. Brander.** 2007. Extensive HLA class I allele promiscuity among viral CTL epitopes. *European journal of immunology* **37**:2419-33.
31. **Goepfert P. a, W. Lumm, P. Farmer, P. Matthews, A. Prendergast, J. M. Carlson, C. a Derdeyn, J. Tang, R. a Kaslow, A. Bansal, K. Yusim, D. Heckerman, J. Mulenga, S. Allen, P. J. R. Goulder, and E. Hunter.** 2008. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *The Journal of experimental medicine* **205**:1009-17.
32. **Goulder P. J., M. Bunce, P. Krausa, K. McIntyre, S. Crowley, B. Morgan, A. Edwards, P. Giangrande, R. E. Phillips, and A. J. McMichael.** 1996. Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS research and human retroviruses* **12**:1691-8.
33. **Goulder P. J. R., and D. I. Watkins.** 2004. HIV and SIV CTL escape: implications for vaccine design. *Nature Reviews Immunology* **4**:630-40.
34. **Goulder P. J. R., and D. I. Watkins.** 2008. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. *Nature Reviews Immunology* **8**:619-30.
35. **Goulder P. J. R., R. E. Phillips, R. A. Colbert, S. McAdam, G. Ogg, M. A. Nowak, P. Giangrande, G. Luzzi, B. Morgana, A. Edwards, A. J. McMichael, and S. Rowland-Jones.** 1997. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nature Medicine* **3**:212-217.

36. **Guindon S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**:307-21.
37. **Holdsworth R., C. K. Hurley, S. G. E. Marsh, M. Lau, H. J. Noreen, J. H. Kempenich, M. Setterholm, and M. Maiers.** 2009. The HLA dictionary 2008: a summary of HLA-A, -B, -C, -DRB1/3/4/5, and -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. *Tissue antigens* **73**:95-170.
38. **Huang K.-H. G., D. Goedhals, J. M. Carlson, M. A. Brockman, S. Mishra, Z. L. Brumme, S. Hickling, C. S. W. Tang, T. Miura, C. Seebregts, D. Heckerman, T. Ndung'u, B. Walker, P. Klenerman, D. Steyn, P. Goulder, R. Phillips, C. van Vuuren, and J. Frater.** 2011. Progression to AIDS in South Africa is associated with both reverting and compensatory viral mutations. *PLoS one* **6**:e19018.
39. **Huang K.-H. G., D. Goedhals, H. Fryer, C. van Vuuren, A. Katzourakis, T. De Oliveira, H. Brown, S. Cassol, C. Seebregts, A. McLean, P. Klenerman, R. Phillips, and J. Frater.** 2009. Prevalence of HIV type-1 drug-associated mutations in pre-therapy patients in the Free State, South Africa. *Antiviral therapy* **14**:975-84.
40. **Iversen A. K. N., G. Stewart-Jones, G. H. Learn, N. Christie, C. Sylvester-Hviid, A. E. Armitage, R. Kaul, T. Beattie, J. K. Lee, Y. Li, P. Chotiyarnwong, T. Dong, X. Xu, M. a Luscher, K. MacDonald, H. Ullum, B. Klarlund-Pedersen, P. Skinhøj, L. Fugger, S. Buus, J. I. Mullins, E. Y. Jones, P. A. van der Merwe, and A. J. McMichael.** 2006. Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope. *Nature immunology* **7**:179-89.
41. **John M., D. Heckerman, I. James, L. P. Park, J. M. Carlson, A. Chopra, S. Gaudieri, D. Nolan, D. W. Haas, S. a Riddler, R. Haubrich, and S. Mallal.** 2010. Adaptive interactions between HLA and HIV-1: highly divergent selection imposed by HLA class I molecules with common supertype motifs. *Journal of immunology* **184**:4368-77.
42. **Kaslow R. A., M. Carrington, R. Apple, L. Park, A. Muñoz, A. J. Saah, J. J. Goedert, C. Winkler, S. J. O'Brien, C. Rinaldo, R. Detels, W. Blattner, J. Phair, H. Erlich, and D. L. Mann.** 1996. Influence of combinations of human major histocompatibility complex genes on the course of HIV-1 infection. *Nature Medicine* **2**:405-411.
43. **Kelleher A. D., C. Long, E. C. Holmes, R. L. Allen, J. Wilson, C. Conlon, C. Workman, S. Shaunak, K. Olson, P. Goulder, C. Brander, G. Ogg, J. S. Sullivan, W. Dyer, I. Jones, a J. McMichael, S. Rowland-Jones, and R. E. Phillips.** 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**:375-86.
44. **Kiepiela P., A. J. Leslie, I. Honeyborne, D. Ramduth, C. Thobakgale, S. Chetty, P. Rathnavalu, C. Moore, K. J. Pfafferott, L. Hilton, P. Zimbwa, S. Moore, T. Allen, C. Brander, M. M. Addo, M. Altfeld, I. James, S. Mallal, M. Bunce, L. D. Barber, J. Szinger,**

- C. Day, P. Klenerman, J. Mullins, B. Korber, H. M. Coovadia, B. D. Walker, and P. J. R. Goulder.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**:769-75.
45. **Kiepiela P., K. Ngumbela, C. Thobakgale, D. Ramduth, I. Honeyborne, E. Moodley, S. Reddy, C. de Pierres, Z. Mncube, N. Mkhwanazi, K. Bishop, M. van der Stok, K. Nair, N. Khan, H. Crawford, R. Payne, A. Leslie, J. Prado, A. Prendergast, J. Frater, N. McCarthy, C. Brander, G. H. Learn, D. Nickle, C. Rousseau, H. Coovadia, J. I. Mullins, D. Heckerman, B. D. Walker, and P. Goulder.** 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nature medicine* **13**:46-53.
46. **Kloverpris H. N., A. Stryhn, M. Harndahl, M. van der Stok, R. P. Payne, P. C. Matthews, F. Chen, L. Riddell, B. D. Walker, T. Ndung'u, S. Buus, and P. Goulder.** 2012. HLA-B*57 Micropolymorphism Shapes HLA Allele-Specific Epitope Immunogenicity, Selection Pressure, and HIV Immune Control. *Journal of virology* **86**:919-29.
47. **Larder B. A., A. Kohli, P. Kellam, S. D. Kemp, M. Kronick, and R. D. Henfrey.** 1993. Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. *Nature* **365**:671-3.
48. **Leitner T., E. Halapi, G. Scarlatti, P. Rossi, J. Albert, E. M. Fenyö, and M. Uhlen.** 1993. Analysis of heterogeneous viral populations by direct DNA sequencing. *BioTechniques* **15**:120-7.
49. **Leslie A., P. C. Matthews, J. Listgarten, J. M. Carlson, C. Kadie, T. Ndung'u, C. Brander, H. Coovadia, B. D. Walker, D. Heckerman, and P. J. R. Goulder.** 2010. Additive contribution of HLA class I alleles in the immune control of HIV-1 infection. *Journal of virology* **84**:9879-88.
50. **Leslie A., D. A. Price, P. Mkhize, A. Rathod, C. Day, H. Crawford, I. Honeyborne, T. E. Asher, G. Luzzi, C. M. Rosseau, J. I. Mullins, V. Novelli, C. Brander, D. C. P. Kiepiela, B. D. Walker, P. J. R. K. Bishop, A. Edwards, G. Tudor-williams, D. C. Douek, and P. J. R. Goulder.** 2006. Differential Selection Pressure Exerted on HIV by CTL Targeting Identical Epitopes but Restricted by Distinct HLA Alleles from the Same HLA Supertype. *The Journal of Immunology* **177**:4699-4708.
51. **Leslie a J., K. J. Pfafferott, P. Chetty, R. Draenert, M. M. Addo, M. Feeney, Y. Tang, E. C. Holmes, T. Allen, J. G. Prado, M. Altfeld, C. Brander, C. Dixon, D. Ramduth, P. Jeena, S. a Thomas, A. St John, T. a Roach, B. Kupfer, G. Luzzi, A. Edwards, G. Taylor, H. Lyall, G. Tudor-Williams, V. Novelli, J. Martinez-Picado, P. Kiepiela, B. D. Walker, and P. J. R. Goulder.** 2004. HIV evolution: CTL escape mutation and reversion after transmission. *Nature medicine* **10**:282-9.
52. **Llano A., N. Frahm, and C. Brander.** 2009. How to optimally define optimal cytotoxic T lymphocyte epitopes in HIV infection?, p. 3-24. *In* K. Yusim, B.T. Korber, C. Brander, B.F. Haynes, R.A. Koup, J.P. Moore, B.D. Walker, and D.I. Watkins (eds.), *HIV Molecular*

Immunology. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico.

53. **Martinez-picado J., J. G. Prado, E. E. Fry, K. Pfafferoth, A. Leslie, S. Chetty, C. Thobakgale, I. Honeyborne, H. Crawford, P. Matthews, T. Pillay, C. Rousseau, J. I. Mullins, C. Brander, B. D. Walker, D. I. Stuart, P. Kiepiela, and P. Goulder.** 2006. Fitness Cost of Escape Mutations in p24 Gag in Association with Control of Human Immunodeficiency Virus Type 1. *Journal of virology* **80**:3617-3623.
54. **Masemola A. M., T. N. Mashishi, G. Khoury, H. Bredell, M. Paximadis, T. Mathebula, D. Barkhan, A. Puren, E. Vardas, M. Colvin, L. Zijenah, D. Katzenstein, R. Musonda, S. Allen, N. Kumwenda, T. Taha, G. Gray, J. McIntyre, S. A. Karim, H. W. Sheppard, and C. M. Gray.** 2004. Novel and Promiscuous CTL Epitopes in Conserved Regions of Gag Targeted by Individuals with Early Subtype C HIV Type 1 Infection from Southern Africa. *Journal of immunology* **173**:4607-4617.
55. **Matthews P. C., E. Adland, J. Listgarten, A. Leslie, N. Mkhwanazi, J. M. Carlson, M. Harndahl, A. Stryhn, R. P. Payne, A. Ogwu, K.-H. G. Huang, J. Frater, P. Paioni, H. Klooverpris, P. Jooste, D. Goedhals, C. van Vuuren, D. Steyn, L. Riddell, F. Chen, G. Luzzi, T. Balachandran, T. Ndung'u, S. Buus, M. Carrington, R. Shapiro, D. Heckerman, and P. J. R. Goulder.** 2011. HLA-A*7401-mediated control of HIV viremia is independent of its linkage disequilibrium with HLA-B*5703. *Journal of immunology* **186**:5675-86.
56. **Matthews P. C., A. Prendergast, A. Leslie, H. Crawford, R. Payne, C. Rousseau, M. Rolland, I. Honeyborne, J. M. Carlson, C. Kadie, C. Brander, K. Bishop, N. Mlotshwa, J. I. Mullins, H. Coovadia, T. Ndung'u, B. D. Walker, D. Heckerman, and P. J. R. Goulder.** 2008. Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. *Journal of virology* **82**:8548-59.
57. **Migueles S. A., A. C. Laborico, H. Imamichi, W. L. Shupert, C. Royce, M. McLaughlin, L. Ehler, J. Metcalf, S. Liu, C. W. Hallahan, and M. Connors.** 2003. The Differential Ability of HLA B * 5701+ Long-Term Nonprogressors and Progressors To Restrict Human Immunodeficiency Virus Replication Is Not Caused by Loss of Recognition of Autologous Viral gag Sequences. *Journal of virology* **77**:6889-6898.
58. **Miura T., M. A. Brockman, A. Schneidewind, M. Lobritz, F. Pereyra, A. Rathod, B. L. Block, Z. L. Brumme, C. J. Brumme, B. Baker, A. C. Rothchild, B. Li, A. Trocha, E. Cutrell, N. Frahm, C. Brander, I. Toth, E. J. Arts, T. M. Allen, and B. D. Walker.** 2009. HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte recognition. *Journal of virology* **83**:2743-55.
59. **Miura T., C. J. Brumme, M. A. Brockman, Z. L. Brumme, F. Pereyra, B. L. Block, A. Trocha, M. John, S. Mallal, P. R. Harrigan, and B. D. Walker.** 2009. HLA-associated viral mutations are common in human immunodeficiency virus type 1 elite controllers. *Journal of virology* **83**:3407-12.

60. **Moore C. B., M. John, I. R. James, F. T. Christiansen, C. S. Witt, and S. a Mallal.** 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* **296**:1439-43.
61. **Neeffjes J., M. L. M. Jongsma, P. Paul, and O. Bakke.** 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology* **11**:823-836.
62. **Peyerl F. W., H. S. Bazick, M. H. Newberg, D. H. Barouch, J. Sodroski, and N. L. Letvin.** 2004. Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *Journal of virology* **78**:13901-10.
63. **Phillips R. E., S. Rowland-Jones, D. F. Nixon, F. M. Gotch, J. P. Edwards, A. O. Ogunlesi, J. G. Elvin, J. A. Rothbard, C. R. Bangham, and C. R. Rizza.** 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. *Nature* **354**:453-9.
64. **Pounds S., and C. Cheng.** 2006. Robust estimation of the false discovery rate. *Bioinformatics* **22**:1979-87.
65. **Rousseau C. M., M. G. Daniels, J. M. Carlson, C. Kadie, H. Crawford, A. Prendergast, P. C. Matthews, R. P. Payne, M. Rolland, D. N. Raugi, B. S. Maust, G. H. Learn, D. C. Nickle, H. M. Coovadia, P. J. R. Goulder, T. Bhattacharya, D. E. Heckerman, B. T. Korber, J. I. Mullins, T. Ndung'u, N. Frahm, C. Brander, and B. D. Walker.** 2008. HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. *Journal of virology* **82**:6434-6446.
66. **Sabbaj S., A. Bansal, G. D. Ritter, C. Perkins, B. H. Edwards, E. Gough, J. Tang, J. J. Szinger, B. Korber, C. M. Wilson, R. a Kaslow, M. J. Mulligan, and P. a Goepfert.** 2003. Cross-reactive CD8+ T cell epitopes identified in US adolescent minorities. *Journal of acquired immune deficiency syndromes (1999)* **33**:426-38.
67. **Schneidewind A., M. A. Brockman, J. Sidney, Y. E. Wang, H. Chen, T. J. Suscovich, B. Li, R. I. Adam, R. L. Allgaier, B. R. Mothé, T. Kuntzen, C. Oniangue-Ndza, A. Trocha, X. G. Yu, C. Brander, A. Sette, B. D. Walker, and T. M. Allen.** 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**:5594-605.
68. **Schneidewind A., M. A. Brockman, R. Yang, R. I. Adam, B. Li, S. Le Gall, C. R. Rinaldo, S. L. Craggs, R. L. Allgaier, K. a Power, T. Kuntzen, C.-S. Tung, M. X. LaBute, S. M. Mueller, T. Harrer, A. J. McMichael, P. J. R. Goulder, C. Aiken, C. Brander, A. D. Kelleher, and T. M. Allen.** 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**:12382-93.

69. Shapiro R. L., M. D. Hughes, A. Ogwu, D. Kitch, S. Lockman, C. Moffat, J. Makhema, S. Moyo, I. Thior, K. McIntosh, E. van Widenfelt, J. Leidner, K. Powis, A. Asmelash, E. Tumbare, S. Zwierski, U. Sharma, E. Handelsman, K. Mburu, O. Jayeoba, E. Moko, S. Souda, E. Lubega, M. Akhtar, C. Wester, R. Tuomola, W. Snowden, M. Martinez-Tristani, L. Mazhani, and M. Essex. 2010. Antiretroviral Regimens in Pregnancy and Breast-Feeding in Botswana. *New England Journal of Medicine* **362**:2282-2294.
70. Sidney J., M. del Guercio, S. Southwood, V. Engelhard, E. Appella, H. Rammensee, K. Falk, O. Rotzschke, M. Takiguchi, and R. Kubo. 1995. Several HLA alleles share overlapping peptide specificities. *Journal of immunology* **154**:247-259.
71. Sidney J., B. Peters, N. Frahm, C. Brander, and A. Sette. 2008. HLA class I supertypes: a revised and updated classification. *BMC immunology* **9**:1.
72. Sidney J., S. Southwood, and A. Sette. 2005. Classification of A1- and A24-supertype molecules by analysis of their MHC-peptide binding repertoires. *Immunogenetics* **57**:393-408.
73. Storey J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**:479-498.
74. Storey J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100**:9440-5.
75. Tenzer S., E. Wee, A. Burgevin, G. Stewart-Jones, L. Friis, K. Lamberth, C.-hao Chang, M. Harndahl, M. Weimershaus, J. Gerstoft, N. Akkad, P. Klenerman, L. Fugger, E. Y. Jones, A. J. McMichael, S. Buus, H. Schild, P. van Endert, and A. K. N. Iversen. 2009. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nature immunology* **10**:636-46.
76. The International HIV Controllers Study. 2010. The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation. *Science* **330**:1551-1557.
77. Threlkeld S., P. Wentworth, S. Kalams, B. Wilkes, D. Ruhl, E. Keogh, J. Sidney, S. Southwood, B. Walker, and A. Sette. 1997. Degenerate and promiscuous recognition by CTL of peptides presented by the MHC class I A3-like superfamily: implications for vaccine development. *Journal of immunology* **159**:1648-1657.
78. Troyer R. M., J. McNevin, Y. Liu, S. C. Zhang, R. W. Krizan, A. Abraha, D. M. Tebit, H. Zhao, S. Avila, M. a Lobritz, M. J. McElrath, S. Le Gall, J. I. Mullins, and E. J. Arts. 2009. Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. *PLoS pathogens* **5**:e1000365.
79. Wang Y. E., B. Li, J. M. Carlson, H. Streeck, A. D. Gladden, R. Goodman, A. Schneidewind, K. a Power, I. Toth, N. Frahm, G. Alter, C. Brander, M. Carrington, B. D. Walker, M. Altfeld, D. Heckerman, and T. M. Allen. 2009. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in

highly conserved regions of human immunodeficiency virus type 1. *Journal of virology* **83**:1845-55.

80. Wang Y. E., C. Oniangue-Ndza, Y. Qi, A. Schneidewind, M. Kemper, E. Mellors, A. D. Gladden, K. A. Power, B. G. Pierce, P. McLaren, P. I. W. de Bakker, A. Rauch, A. Telenti, S. G. Deeks, D. W. Haas, X. Gao, F. Pereyra, B. D. Walker, M. Carrington, and T. M. Allen. Cooperative effects of combinations of HLA class I alleles on the control of HIV. In review.

81. Yang O. O., J. Church, C. M. R. Kitchen, R. Kilpatrick, A. Ali, Y. Geng, M. S. Killian, R. L. Sabado, H. Ng, J. Suen, Y. Bryson, B. D. Jamieson, and P. Krogstad. 2005. Genetic and Stochastic Influences on the Interaction of Human Immunodeficiency Virus Type 1 and Cytotoxic T Lymphocytes in Identical Twins. *Journal of virology* **79**:15368-15375.

82. Yu X. G., M. Lichterfeld, S. Chetty, K. L. Williams, S. K. Mui, T. Miura, N. Frahm, M. E. Feeney, Y. Tang, F. Pereyra, M. X. Labute, K. Pfafferoth, A. Leslie, H. Crawford, R. Allgaier, W. Hildebrand, R. Kaslow, C. Brander, T. M. Allen, E. S. Rosenberg, P. Kiepiela, M. Vajpayee, P. a Goepfert, M. Altfeld, P. J. R. Goulder, and B. D. Walker. 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**:1619-31.

693 **Figure Legends**

694 **Figure 1: Per-site differential escape between HLA alleles that restrict the same epitope.** Bars

695 represent the natural logarithm of the phylogenetically-corrected odds ratio. Values between -20 (red,
 696 extending to the left) and +20 (blue, extending to the right) are shown; infinite log-odds ratios are set to
 697 values of +/- 20. Associations that are individually significant are labeled with * (q<0.2) or ** (q<0.05).
 698 Polymorphisms are denoted in the SNP column; underlined amino acids signify cohort consensus. A
 699 phylogenetically-corrected logistic model was used to derive a p-value that tests the null hypothesis that
 700 both alleles select for escape with the same odds ratio. Comparisons with p<0.005 (q<0.006) are
 701 reported. The complete list of all comparisons with p<0.05 is available in Table S4.

702 **Figure 2: Differential escape among protective B58-supertype alleles.** The phylogenetically-corrected
 703 log-odds ratio of each B*57:02-, B*57:03-, or B*58:01-associated polymorphism from Table 1 was tested
 704 for differential selection against the other three alleles. Bars represent the log-odds of observing the

705 indicated polymorphism. Stars indicate that the magnitude of the odds ratio is significantly ($p < 0.05$,
 706 $q < 0.06$) greater than the allele indicated by the color of the star. Large amino acid letters indicate cohort
 707 consensus; small letters indicate alternative polymorphisms associated with at least one allele.

708 **Figure 3: Relative contributions (β parameters) of HLA alleles to log VL.** (A) all HLA alleles identified at
 709 $q < 0.2$ in a forward selection procedure, and (B) HLA-A and HLA-B alleles grouped by supertype, that
 710 were discordantly association with VL compared to an allele that was associated with differential escape
 711 in the same epitope, conditioned on the alleles from (A). Relative bar heights depict the maximum
 712 likelihood β estimates from a joint linear model (estimated corrected average VL among individuals
 713 expressing that allele), conditioned on cohort labels. Error bars represent standard error estimates for β .

714 **Figure 4: HLA alleles that select for escape are associated with reduced viral load.** Log viral load was
 715 modeled as a linear function of the HLA-A and -B alleles from Figure 3, and the resulting β estimates
 716 were correlated against (A) the number of OLP responses that associated with that allele, (B) the
 717 number of optimal epitopes that were targeted by that allele as determined by the presence of
 718 associated escape polymorphisms, and (C) the total number of escape polymorphisms per targeted
 719 optimal epitope. Spearman rank correlation coefficients (ρ) are reported for each plotted dataset.

Table 1: Associations between supertype members and polymorphisms in optimal epitopes at q<0.2.

Name	Optimal ^a	Consensus ^b	Location	HLA ^c	Associations ^d	Super-types ^e
KK9	KIRLRPGGK	KIRLRPGGK	18-26	A*03:01	A*03:01(R18K,K28X); A*30:01(X20S,C/K28Q);	A3
	RLRPGGKKK	RLRPGGKKH	20-28		A*31:01(X18Q); A*33:01(X18Q,X28Q);	
	RLRPGGKKKY	RLRPGGKKHY	20-29		A*74:01(R20K,R28X,Q30X)	
KW9	KYKLVHIVW	HYMLKHLVW	28-36	A*24:02	A*23:01(R28X,L34I); A*24:02(R28X)	A24
RY11	RSLYNTVATLY	RSLYNTVATLY	76-86	A*30:02 B58 B63	A*01:01(E73K,Y79F/H); A*29:02(L85I); A*29:11(C87X); A*30:02(T81A); A*36:01(Y79H); B*58:02(C87X)	A1,A24
LY9	LYNTVATLY	LYNTVATLY	78-86	A*29:02 B*44:03	A*01:01(Y79F/H); A*29:02(L85I); A*29:11(C87X); A*30:02(T81A); A*36:01(Y79H)	A1,A24
TK8	TLYCVHQB	TLYCVHEK	84-91	A*11:01	A*03:01(X91S); A*34:02(K90R); A*68:01(K90X); A*66:03(T81A,A83V); A*74:01(R91K,V94I)	A3
IL10	IEIKDTKEAL	IEV ^U RD ^T KEAL	92-101	B*40:01	B*41:01(E93V); B*44:03(X91Q); B*45:01(E93X)	B44
VL10	VHQAI ^S PR ^T L	VHQAI ^S PR ^T L	143-152	B*15:10	B*14:02(V143I,L147I); B*15:10(A146S)	B27
IW9	ISPRTLN ^A W	ISPRTLN ^A W	147-155	B*57:01 B63	B*57:02(A146P,I147L); B*57:03(A146P,I147L/M); B*58:01(X146P); B*58:02(S146X)	B58
KF11	KAFSPEVI KAFSPEVIPMF	KAFSPEVI KAFSPEVIPMF	162-169 162-172	B*57:03	B*57:03(X161D,A163G/S,S165K/N); B*58:01(V168X)	B58
EL9	EVIPMF ^S AL	EVIPMF ^I AL	167-175	A*26:01	A*26:01(V168X,T173M); A*29:02(X173M)	A1
TL9	TPQDLN ^T ML	TPQDLN ^T ML	180-188	B*07:02 B*39:10	B*39:10(E177D,T186X); B*42:01(X182T); B*67:01(T190A); C*08:02(X182H);	B7
	TPYDIN ^Q ML	TPQ ^D LN ^I ML		B*42:01 B*81:01 C*08:02	B*81:01(E177D,Q/T182E/G/S,T186S,L188F, T190X,V191I);	
TW10	TSTLQE ^Q I ^G W	TSTLQE ^Q I ^A W	240-249	B*57:01 B*58:01	B*57:02(T242N,X247M); B*57:03(T242N,I247V); B*58:01(T242N,L243X,X248A)	B58
NY10	NPPIPV ^G DIY	NPPIPV ^G DIY	253-262	B*35:01	B*35:01(D260E); B*39:10(I250M); B*53:01(X256T,R264X); B*81:01(X252A)	B7
QW9	QASQE ^V KNW	QAT ^Q D ^V KNW	308-316	B*53:01 B*57:01	B*53:01(A309X,N315G); B*58:01(S309A,T310S)	
DL9	DTVLEE ^W NL	DTVLEE ^I NL	30-38	A*68:02	A*02:02(X39S); A*02:05(X36V)	A2
AM9	ALVEI ^C TEM	AL ^T AI ^C E ^E EM	33-41	A*02:01	A*02:01(X35I,X36V,X40D); A*02:02(E36A,X41I)	A2
T18	TAFTI ^P SI	TAFTI ^P SI	128-135	B*51:01	B*51:01(I135T); B*81:01(X134G)	B7
KY9	KQNPDI ^V IY	A ^Q NP ^E I ^V IY	173-181	A*30:02	A*29:02(I178X); A*30:01(E173X); A*36:01(X178V)	A1
IL9	IEELR ^Q HLL	IEELR ^E HLL	202-210	B*40:01	B*18:01(E207X); B*44:03(E204X,E207K/N)	B44
QR9	QIYPGI ^K VR	QIYPGI ^K VR	269-277	A*03:01	A*03:01(K277R); A*30:01(Q278X); A*33:03(X273R); A*34:02(R275X)	A3
IW9	IAMESI ^V IW	IAMESI ^V IW	375-383	B*58:01	B*15:16(X386V); B*58:01(I375V,X377Q/T,X379A)	B58
EY9	ETKLG ^K AGY	ETK ^I G ^K AGY	449-457	A*26:01	A*26:01(K451R,M452X); A*29:01(T450N)	A1
RM9	RPQV ^P LR ^P M	RPQV ^P LR ^P M	71-79	B*42:01 B*07:02	B*35:01(Y81F); B*81:01(X71T,L76T/V)	B7
QK10	QVPLR ^P M ^T YK	QVPLR ^P M ^T YK	73-82	A*03:01 A*11:01	A*03:01(A83G,X85L); A*33:01(X71K); A*34:02(A83G); A*66:03(V70I); A*68:01(X82Q)	A3
RL9	VPLR ^P M ^T Y RPMTY ^K AAL	VPLR ^P M ^T Y RPMTY ^K AAL	74-81 77-85	B*35:01 B*07:02	B*35:01(Y81F); B*81:01(X71T,L76T/V)	B7
PK8	PLR ^P M ^T YK	PLR ^P M ^T YK	75-82	A*11:01	A*03:01(A83G,X85L); A*34:02(A83G); A*68:01(X82Q)	A3
KL10	KA ^A FD ^L S ^F FL	KA ^A FD ^L S ^F FL	82-91	B*57:03	B*57:02(A83X); B*57:03(A83X); B*58:01(A83G)	B58
AK9	AVD ^L SH ^F LK	A ^F D ^L S ^E ^F LK	84-92	A*03:01 A*11:01	A*03:01(A83G,X85L); A*34:02(A83G); A*68:01(X82Q)	A3
HW9	HTQGY ^F PDW	HTQGE ^F PDW	116-124	B*57:01	B*57:03(H116N); B*58:01(X116N)	B58
TL10	TPGPG ^V RY ^P L	TPGPG ^V RY ^P L	128-137	B*07:02 B*42:01	B*35:01(V133T); B*53:01(V133I)	B7
RW8	RYPL ^T FGW	RYPL ^T FGW	134-141	A*24:02	A*23:01(F143Y); A*24:02(Y135F/L)	A24
YY9	YPL ^T FG ^W CY	YPL ^T FG ^W C ^E	135-143	B*18:01 B*53:01	B*35:01(V133T); B*53:01(V133I)	B7

^a Optimally defined epitopes as defined in (http://www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html). Only optimal epitopes that were restricted by least two HLA alleles are shown. Overlapping optimals with published alleles in the same supertype are grouped together. ^b Consensus in the present cohort. Underlined residues mark differences from published optimal. ^c HLA alleles associated with the epitope as defined in the optimal epitope definitions. ^d HLA alleles that are related to the published alleles via supertype were tested for associations with polymorphisms within or flanking the optimal epitope. Associations are of the form (YposZ), where Y is a residue that is negatively associated with the HLA allele and Z is a residue that is positively associated with the allele. X indicates no association in the positive/negative direction. All associations with $q < 0.2$ are reported. ^e Supertypes that are represented by HLA alleles that are associated with escape.







