Supplementary Materials for

Influence of HLA-C Expression Level on HIV Control

Richard Apps, Ying Qi, Jonathan M. Carlson, Haoyan Chen, Xiaojiang Gao, Rasmi Thomas, Yuko Yuki, Greg Q. Del Prete, Philip Goulder, Zabrina L. Brumme, Chanson J. Brumme, Mina John, Simon Mallal, George Nelson, Ronald Bosch, David Heckerman, Judy L. Stein, Kelly A. Soderberg, M. Anthony Moody, Thomas N. Denny, Xue Zeng, Jingyuan Fang, Ashley Moffett, Jeffrey D. Lifson, James J. Goedert, Susan Buchbinder, Gregory D. Kirk, Jacques Fellay, Paul McLaren, Steven G. Deeks, Florencia Pereyra, Bruce Walker, Nelson L. Michael, Amy Weintrob, Steven Wolinsky, Wilson Liao, Mary Carrington*

*Corresponding author. E-mail: carringm@mail.nih.gov

Published 5 April 2013, *Science* **340**, 87 (2013)
DOI: 10.1126/science.1232685

This PDF file includes:

Materials and Methods

Samples

200 healthy donors of African American ancestry recruited at the Duke Human Vaccine Institute as part of the CHAVI 008A study were used for determination of HLA-C expression levels. Genetic effects of HLA-C expression were tested on clinical outcomes of 5,243 HIV^+ individuals derived from 9 cohorts: Multicenter AIDS Cohort Study (MACS) (27), Multicenter Hemophilia Cohort Study (MHCS) (28), San Francisco City Clinic Cohort (SFCCC) (29), AIDS Linked to Intravenous Experience (ALIVE) (30), Swiss HIV Cohort Study (SHCS) (http://www.shcs.ch), Study on the Consequences of Protease Inhibitor Era (SCOPE) (31), Massachusetts General Hospital (MGH) Controller Cohort (http://www.hivcontrollers.org/hivcontrollers), United States Military HIV Natural History Study (NHS) (7) and the DC gay cohort study (DCG) (32). Time from seroconversion to a $CD4^+$ cell count <200 per mm^3 was determined in 1,069 ART-naïve patients. Analysis of longitudinal VL in chronic infection was performed by averaging all available viral load readings prior to ART for both seroincident and seroprevalent subjects. Longitudinal mVL was analyzed as a continuous variable for 4,739 patients and also used to define 1,258 controllers with mVL <2,000 and 2,478 non-controllers with mVL >10,000 viral copies/ml plasma. Subjects used in the longitudinal VL analyses were largely independent from the patients used in progression analyses, with 87% of the 3,280 European American patients and 82% of the 1,459 African American patients unique to the longitudinal VL studies. HLA-C expression showed a significant independent effect even when only those individuals unique to the longitudinal viral load study were included in regression analyses that included all HLA class I alleles as covariables (≥2% phenotypic frequency). This held true in both European and African American populations tested separately when using mean longitudinal VL as a continuous variable and when treating controllers and non-controllers as a dichotomous variable (Table S7).

Nearly full length viral genomes and HLA genotypes from 1,888 individuals from the British Columbia HOMER cohort (33), the Western Australian HIV Cohort Study (WAHCS) (34), and the US AIDS Clinical Trials Group (ACTG) protocol 5142 (35) participants who also provided human DNA under ACTG protocol 5128 (36) were used to measure the strength of selection imposed by HLA-C alleles. The HIV sequences have previously been deposited in GenBank (11) and for these subjects linked HLA/HIV data sets are available for sharing with interested researchers in accordance with cohort-specific Research Ethics Board protocols (please contact the corresponding author for information). CTL responses were measured in a cohort of 1010 individuals from Durban, South Africa (14). Wellcome Trust Case-Control Consortium data of inflammatory bowel disease were obtained for subjects of European descent from the WTCCC official website (http://www.wtccc.org.uk/), including 1999 Crohn's disease cases and 3004 controls from the WTCCC1 study, and 2361 ulcerative colitis cases and 5417 controls from WTCCC2 study. Further details for these samples have been previously described (17,18). The representative institutional review boards (IRB) approved this study. All subjects gave written informed consent and/or specimens were anonymized by IRB-approved procedures.

Genotyping

HIV patients were genotyped for HLA-A/B/C by the PCR-SSOP (sequence-specific oligonucleotide probing) typing protocol and PCR-SBT (sequence based typing) as recommended by the 13th International Histocompatibility Workshop (37) (http://www.ihwg.org/tmanual/TMcontents.htm). All HLA class I loci were defined to 4-digit resolution with the exceptions of A*74:01/2, B*81, Cw*17 and Cw*18, which were determined to 2-digits. SNP rs9264942 was genotyped using an ABI TaqMan allelic discrimination assay as previously described (4). Four digit imputation of the HLA-A, B, and C class I alleles, and the HLA-DQA1, DQB1, and DRB1 class II alleles was performed for all inflammatory bowel disease patients. The method of imputation and validation of imputation accuracy was described previously (38).

HLA-C expression levels

Surface protein HLA-C expression levels were measured by flow cytometry as previously described (4). Briefly, the monoclonal antibody DT9 was shown to be HLA-C specific by screening against 100 common HLA class I alleles and used to stain CD3+ mononuclear cells isolated from freshly drawn blood by Lymphoprep. This antibody also binds HLA-E, but expression levels of this non-classical HLA molecule are very low and likely contribute negligibly to DT9 binding of CD3+ cells (4). The level of HLA-C expression for each individual in our cohorts of HIV patients was then predicted by summing the average expression observed for each of the two corresponding alleles in our staining of HLA-C levels on cells from 200 normal African Americans. The legitimacy of our prediction algorithm was verified by accurately predicting the expression level of 50 of the African Americans (the test set) based on a training set of 150 African Americans ($p<0.0001$, $r=0.63$). Importantly, the prediction based on the African American training set is equally as proficient when applied to the 50 Europeans Americans where we have measured HLA-C ($p<0.0001$, $r=0.64$) (Fig. S2). HLA-C levels in HIV patients are not known to differ from that in healthy donors. It is well established that although HIV nef downregulates HLA-A and –B, it does not downregulate HLA-C (39). Further, we do not detect any alteration in HLA-C levels on primary CD4+ cells infected with HIV in vitro (Fig. S2) in an assay where CD4+ cells isolated from PBL using anti-CD4 mAb and magnetic selection (EasySep) were expanded for 3 days with IL-2 (Peprotech) and anti-CD3/28 beads (Invitrogen) before infection with NL4-3 at MOI=0.1. Cytometry analysis was conducted 4 days later. The same results were obtained for a second viral strain JR-FL (results are not shown).

Statistical analyses

The predicted HLA-C expression level as a continuous variable and presence versus absence of all individual HLA class I alleles of frequency ≥2% were included in regression analyses with stepwise selection for all models testing effects on outcomes of HIV infection. In addition, all HLA class II alleles with frequency ≥2% were included in analyses of inflammatory disease. SAS 9.1 (SAS Institute) was used for data management and statistical analyses. PROC FREQ was used to compute frequencies on categorical variables. PROC LOGISTIC was used to calculate odds ratios and 95% confidence intervals, with an odds ratio below 1 indicating protection. PROC REG was

used for linear regression of continuous variables, taking the OR determined for an increase of a single unit in the continuous variable and converting this to a difference of "x" units in the continuous variable by following: $e^{(\ln(OR) \text{ multiplied by } x)}$. PROC PHREG and LIFETEST were used for Cox model analyses. Analyses were either performed separately for European and African Americans, or by combining the groups, in which case the analyses were adjusted for race. Results are shown for models using a threshold of a two-sided P value <0.05 for inclusion of a covariate as a significant independent effect.

The stability of regression model analyses of longitudinal VL in European and African Americans separately, progression to CD4<200 in European and African Americans combined and the case/control analysis of CD was also tested using more stringent thresholds of $p<0.01$ and $p<0.001$ for covariate inclusion in the model. This resulted in reduced sets of independently significant allelic variables, but HLA-C expression remained a significant independent effect in every case. Since HLA-C alleles can influence control of HIV independently of expression level, we tested the likelihood that the distribution of allelic expression values we observe correlate with the influence of these HLA-C alleles on HIV infection by chance. The true expression levels for each of the fourteen 2-digit HLA-C alleles were randomly re-assigned in 1000 tests. Each iteration was used to predict expression in our cohort of 2,527 European Americans categorized as controllers or non-controllers. Each of the 1000 pseudo-expression distributions was then tested as a continuous variable along with all HLA class I alleles in a regression model, as used for the analysis shown in Table 1 where the true expression values were used. This permutation test indicated no significant likelihood for an effect as significant as that observed when using the true expression values observed for each allele ($p=0.02$).

Strength of selection pressure on the virus as indicated by viral mutations was measured using a phylogenetically-corrected logistic regression model, as previously described (10,11). Briefly, for each consensus residue across the viral genome, a stepwise procedure was employed to identify HLA alleles that were associated with escape at $q<0.2$. For each HLA-C associated escape, the most likely epitope was determined using an epitope prediction algorithm and only those mutations within or flanking high confidence epitopes (posterior epitope probability > 50%) were included (12). Nearby mutations that associate with the same HLA allele were assumed to be alternative routes of escapes from the same epitope. Thus, associations of viral mutations with a single HLA allele were determined for occurrence of any escape within 10 amino acid residues (Table S3). The strength of selection for a given HLA-C allele was then calculated as the median log odds ratio for escape at independent epitopes associated with each allele. The correlation between strength of selection and HLA-C expression level persists, although more weakly, if nearby mutations are considered independently or if mutations are not filtered for occurrence in predicted epitopes.

CTL responses were detected for each individual by stimulation of PBMC with a panel of 410 overlapping 18-mer peptides (OLP) in ELISPOT assays performed as previously described (40). HLA restriction was determined for each OLP using a stepwise Fisher's exact procedure to control for linkage disequilibrium. The most significantly associated HLA allele in each iteration was determined using Fisher's exact test, then all individuals who expressed that allele were removed and

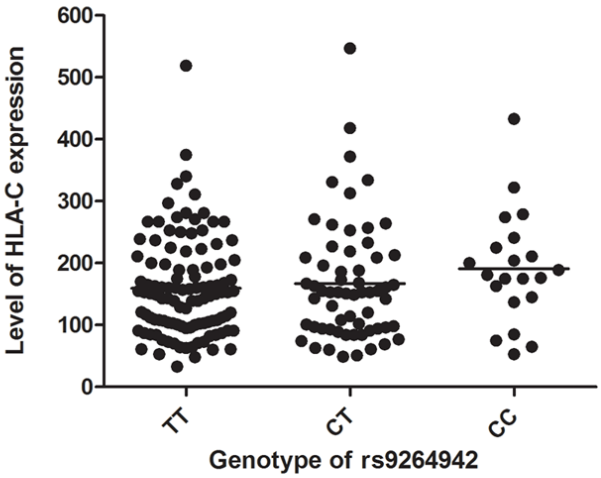
the next most significant allele (with corresponding p-value and odds ratio) was identified. All alleles associating with the OLP at $p<0.05$ ($q<0.26$) (http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/FalseDiscoveryRate/) were considered restricting alleles, although the correlation of HLA-C expression with odds of CTL response (Fig. 2B) was no weaker or less significant when using a more stringent threshold of $p<0.007$ ($q<0.05$). Five pairs of overlapping peptides associated with CTL responses restricted by the same HLA allele, which were likely to be the same CTL response detected by both peptides. Only the stronger association was included in the calculation of median lg(OR) of independent associations with each allele (Fig. 2B). Both members of these pairs were included for identification of peptides eliciting CTL responses with multiple HLA alleles (Fig. 2C).

Fig. S1

A

HLA-C allele	rs9264942 allele	Haplotype frequency	D'	p value
04:01	T	0.176	0.95	<0.0001
07:01	T	0.091	0.06	NS
16:01	T	0.086	0.85	<0.0001
06:02	C	0.085	0.94	<0.0001
17	T	0.071	1.00	<0.0001
07:02	T	0.061	0.99	<0.0001
03:04	T	0.055	0.90	<0.0001
18	T	0.044	0.93	<0.0001
08:02	C	0.033	0.83	<0.0001
05:01	C	0.028	0.97	<0.0001
02:02	C	0.025	0.31	<0.0001
02:10	T	0.021	0.26	NS
03:02	C	0.021	0.92	<0.0001
12:03	C	0.019	0.95	<0.0001
14:02	C	0.019	0.95	<0.0001
15:05	T	0.014	0.54	NS
08:04	C	0.013	0.97	<0.0001
03:03	T	0.010	0.97	0.03

B

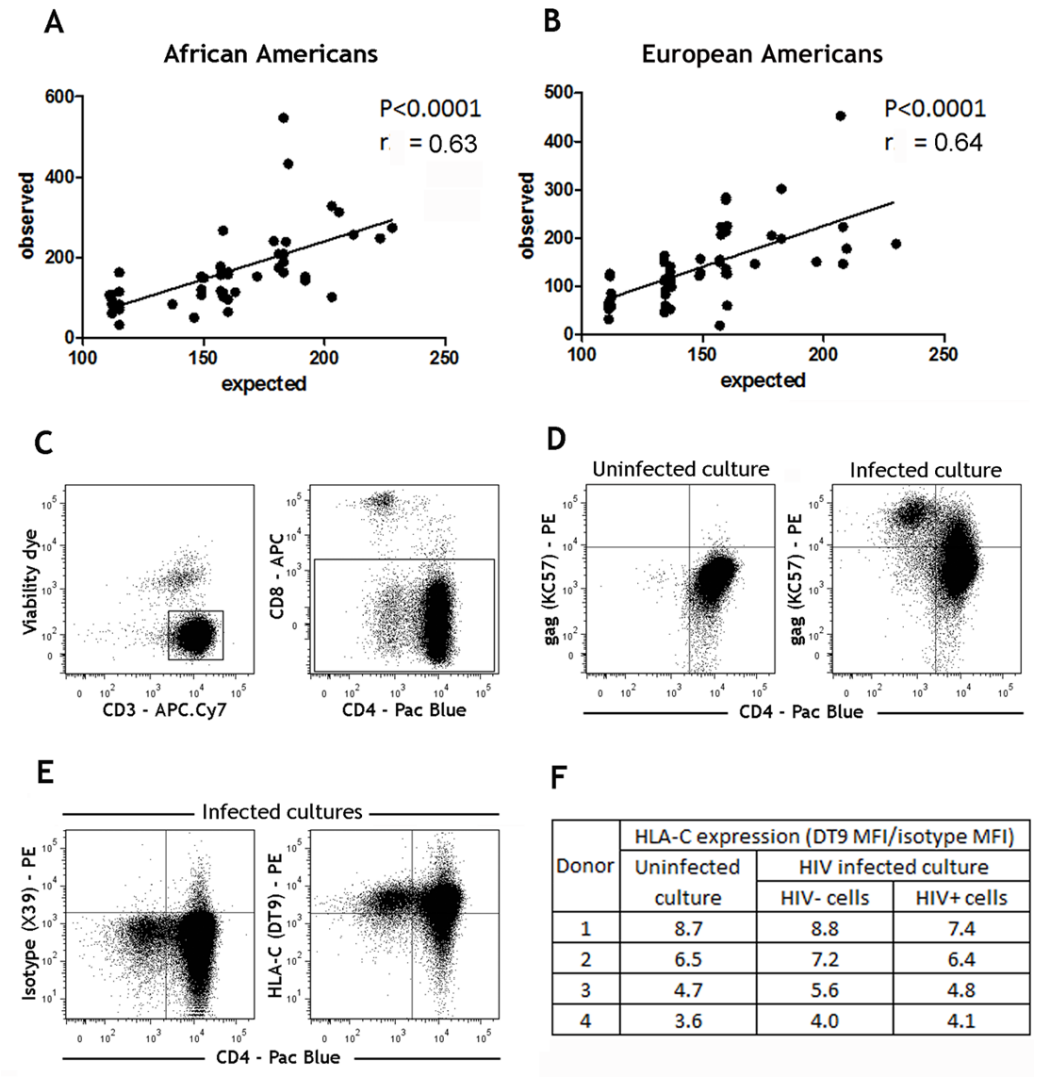


C

Covariate	p	OR	95%CI	
B*57:03	4E-22	0.11	0.07	0.17
B*81	7E-10	0.13	0.07	0.25
B*57:01	5E-04	0.15	0.05	0.44
A*36:01	6E-04	5.10	2.01	12.96
B*58:01	1E-03	0.40	0.23	0.71
B*45:01	2E-03	3.21	1.51	6.83
B*35:01	4E-03	2.19	1.29	3.75
B*57:02	6E-03	0.23	0.08	0.66
A*23:01	6E-03	1.78	1.18	2.68
B*39:10	6E-03	0.25	0.09	0.67
B*14:02	8E-03	0.38	0.18	0.77
C*05:01	8E-03	0.39	0.20	0.79
A*01:01	1E-02	2.04	1.15	3.62
B*27:05	2E-02	0.29	0.11	0.79
C*02:10	2E-02	0.49	0.27	0.89
B*52:01	2E-02	0.38	0.17	0.86
B*44:03	3E-02	0.57	0.35	0.95

The SNP rs9264942 cannot be used to test an effect of expression level in African Americans. **(A)** LD between rs9264942 and HLA-C was determined in 1,931 African Americans. Common HLA-C alleles, such as Cw*07:01, show no significant LD with rs9264942. Among HLA-C alleles that do show significant LD, the strength is substantially weaker than that observed in European Americans, where D' = 1 for virtually all HLA-C alleles (4). **(B)** HLA-C surface expression was measured on peripheral blood CD3+ cells by flow cytometry in 192 normal individuals in whom rs9264942 genotype was also determined. Level of HLA-C does not correlate with rs9264942 genotype, unlike that observed in European Americans (4). **(C)** The rs9264942 genotype has no effect on outcome of HIV infection. African Americans (n = 920) were categorized as controllers or non-controllers (8), and all HLA class I alleles with ≥2% phenotypic frequency as well as rs9264942 genotype (as a continuous variable: CC vs CT vs TT) were used as covariates in a logistic regression model with stepwise selection. Independently significant covariates are reported with significance (p), odds ratio (OR) and 95% confidence inteverval (CI).

Fig. S2



F

Donor	HLA-C expression (DT9 MFI/isotype MFI)		
	Uninfected culture	HIV infected culture	
		HIV- cells	HIV+ cells
1	8.7	8.8	7.4
2	6.5	7.2	6.4
3	4.7	5.6	4.8
4	3.6	4.0	4.1

Use of observed allelic expression to predict HLA-C levels in HIV patients. Mean expression levels for each 2-digit HLA-C allotype, as determined by DT9 staining in a training set of 150 African Americans, were used to generate an expected expression value (i.e. the mean of the two HLA-C allotype expression values corresponding to the HLA-C genotype) for each subject of a test set (8). Expected values correlated significantly with the actual observed measurements of expression (by DT9 staining) in test sets of 50 African Americans **(A)** and 50 European Americans **(B)**. HLA-C expression levels are not altered after HIV infection (8). CD4+ cells were purified from peripheral blood, infected with HIV *in vitro* and analyzed by flow cytometry after 4 days. After gating on CD3+ CD8- cells **(C)**, CD4 staining can be used to discriminate infected cells **(D)**. The majority of gag+ infected cells have downregulated CD4, whereas most CD4 positive cells remain uninfected. HLA-C staining between CD4 positive (uninfected) as compared to CD4 negative (infected) cells shows no significant difference **(E)**. This finding was replicated in 4 independent donors **(F)**.

Table S1

European Americans (n=3280)				African Americans (n=1459)			
Covariate	**p**	**Δ**	**SE**	**Covariate**	**p**	**Δ**	**SE**
B*57:01	3E-26	-0.92	0.09	B*57:03	3E-32	-0.74	0.06
A*01:01	9E-13	0.38	0.05	B*58:02	3E-07	0.46	0.09
C*04:01	2E-08	0.29	0.05	B*45:01	1E-06	0.40	0.08
B*27:05	1E-06	-0.34	0.07	**C expression**	**2E-05**	**-0.15**	**0.03**
C*07:01	4E-06	-0.37	0.08	B*35:01	2E-05	0.28	0.07
C expression	**4E-06**	**-0.20**	**0.04**	A*23:01	5E-05	0.22	0.05
B*07:02	1E-05	0.26	0.06	B*81	2E-04	-0.39	0.11
B*13:02	2E-05	-0.46	0.11	B*14:02	1E-03	-0.33	0.10
B*52:01	3E-05	-0.47	0.11	A*36:01	1E-03	0.35	0.11
B*49:01	7E-05	0.49	0.12	B*15:10	2E-03	0.27	0.09
A*02:01	8E-05	0.17	0.04	B*39:10	2E-03	-0.42	0.14
B*08:01	2E-04	0.32	0.09	B*53:01	1E-02	0.14	0.06
B*18:01	3E-04	0.28	0.08	B*58:01	2E-02	-0.20	0.08
A*25:01	3E-04	-0.32	0.09	A*74	2E-02	-0.15	0.06
B*14:02	7E-04	-0.27	0.08	C*08:04	2E-02	-0.32	0.14
A*26:01	4E-03	0.24	0.08	A*03:01	2E-02	-0.13	0.06
A*24:02	4E-03	0.16	0.05				
B*15:01	6E-03	-0.18	0.06				
C*06:02	6E-03	0.22	0.08				
A*31:01	6E-03	-0.21	0.08				
B*40:01	1E-02	0.19	0.08				
B*55:01	2E-02	0.29	0.13				
A*68:01	3E-02	0.18	0.08				
A*29:02	4E-02	0.16	0.08				

Effect of HLA-C expression level on longitudinal mean viral load as a continuous variable. HLA-C expression as a continuous variable and all HLA class I alleles with frequency ≥ 2% were tested with stepwise selection in a linear regression model for an effect on mean VL. Independently significant covariates are reported with significance (p), change in log10 viral load (Δ) and standard error (SE). The change in log10 VL shown for HLA-C expression represents an increase of 100 MFI expression units.

Table S2

Covariate	p	HR	95% CI	
C expression	**7E-06**	**0.61**	**0.50**	**0.74**
B*57:01	9E-05	0.31	0.17	0.56
A*26:01	3E-04	0.32	0.17	0.59
A*11:01	2E-03	0.59	0.43	0.82
A*32:01	4E-03	0.52	0.33	0.82
B*15:01	5E-03	0.55	0.36	0.83
B*35:03	5E-03	2.01	1.23	3.30
C*06:02	1E-02	1.61	1.12	2.31
C*04:01	3E-02	1.38	1.03	1.85
C*03:03	5E-02	1.51	1.01	2.26

HLA-C expression level affects time for progression to earlier AIDS outcomes in European Americans. HLA-C expression as a continuous variable and all HLA class I alleles with phenotypic frequencies ≥ 2% were tested in a cox model with stepwise selection for an effect on progression to CD4<200 in an analysis restricted to 783 ART-naïve European Americans. Independently significant covariates are reported with significance (p), hazard ratio (HR) and 95% confidence interval (CI). The change in HR shown for HLA-C expression represents an increase of 100 MFI expression units. Thus, the effect of HLA-C expression level is significant in an analysis restricted to 783 European Americans. Inclusion of the 286 African American samples resulted in greater significance of the effect for HLA-C expression level ($p=1x10^{-6}$, Table 2) with negligible change in the strength of the effect, consistent with a similar strength of the effect across ethnicities.

Table S3

HLA-C allele	Proteome location (protein, residue)	Consensus residue	Mutation ln(OR)	p value	Mutation at any site within 10 amino acids ln(OR)	Median association for each 2-digit HLA-C allele ln(OR)
C*04:01	Pol , 821	G	15.8	1.5E-07	15.8	10.9
C*04:04	Nef , 125	Q	6.1	1.0E-04	6.1	
C*06:02	Nef , 179	E	16.4	5.2E-04	16.4	14.1
C*06:02	Pol , 709	A	11.8	1.0E-05	11.8	
C*07	Nef , 114	V	2.3	6.4E-06	4.9	4.9
C*07:01	Nef , 105	K	3.7	6.0E-13		
C*07:01	Nef , 108	D	1.5	3.9E-06		
C*07:01	Nef , 115	Y	3.1	3.3E-10		
C*07:02	Nef , 105	K	1.7	1.5E-08		
C*08:01	gp41 , 54	L	5.5	1.1E-04	5.5	5.5
C*08:02	Nef , 82	K	3.7	3.1E-20	2.7	
C*08:02	Nef , 85	L	4.1	4.2E-11		
C*08:02	Nef , 86	D	9.4	2.3E-11		
C*08:02	Nef , 89	H	12.5	4.7E-04		
C*08:02	Nef , 203	Y	10.5	2.0E-11	10.5	
C*14	Nef , 125	Q	11.3	9.1E-07	27.6	27.6
C*14:02	Nef , 120	Y	12.4	1.2E-10		
C*14:02	Nef , 126	N	5.5	1.5E-04		
C*15	Pol , 709	A	11.2	5.4E-08	7.6	7.6
C*15:05	Pol , 711	I	12.2	2.4E-59		
C*18:01	Nef , 76	L	11.9	3.4E-11	11.9	17.8
C*18:01	gp120 , 252	R	23.7	4.4E-04	23.7	

List of the 22 viral mutations associating with HLA-C alleles and the associations for escape at the 12 independent epitopes. Mutation ln(OR) is the log OR for association of the viral mutation and HLA-C allele shown, after adjusting for both HLA-A/B alleles and the transmitted viral sequence (8). All p-values are significant at $q<0.2$ after correction for multiple testing. Multiple nearby escapes associate with the same HLA-C allele for several alleles (C*07, 08, 14, 15). These are likely to be alternative routes of escape from the same immune response, so the association between each of these HLA alleles and any of the corresponding mutations was also determined. Median log OR are then calculated for mutation at all independent epitopes associated with the same HLA-C allele.

Table S4

HLA-C allele	HIV peptide	Proteome location (protein, residue)	Patients with associated HLA allele		Patients without associated HLA allele		log10 (OR)	p value
			No CTL response	CTL response	No CTL response	CTL response		
C01	ELAENREILKEPVHGVYY	Pol , 457	3	1	768	6	1.74	3.6E-02
C02	PRTLNAWVKVIEEKAF	Gag , 149	129	73	764	20	1.32	1.4E-37
C02	RQANFLGKIWPSHKGR	Gag , 429	160	4	557	2	0.76	2.6E-02
C02	GKKAIGTVLVGPTPVNII	Pol , 124	177	25	775	9	1.05	1.6E-11
C02	ELAENREILKEPVHGVYY	Pol , 457	185	17	775	9	0.88	5.9E-07
C02	VKAACWWAGIQQEFGIPY	Pol , 841	155	47	769	15	1.17	2.6E-22
C02	AVFIHNFKRKGGIGGYSA	Pol , 894	94	108	755	29	1.46	9.3E-60
C02	KVVPRRKAKIIKDYGKQM	Pol , 973	155	25	604	10	0.96	4.1E-10
C02	YHCLVCFQTKGLGISYGR	Tat , 32	184	18	779	5	1.13	2.7E-09
C03	GKKHYMLKHLVWASREL	Gag , 25	83	23	652	48	0.58	7.7E-06
C03	GPKEPFRDYVDRFFKTLR	Gag , 288	87	48	713	22	1.24	2.9E-26
C03	TILRALGPGASLEEMMTA	Gag , 332	121	6	704	10	0.57	2.3E-02
C03	IVQQQSNLLRAIEAQQHM	gp160 , 548	122	7	722	12	0.56	1.5E-02
C03	EVGFPVRPQVPLRPMTFK	Nef , 65	63	14	330	29	0.41	1.1E-02
C03	ARRHMARELHPEFYKDC	Nef , 190	136	6	833	12	0.52	3.4E-02
C04	KVSFDPIPIHYCAPAGYA	gp160 , 207	175	22	628	4	1.22	4.3E-11
C04	RVSKKLAEHFPNKTIKF	gp160 , 344	223	4	758	2	0.75	2.8E-02
C04	IWDNMTWMQWDREISNY	gp160 , 622	215	12	759	1	1.36	1.8E-07
C04	EVGFPVRPQVPLRPMTFK	Nef , 65	20	4	138	6	0.67	3.8E-02
C04	FKGAFDLSFFLKEKGGL	Nef , 81	55	22	211	36	0.37	9.9E-03
C04	KKKSVTVLDVGDAYFSV	Pol , 257	196	31	757	3	1.49	9.5E-18
C05	NANPDCKTILRALGPGA	Gag , 325	10	2	923	9	1.40	7.7E-03
C05	SFFLKEKGGLEGLIYSKK	Nef , 88	9	3	964	11	1.51	4.6E-04
C05	KKRQEILDLWVYHTQGYF	Nef , 104	8	2	538	11	1.18	2.0E-02
C06	IRSENLTNNAKTIIVHL	gp160 , 272	234	31	718	3	1.39	1.4E-15
C06	AASITLTVQARQLLSGIV	gp160 , 532	256	9	720	1	1.15	5.1E-05
C06	GIKQLQTRVLAIERYLK	gp160 , 572	218	47	720	1	1.90	2.2E-27
C06	WLRAQEEEEEVGFPVRPQV	Nef , 57	225	5	676	3	0.65	2.8E-02
C06	FFREDLAFPQGEAREF	Pol , 1	252	7	669	3	0.72	6.5E-03
C06	LKTGKYAKMRTAHTNDVK	Pol , 504	255	10	717	4	0.79	5.9E-04
C06	AVFIHNFKRKGGIGGYSA	Pol , 894	188	6	381	2	0.67	2.0E-02
C07	TILRALGPGASLEEMMTA	Gag , 332	282	8	422	2	0.65	1.8E-02
C07	VMGNLWVTVYYGVPVWK	gp160 , 30	325	6	578	2	0.62	3.0E-02
C07	KKRQEILDLWVYHTQGYF	Nef , 104	221	130	607	29	1.08	2.9E-39
C08	PMFTALSEGATPQDLNTM	Gag , 170	95	5	626	8	0.64	2.2E-02
C08	NANPDCKTILRALGPGA	Gag , 325	116	4	807	5	0.76	2.0E-02
C08	SHKARVLAEAMSQANSA	Gag , 357	120	4	859	4	0.85	1.1E-02
C08	RAVGIGAVFLGFLGAA	gp160 , 511	119	5	861	2	1.16	4.9E-04
C08	IVQQQSNLLRAIEAQQHM	gp160 , 548	114	10	844	19	0.61	1.7E-03
C08	DIIATDIQTKELQKQIIK	Pol , 917	115	9	861	2	1.39	2.6E-07

C12	NANPDCKTILRALGPGA	Gag , 325	28	15	933	11	1.63	1.5E-15
C12	IHYCAPAGYAILKCNNKTF	gp160 , 215	40	3	939	5	1.18	3.7E-03
C12	EVGFPVRPQVPLRPMTFK	Nef , 65	19	11	393	43	0.73	1.8E-04
C12	QITLWQRPLVSIKVGGQI	Pol , 58	40	3	884	5	1.16	4.4E-03
C12	QGWKGSPAIFQSSMTKIL	Pol , 306	31	11	799	35	0.92	3.2E-06
C12	RAQNPEIVIYQYMDDLYV	Pol , 327	27	7	639	17	1.01	7.1E-05
C12	QKIAMESIVIWGKTPKFR	Pol , 528	24	9	597	6	1.53	3.5E-09
C12	QAVRIIKILYQSNPY	Rev , 14	31	4	719	14	0.88	7.0E-03
C12	HIPLGEARLVIKTYWGL	Vif , 56	40	3	940	4	1.26	2.4E-03
C14	TGTEELRSLYNTVATLY	Gag , 70	2	2	504	20	1.38	9.5E-03
C14	AFSPEVIPMFTALSEGA	Gag , 163	5	5	931	20	1.65	1.8E-06
C14	KKRQEILDLWVYHTQGYF	Nef , 104	2	4	605	25	1.59	5.0E-05
C14	NYTPGPGVRYPLTFGWCF	Nef , 126	4	3	435	8	1.59	3.6E-04
C14	QITLWQRPLVSIKVGGQI	Pol , 58	9	2	924	8	1.49	5.3E-03
C14	LVGPTPVNIIGRNMLTQL	Pol , 132	4	1	666	5	1.65	4.4E-02
C15	MASEFNLPPIVAKEIVA	Pol , 737	8	5	761	21	1.36	2.7E-05
C16	WVKVIEEKAFSPEVIPMF	Gag , 155	94	10	793	15	0.76	1.8E-04
C16	ELKKSAVSLLDTIAIAVA	gp160 , 806	104	5	873	5	0.92	2.4E-03
C16	QGYFPDWQNYTPGPGVRY	Nef , 118	63	4	501	5	0.82	1.4E-02
C16	ETKIGKAGYVTDRGRQKI	Pol , 604	100	4	805	3	1.00	4.3E-03
C16	QAVRIIKILYQSNPY	Rev , 14	57	5	662	9	0.84	4.1E-03
C16	ILSTCLGRPAEPVPLQL	Rev , 59	102	7	865	13	0.68	3.8E-03
C17	EKIRLRPGGKKHYMLKHL	Gag , 17	64	7	360	3	1.05	1.9E-04
C17	WIILGLNKIVRMYSPVSI	Gag , 265	182	23	639	26	0.49	2.1E-04
C17	TALIKPKKIKPPLPSVRK	Vif , 151	205	15	760	7	0.87	4.2E-06
C18	WIILGLNKIVRMYSPVSI	Gag , 265	93	24	821	49	0.64	5.7E-07
C18	GPKEPFRDYVDRFFKTLR	Gag , 288	51	66	800	70	1.16	5.3E-33
C18	TVYYGVPVWKEAKTTLF	gp160 , 37	86	8	753	20	0.57	7.0E-03
C18	IRIGPGQTFYATGDII	gp160 , 307	111	6	870	0	1.74	2.5E-06
C18	MLDLLAKVDYRLGVGALI	Pol , 596	107	10	867	3	1.34	8.1E-08
C18	ELKKIIGQVRDQAEHLK	Pol , 872	87	17	660	4	1.43	2.0E-12

Supplementary Table 4: List of the 71 HIV peptide and HLA-C allele associations with occurrence of a CTL response ($p<0.05$). Numbers of people in columns 4-7 are raw counts. Log10(OR) is a smoothed OR for association of CTL response to a peptide with an allele, after adjusting for linked HLA using a stepwise Fisher's exact procedure (8). All p-values shown are significant at $q<0.26$ after correction for multiple testing. Use of a more stringent threshold for identification of HLA-associated CTL responses of $p<0.007$ ($q<0.05$) resulted in an equally strong and significant correlation of HLA-C expression with odds of CTL response.

Table S5

HIV peptide	Proteome location (protein, residue)	Restricting alleles compared		Difference in log10(OR)	Difference in expression
		Allele 1	Allele 2		
WIILGLNKIVRMYSPVSI	Gag , 265	C*17	C*18	-0.15	-124
GPKEPFRDYVDRFFKTLR	Gag , 288	C*03	C*18	0.08	-123
NANPDCKTILRALGPGA	Gag , 325	C*05	C*08	0.64	-22
		C*05	C*12	-0.23	-39
		C*08	C*12	-0.87	-17
TILRALGPGASLEEMMTA	Gag , 332	C*03	C*07	-0.09	2
		C*03	C*08	-0.08	-60
		C*07	C*08	0.00	-62
EVGFPVRPQVPLRPMTFK	Nef , 65	C*03	C*04	-0.26	-85
		C*03	C*12	-0.32	-77
		C*04	C*12	-0.06	8
KKRQEILDLWVYHTQGYF	Nef , 104	C*05	C*07	0.10	40
		C*05	C*14	-0.41	-140
		C*07	C*14	-0.51	-180
QGYFPDWQNYTPGPGVRY	Nef , 118	C*14	C*16	0.30	113
QAVRIIKILYQSNPY	Rev , 14	C*12	C*16	0.04	12
QITLWQRPLVSIKVGGQI	Pol , 58	C*12	C*14	-0.33	-101
ELAENREILKEPVHGVYY	Pol , 457	C*01	C*02	0.86	87
AVFIHNFKRKGGIGGYSA	Pol , 894	C*02	C*06	0.79	-59
IVQQQSNLLRAIEAQQHM	gp160 , 548	C*03	C*08	-0.05	-60

List of the 20 comparisons involving a single HIV peptide associating with occurrence of a CTL response involving distinct HLA-C alleles. Difference in log10(OR) compares the difference in association with CTL response between the two alleles shown. The difference in expression is then also compared between these two alleles.

Table S6

Crohn's Disease				
Covariate	**p**	**OR**	**95%CI**	
DRB1*01:03	5E-13	2.70	2.06	3.54
C expression	**3E-07**	**1.35**	**1.22**	**1.49**
DRB1*15:01	1E-05	0.71	0.60	0.83
C*01:02	4E-04	0.63	0.49	0.81
A*03:01	9E-04	1.27	1.10	1.45
DQA1*03:01	1E-03	0.80	0.69	0.92
A*01:01	3E-03	1.24	1.08	1.43
C*05:01	4E-03	1.24	1.07	1.44
B*40:01	5E-03	0.74	0.59	0.91
B*37:01	8E-03	0.60	0.41	0.87
DRB1*08:01	1E-02	5.11	1.41	18.54
DQA1*01:01	1E-02	0.82	0.70	0.96
DQA1*04:01	4E-02	0.26	0.07	0.93
B*35:01	4E-02	0.79	0.63	0.99
DRB1*03:01	4E-02	0.84	0.72	0.99

Ulcerative Colitis				
Covariate	**p**	**OR**	**95%CI**	
DQA1*02:01	2E-15	0.59	0.52	0.67
DQB1*02:01	8E-13	0.53	0.44	0.63
DRB1*01:03	1E-12	2.03	1.67	2.47
DRB1*13:02	2E-07	0.58	0.47	0.71
A*02:01	2E-04	1.21	1.1	1.35
DQA1*03:01	7E-04	0.73	0.61	0.88
B*08:01	9E-04	1.31	1.12	1.54
DQB1*06:03	2E-03	0.23	0.09	0.57
DRB1*14:01	2E-03	0.68	0.53	0.87
DRB1*04:04	4E-03	0.71	0.56	0.9
DQA1*01:03	5E-03	3.78	1.51	9.45
DRB1*11:01	5E-03	0.73	0.59	0.91
DRB1*01:01	6E-03	0.81	0.7	0.94
C*07:02	1E-02	0.86	0.77	0.97
A*32:01	2E-02	1.24	1.04	1.48
DQA1*04:01	2E-02	0.75	0.58	0.96
DRB1*04:01	3E-02	0.8	0.65	0.98
B*49:01	3E-02	1.38	1.02	1.86

HLA-C expression level affects Crohn's disease but not ulcerative colitis. HLA-C expression level as a continuous variable and all HLA class I and class II alleles with phenotypic frequency ≥ 2% were tested by stepwise selection in a logistic regression model comparing 1999 Crohn's disease patients to 3004 controls, and comparing 2361 ulcerative colitis patients to 5175 controls, all of whom were of European descent. Independently significant covariates are reported with significance (p), odds ratio (OR) and 95% confidence interval (CI) where the OR of HLA-C expression represents an increase of 100 MFI expression units.

Table S7

A

Covariate	p	OR	95%CI	
B*57:01	2E-19	0.23	0.17	0.32
C*04:01	9E-10	2.31	1.77	3.02
A*01:01	2E-09	2.18	1.69	2.80
B*07:02	9E-08	2.18	1.64	2.89
A*02:01	2E-07	1.74	1.41	2.14
C expression	**6E-07**	**0.64**	**0.54**	**0.77**
B*27:05	1E-05	0.48	0.35	0.67
C*12:02	1E-04	0.36	0.21	0.61
A*24:02	4E-04	1.65	1.25	2.19
A*26:01	9E-04	2.00	1.33	3.00
B*40:01	3E-03	1.80	1.22	2.66
A*29:02	5E-03	1.87	1.21	2.89
B*58:01	8E-03	0.43	0.23	0.81
B*14:02	9E-03	0.62	0.43	0.89
B*50:01	2E-02	2.44	1.17	5.11

B

Covariate	p	OR	95%CI	
B*57:03	2E-20	0.14	0.09	0.21
B*35:01	2E-05	3.14	1.86	5.29
C expression	**4E-05**	**0.61**	**0.48**	**0.77**
B*45:01	1E-03	3.25	1.61	6.56
A*36:01	1E-03	4.29	1.78	10.34
A*23:01	1E-03	1.89	1.28	2.79
B*27:05	4E-03	0.15	0.04	0.54
B*81	4E-03	0.36	0.18	0.72
B*58:02	1E-02	2.43	1.24	4.78
C*01:02	1E-02	4.92	1.41	17.13
B*39:10	2E-02	0.33	0.13	0.81
B*58:01	2E-02	0.54	0.32	0.90
C*08:04	2E-02	0.32	0.12	0.86

C

Covariate	p	Δ	SE
B*57:01	2E-31	-0.84	0.07
B*27:05	1E-07	-0.40	0.08
A*01:01	1E-07	0.27	0.05
C*07:02	2E-06	0.28	0.06
B*52:01	5E-06	-0.56	0.12
A*31:01	4E-05	-0.35	0.09
B*14:02	8E-05	-0.33	0.08
B*58:01	1E-04	-0.56	0.14
C*04:01	1E-04	0.22	0.06
A*25:01	2E-04	-0.34	0.09
B*40:01	6E-04	0.27	0.08
B*15:01	3E-03	-0.20	0.07
C expression	**4E-03**	**-0.13**	**0.04**
A*03:01	8E-03	-0.14	0.05
A*11:01	8E-03	-0.17	0.06
B*13:02	1E-02	-0.27	0.11
A*30:01	2E-02	-0.33	0.14
A*32:01	2E-02	-0.18	0.08
C*14:02	4E-02	-0.29	0.14

D

Covariate	p	Δ	SE
B*57:03	8E-27	-0.72	0.07
A*23:01	1E-05	0.27	0.06
B*58:02	2E-05	0.44	0.10
A*36:01	1E-04	0.47	0.12
B*35:01	2E-04	0.28	0.08
B*81	6E-04	-0.40	0.12
C*12:03	7E-04	-0.36	0.11
B*27:05	2E-03	-0.56	0.18
B*45:01	2E-03	0.30	0.10
C expression	**6E-03**	**-0.11**	**0.04**
B*57:01	1E-02	-0.41	0.16
B*15:10	1E-02	0.26	0.10
C*01:02	2E-02	0.44	0.18
C*08:04	2E-02	-0.37	0.15
B*14:02	2E-02	-0.25	0.11
A*01:01	2E-02	0.20	0.09
A*24:02	2E-02	0.26	0.11
A*32:01	5E-02	0.25	0.13

Significant effects of HLA-C expression level in the subset of subjects with longitudinal viral load measurements that are not included in progression analyses. HLA-C expression as a continuous variable and all HLA class I alleles with phenotypic frequency ≥2% were tested using stepwise selection in regression models for an effect on the categorical distinction of mean VL <2,000 or >10,000 **(A,B)** or continuous outcome of mean VL **(C,D)** using longitudinal viral load measurements prior to ART. Analyses were performed separately for European American **(A,C)** and African American **(B,D)** individuals. HLA-C expression remains a significant independent effect in the analyses of categorical outcome of HIV infection for 2173 European **(A)** and 991 African Americans **(B)** and in the analyses of continuous mean VL in 2866 European **(C)** and 1201 African Americans **(D)**. Independently significant covariates are reported with significance (p), change in log10 viral load (Δ), odds ratio (OR), standard error (SE) and 95% confidence inteverval (CI). Δ and OR shown for HLA-C expression represent an increase of 100 MFI expression units.

References

1. A. A. Bashirova, R. Thomas, M. Carrington, HLA/KIR restraint of HIV: Surviving the fittest. *Annu. Rev. Immunol.* **29**, 295 (2011). doi:10.1146/annurev-immunol-031210-101332 Medline

2. International HIV Controllers Study, The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330**, 1551 (2010). doi:10.1126/science.1195271 Medline

3. J. Fellay *et al*., A whole-genome association study of major determinants for host control of HIV-1. *Science* **317**, 944 (2007). doi:10.1126/science.1143767 Medline

4. R. Thomas *et al*., HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nat. Genet.* **41**, 1290 (2009). doi:10.1038/ng.486 Medline

5. T. W. Corrah *et al*., Reappraisal of the relationship between the HIV-1-protective single-nucleotide polymorphism 35 kilobases upstream of the HLA-C gene and surface HLA-C expression. *J. Virol.* **85**, 3367 (2011). doi:10.1128/JVI.02276-10 Medline

6. F. F. Gonzalez-Galarza, S. Christmas, D. Middleton, A. R. Jones, Allele frequency net: A database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res.* **39**, (Database), D913 (2011). doi:10.1093/nar/gkq1128 Medline

7. K. Pelak *et al*., Host determinants of HIV-1 control in African Americans. *J. Infect. Dis.* **201**, 1141 (2010). doi:10.1086/651382 Medline

8. Materials and methods are available as supplementary materials on *Science* Online

9. J. K. Wright *et al*., Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: Associations with HLA type and clinical parameters. *J. Virol.* **84**, 10820 (2010). doi:10.1128/JVI.01084-10 Medline

10. J. M. Carlson *et al*., Phylogenetic dependency networks: Inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLOS Comput. Biol.* **4**, e1000225 (2008). doi:10.1371/journal.pcbi.1000225 Medline

11. J. M. Carlson *et al*., Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. *J. Virol.* **86**, 5230 (2012). doi:10.1128/JVI.06728-11 Medline

12. D. Heckerman, C. Kadie, J. Listgarten, Leveraging information across HLA alleles/supertypes improves epitope prediction. *J. Comput. Biol.* **14**, 736 (2007). doi:10.1089/cmb.2007.R013 Medline

13. M. E. Blais *et al*., High frequency of HIV mutations associated with HLA-C suggests enhanced HLA-C-restricted CTL selective pressure associated with an AIDS-protective polymorphism. *J. Immunol.* **188**, 4663 (2012). doi:10.4049/jimmunol.1103472 Medline

14. P. Kiepiela *et al*., Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* **432**, 769 (2004). doi:10.1038/nature03113 Medline

15. P. Kiepiela *et al*., $CD8^+$ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat. Med.* **13**, 46 (2007). doi:10.1038/nm1520 Medline

16. P. C. Stokkers, P. H. Reitsma, G. N. Tytgat, S. J. van Deventer, HLA-DR and -DQ phenotypes in inflammatory bowel disease: A meta-analysis. *Gut* **45**, 395 (1999). doi:10.1136/gut.45.3.395 Medline

17. Wellcome Trust Case Control Consortium, Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661 (2007). doi:10.1038/nature05911 Medline

18. C. A. Anderson *et al.*, Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246 (2011). doi:10.1038/ng.764 Medline

19. B. Newman *et al.*, CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn's disease. *Am. J. Gastroenterol.* **99**, 306 (2004). doi:10.1111/j.1572-0241.2004.04038.x Medline

20. G. Bouma, W. Strober, The immunological and genetic basis of inflammatory bowel disease. *Nat. Rev. Immunol.* **3**, 521 (2003). doi:10.1038/nri1132 Medline

21. H. P. Koeffler, J. Ranyard, L. Yelton, R. Billing, R. Bohman, Gamma-interferon induces expression of the HLA-D antigens on normal and leukemic human myeloid cells. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4080 (1984). doi:10.1073/pnas.81.13.4080 Medline

22. M. A. Purbhoo, D. J. Irvine, J. B. Huppa, M. M. Davis, T cell killing does not require the formation of a stable mature immunological synapse. *Nat. Immunol.* **5**, 524 (2004). doi:10.1038/ni1058 Medline

23. M. Faroudi *et al.*, Lytic versus stimulatory synapse in cytotoxic T lymphocyte/target cell interaction: manifestation of a dual activation threshold. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14145 (2003). doi:10.1073/pnas.2334336100 Medline

24. S. Kageyama, T. J. Tsomides, Y. Sykulev, H. N. Eisen, Variations in the number of peptide-MHC class I complexes required to activate cytotoxic T cell responses. *J. Immunol.* **154**, 567 (1995). Medline

25. E. A. Reits *et al.*, Radiation modulates the peptide repertoire, enhances MHC class I expression, and induces successful antitumor immunotherapy. *J. Exp. Med.* **203**, 1259 (2006). doi:10.1084/jem.20052494 Medline

26. C. O'Huigin *et al.*, The molecular origin and consequences of escape from miRNA regulation by HLA-C alleles. *Am. J. Hum. Genet.* **89**, 424 (2011). doi:10.1016/j.ajhg.2011.07.024 Medline

27. J. Phair *et al.*, Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: The Multicenter AIDS Cohort Study. *J. Acquir. Immune Defic. Syndr.* **5**, 490 (1992). doi:10.1097/00126334-199205000-00010 Medline

28. J. J. Goedert *et al.*, A prospective study of human immunodeficiency virus type 1 infection and the development of AIDS in subjects with hemophilia. *N. Engl. J. Med.* **321**, 1141 (1989). doi:10.1056/NEJM198910263211701 Medline

29. S. P. Buchbinder, M. H. Katz, N. A. Hessol, P. M. O'Malley, S. D. Holmberg, Long-term HIV-1 infection without immunologic progression. *AIDS* **8**, 1123 (1994). doi:10.1097/00002030-199408000-00014 Medline

30. D. Vlahov *et al*., Prognostic indicators for AIDS and infectious disease death in HIV-infected injection drug users: plasma viral load and $CD4^+$ cell count. *JAMA* **279**, 35 (1998). doi:10.1001/jama.279.1.35 Medline

31. B. Emu *et al*., Phenotypic, functional, and kinetic parameters associated with apparent T-cell control of human immunodeficiency virus replication in individuals with and without antiretroviral treatment. *J. Virol.* **79**, 14169 (2005). doi:10.1128/JVI.79.22.14169-14178.2005 Medline

32. J. J. Goedert *et al*., Decreased helper T lymphocytes in homosexual men. I. Sexual contact in high-incidence areas for the acquired immunodeficiency syndrome. *Am. J. Epidemiol.* **121**, 629 (1985). doi:10.1093/aje/121.5.629 Medline

33. Z. L. Brumme *et al*., Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS Pathog.* **3**, e94 (2007). doi:10.1371/journal.ppat.0030094 Medline

34. S. A. Mallal, The Western Australian HIV Cohort Study, Perth, Australia. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **17**, (Suppl 1), S23 (1998). doi:10.1097/00042560-199801001-00008 Medline

35. M. John *et al*., Adaptive interactions between HLA and HIV-1: Highly divergent selection imposed by HLA class I molecules with common supertype motifs. *J. Immunol.* **184**, 4368 (2010). doi:10.4049/jimmunol.0903745 Medline

36. D. W. Haas *et al*., Adult AIDS Clinical Trials Group, A multi-investigator/institutional DNA bank for AIDS-related human genetic studies: AACTG Protocol A5128. *HIV Clin. Trials* **4**, 287 (2003). doi:10.1310/MUQC-QXBC-8118-BPM5 Medline

37. M. P. Martin *et al*., Innate partnership of HLA-B and KIR3DL1 subtypes against HIV-1. *Nat. Genet.* **39**, 733 (2007). doi:10.1038/ng2035 Medline

38. H. Chen *et al*., Psoriasis patients are enriched for genetic variants that protect against HIV-1 disease. *PLoS Genet.* **8**, e1002514 (2012). doi:10.1371/journal.pgen.1002514 Medline

39. G. B. Cohen *et al*., The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity* **10**, 661 (1999). doi:10.1016/S1074-7613(00)80065-5 Medline

40. M. M. Addo *et al*., Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load. *J. Virol.* **77**, 2081 (2003). doi:10.1128/JVI.77.3.2081-2092.2003 Medline