Table 2. Cont.

						q-value of HLA-related association in each strata			Phylogentically corrected Log2 Odds Ratio (pLOR) towards "escape"		
HIV-1 protein	HLA Class I	Population majority consensus amino acid at associated site	HXB2 site and direction of HLA associated polymorphism	Is the amino acid part of known epitope (E), or within flanking (F) region (within 5 amino acid)	Is there evidence for reversion in HLA mismatched host (R)?	Entire population	High CD4 cell count strata (>500 cells/μl)	Low CD4 cell count strata (<100 cells/μl)	High CD4 cell count strata (>500 cells/μl)	Low CD4 cell count strata (<100 cells/μl)	p-value (likelihood ratio test between pLOR of high and low CD4 strata)
RT	A80	E	E276D			0.171		0.092			
RT	B4403	E	E306K/N/R	E	R	0.000		0.001			
RT	B1801	E	E306			0.167		0.131			
RT	C06	E	D306					0.143			
RT	C04	K	R310K		R			0.072	16.98	17.95	0.0001
RT	A74	K	K310		R		0.178		16.39	−10.22	0.0059
RT	A80	D	D349					0.195			
RT	A30	I	I356		R			0.121			
RT	A33	P	P371S					0.100	−12.13	5.66	0.0196
RT	B4201	P	P371A	E				0.000	12.06	15.09	0.0040
RT	C1701	P	P371A/S			0.000	0.054				
RT	B42	I	I373V	E	R	0.000		0.037			
RT	A3001	K	K374	F				0.064			
RT	A03	R	K376R	E		0.000		0.006			
RT	A3001	K	K380R			0.000		0.005			
RT	B57	A	A385V				0.178				
RT	C0210	A	A385T				0.198				
RT	B15	A	A385V		R			0.099			
RT	B45	E	E396K					0.153			
RT	C04	D	E423D		R	0.057		0.143			
RT	B4403	I	I425V					0.003	2.39	20.64	0.0259
RT	C03	M	R456				0.178				
RT	B42	R	457R					0.110	−20.83	(infinity)	0.0039
RT	C06	R	R457K		R	0.000		0.019			
RT	A6601	R	R457K		R			0.097			
RT	B5802	T	458T		R			0.103			
RT	C03	A	A470		R			0.150			
RT	C0302	I	I474V		R	0.165		0.000	- (infinity)	(infinity)	0.0015
RT	B15	L	L476I		R			0.128			
RT	B5801	S	S478C/A	E	R	0.004		0.001	- (infinity)	16.17	0.0001
RT	A32	E	E498	E				0.080			
RT	A3201	T	T499I	E		0.001		0.010			
RT	A29	T	499T					0.081			

Table 2. Cont.

						q-value of HLA-related association in each strata			Phylogentically corrected Log2 Odds Ratio (pLOR) towards "escape"		
HIV-1 protein	HLA Class I	Population majority consensus amino acid at associated site	HXB2 site and direction of HLA associated polymorphism	Is the amino acid part of known epitope (E), or within flanking (F) region (within 5 amino acid)	Is there evidence for reversion in HLA mismatched host (R)?	Entire population	High CD4 cell count strata (>500 cells/μl)	Low CD4 cell count strata (<100 cells/μl)	High CD4 cell count strata (>500 cells/μl)	Low CD4 cell count strata (<100 cells/μl)	p-value (likelihood ratio test between pLOR of high and low CD4 strata)
RT	B13	D	D503N			0.007	0.051				
RT	B39	A	A534V					0.033	- (infinity)	(infinity)	0.0018
RT	B5802	A	I534					0.196			
RT	A6802	A	A536V	E	R	0.007		0.053	15.20	15.75	0.0177
RT	A6601	A	A536V	F		0.001	0.178	0.038			
RT	A2301	N	N546H		R	0.073	0.165		12.30	−5.48	0.0030
RT	B4403	E	E548D					0.025			
RT	A3303	I	551I				0.198		(infinity)	- (infinity)	0.0000
RT	B5802	I	I551A		R			0.072			
RT	B4202	V	V566I					0.143			
RT	B08	V	V566I					0.145			
RT	B08	S	S567T			0.000		0.003			
RT	B08	L	L568					0.190			
RT	A3402	T	T569A		R	0.082	0.155				
RT	B18	Q	Q582H	E		0.001		0.073			
RT	B1503	S	S90S	F	R			0.150	−12.70	5.01	0.0196
RT	B44	S	S590L/P	E	R	0.000		0.000	0.74	3.95	0.0408
RT	B1401	S	S590L	F			0.178				
RT	B4403	E	E591	E		0.018		0.011			
RT	B3910	I	I594V		R	0.000		0.005			
RT	B4202	K	K611N		R	0.071		0.120			
RT	B5802	L	L616I					0.009	−11.95	17.59	0.0084
RT	C0701	K	K626N					0.188	−5.73	17.64	0.0045
Integrase	B4403	E	E669A	E	R	0.000		0.011			
Integrase	B4403	E	E670D	E	R	0.000		0.000			

The table shows the results of the association analysis between HLA Class I alleles and HIV viral polymorphisms in HIV Pol. The columns of the table detail the HIV protein, the associated HLA Class I allele and viral polymorphism, whether the association lies in, or within, the flanking region of a known restricted epitope and whether this association is expected to revert in HLA-mismatched hosts. The q values for the statistically significant associations for the whole cohort, the 'high' CD4 count subgroup (>500 cells/μl) and 'low' CD4 count subgroup (<100 cells/μl) are shown, followed by the $\log_2$-adjusted odds ratios for the 'high' and 'low' groups. In the final column, the p value is reported, where there is a significant difference between the two odds ratios.

doi:10.1371/journal.pone.0019018.t002

Table 3. HLA-related associated polymorphisms in the HIV Nef protein in different CD4 count strata.

						q-value of HLA-related association in each strata			Phylogentically corrected Log2 Odds Ratio (pLOR) towards "escape"		
HIV-1 protein	HLA Class I	Population majority consensus amino acid at associated site	HXB2 site and direction of HLA associated polymorphism	Is the amino acid part of known epitope (E), or within flanking (F) region (within 5 amino acid)	Is there evidence for reversion in HLA mismatched host (R)?	Entire population	High CD4 cell count strata (>500 cells/μl)	Low CD4 cell count strata (<100 cells/μl)	High CD4 cell count strata (>500 cells/μl)	Low CD4 cell count strata (<100 cells/μl)	p-value (likelihood ratio test between pLOR of high and low CD4 strata)
Nef	B07	I	I10R	F				0.178			
Nef	A01	E	E24A					0.172	- (infinity)	15.19	0.0024
Nef	B42	A	A26P		R	0.013		0.053			
Nef	A6802	A	A53E					0.134			
Nef	A0201	C	C55	F				0.033			
Nef	B45	E	E65D	E				0.150	- (infinity)	20.98	0.0498
Nef	B45	E	E66D	E		0.000		0.178			
Nef	C0702	R	R71K		R	0.000	0.110	0.000			
Nef	B8101	L	L76T/V	E	R	0.000		0.018			
Nef	B5801	A	A83G	E		0.159		0.003	11.33	26.74	0.0008
Nef	B4403	Y	Y102H	F		0.000		0.000			
Nef	B08	K	K105	E				0.178	−23.73	19.71	0.0001
Nef	C0701	K	K105R	E	R	0.000	0.124	0.000	21.77	23.34	0.0290
Nef	B4403	E	E108D	E		0.000		0.011			
Nef	B18	E	E108D	E		0.000		0.134			
Nef	C0602	Q	Q125H	E	R	0.149		0.030			
Nef	B42	P	P129	E				0.150	−15.22	22.70	0.0036
Nef	B53	V	V133I	F				0.078	6.21	21.38	0.0240
Nef	A23	F	F143Y	F		0.000		0.000			
Nef	C06	S	S187		R			0.104			

The table shows the results of the association analysis between HLA Class I alleles and HIV viral polymorphisms in HIV Nef. The columns of the table detail the HIV protein, the associated HLA Class I allele and viral polymorphism, whether the association lies in, or within, the flanking region of a known restricted epitope and whether this association is expected to revert in HLA-mismatched hosts. The q values for the statistically significant associations for the whole cohort, the 'high' CD4 count subgroup (>500 cells/μl) and 'low' CD4 count subgroup (<100 cells/μl) are shown, followed by the $\log_2$-adjusted odds ratios for the 'high' and 'low' groups. In the final column, the p value is reported, where there is a significant difference between the two odds ratios.

doi:10.1371/journal.pone.0019018.t003

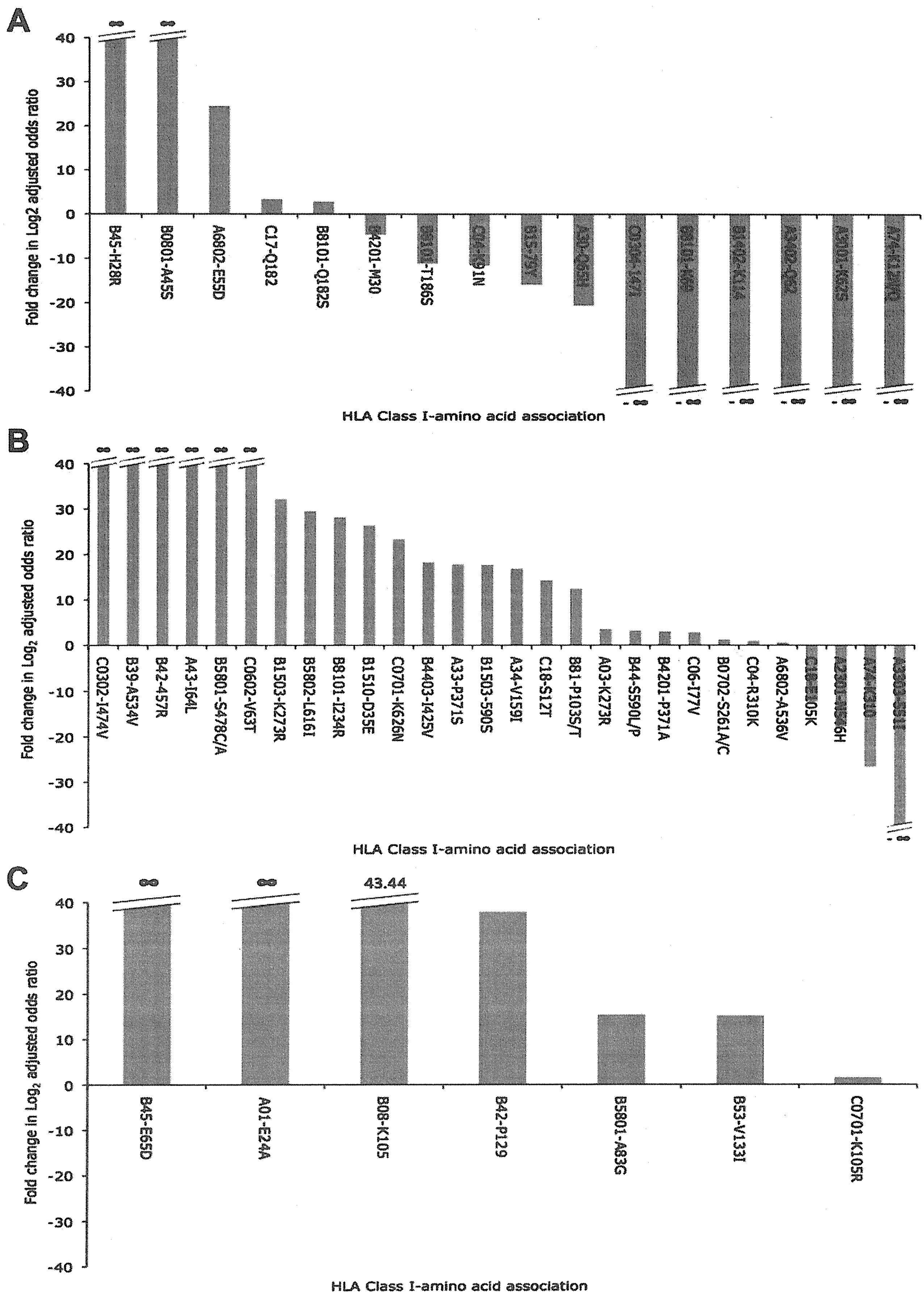
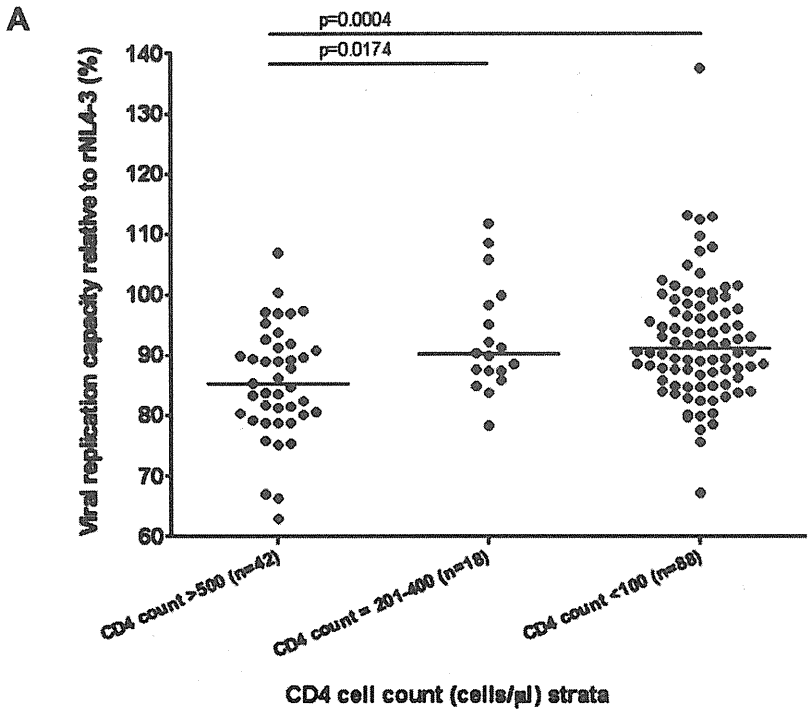

A
Fold change in Log2 adjusted odds ratio
40
30
20
10
0
-10
-20
-30
-40
∞
∞
B45-H28R
B0801-A45S
A6802-E55D
C17-Q182
B8101-Q182S
-∞
HLA Class I-amino acid association
B
Fold change in Log$_2$ adjusted odds ratio
C0302-I474V
B39-A534V
B42-457R
A43-I64L
B5801-S478C/A
C0602-V63T
B1503-K273R
B5802-L616I
B8101-I234R
B1510-D35E
C0701-K626N
B4403-I425V
A33-P371S
B1503-590S
A34-V159I
C18-S12T
B81-P1035/T
A03-K273R
B44-S590L/P
B4201-P371A
C06-I77V
B0702-S261A/C
C04-R310K
A6802-A536V
HLA Class I-amino acid association
C
Fold change in Log$_2$ adjusted odds ratio
43.44
B45-E65D
A01-E24A
B08-K105
B42-P129
B5801-A83G
B53-V133I
C0701-K105R
HLA Class I-amino acid association

Figure 2. Changes in associations between HLA Class I alleles and HIV amino acid polymorphisms for patients with CD4 cell counts <100 cells/μl compared with patients with CD4 cell counts >500 cells/μl. The change in the strength of associations between HLA Class I alleles and specific HIV amino acid polymorphisms in patients with high and low CD4 cell counts is shown for HIV-1 (a.) Gag, (b.) Pol and (c.) Nef. Change in strength of the association is reported on the Y-axis as a log(2)-adjusted value. Associations are recorded on the X-axis according to the restricting HLA Class I allele and the specific viral polymorphism – e.g. B45-H28R in Gag means that in the presence of HLA B*45, Gag amino acid 28 mutates from a Histidine (H) to an Arginine (R). In this case there is at least a 40-fold change in the log(2)-adjusted odds ratio for this association in patients with CD4 counts <100 cells/μl compared with less progressed patients with CD4 cell counts >500 cells/μl. (All fold changes are capped at a value of 40).
doi:10.1371/journal.pone.0019018.g002



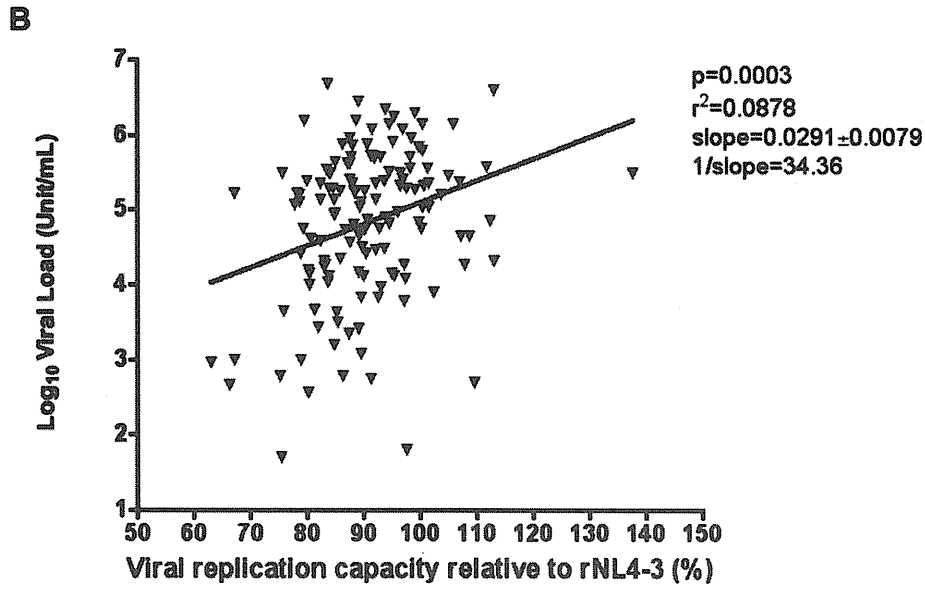
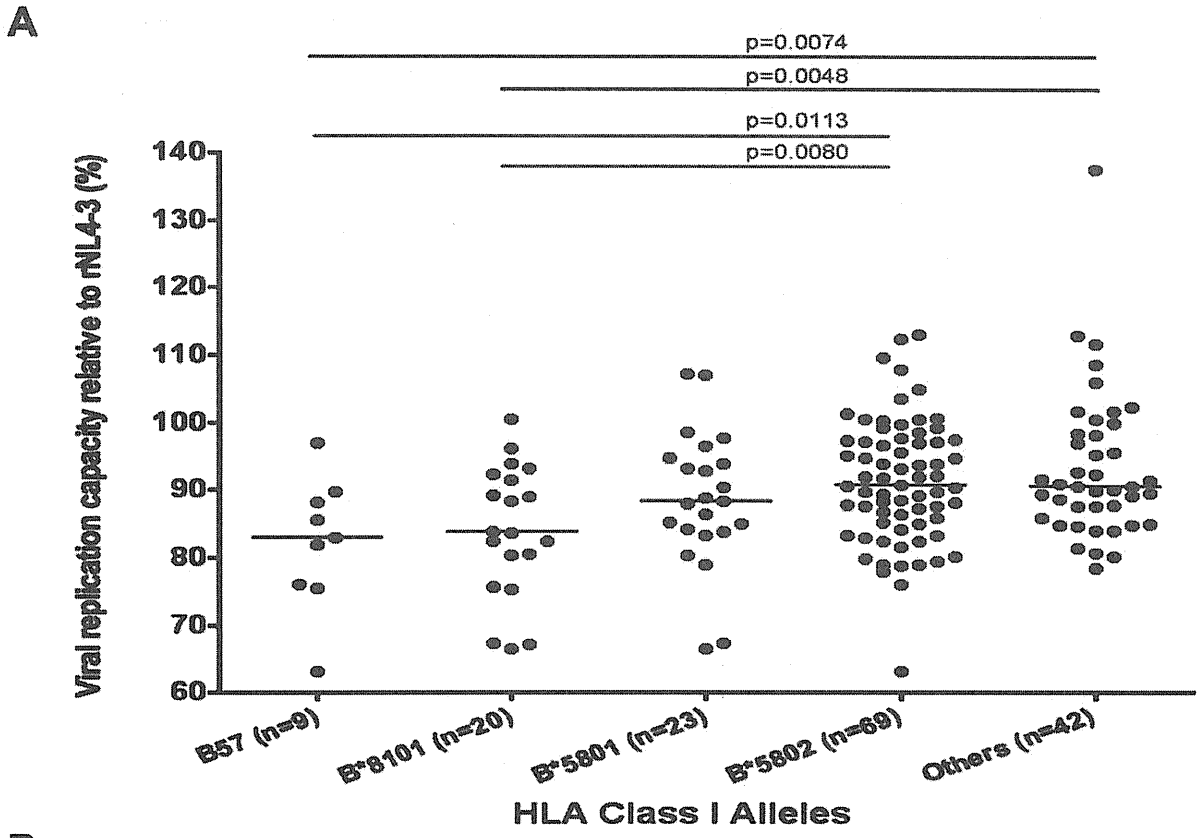


Figure 3. Viral replication capacity (VRC) according to CD4 cell count and plasma viral load. The viral replicative capacity (VRC) of chimeric NL4-3 viruses recombined with patient autologous gag-protease genes is compared with a laboratory recombinant HIV-1 NL4-3 strain. In (a.), The VRC of isolates from the Bloemfontein cohort is presented stratified by CD4 T cell count. The p-values are calculated using the Mann-Whitney test. In (b.), plasma viral load (log_{10} copies/ml) is correlated against the VRC from samples from the Bloemfontein cohort. In both panels VRC is presented as a percentage relative to the fitness of the control NL4-3 virus.
doi:10.1371/journal.pone.0019018.g003


A
p=0.0074
p=0.0048
p=0.0113
p=0.0080
Viral replication capacity relative to rNL4-3 (%)
140
130
120
110
100
90
80
70
60
B57 (n=9)
B*8101 (n=20)
B*5801 (n=23)
B*5802 (n=69)
Others (n=42)
HLA Class I Alleles

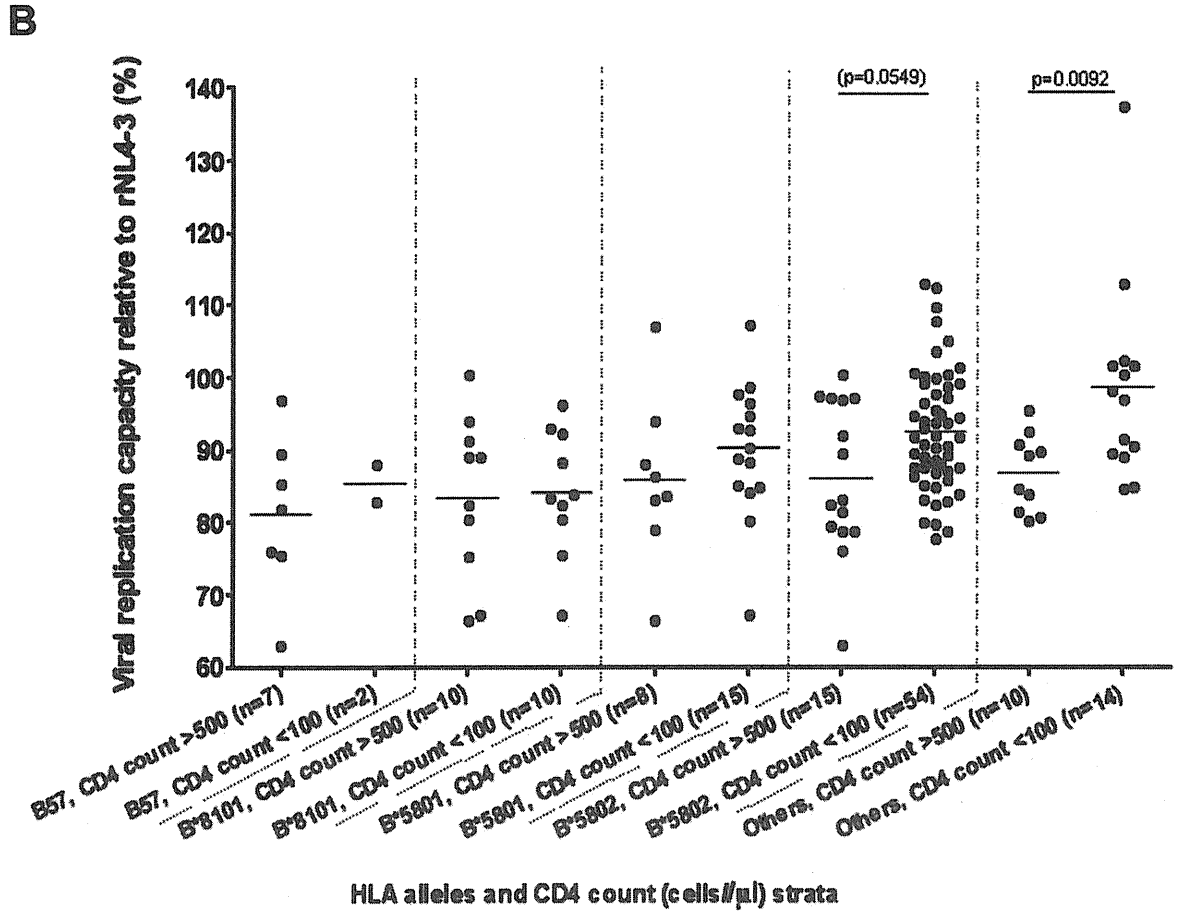

B
(p=0.0549)
p=0.0092
Viral replication capacity relative to rNL4-3 (%)
140
130
120
110
100
90
80
70
60
B57, CD4 count >500 (n=7)
B57, CD4 count <100 (n=2)
B*8101, CD4 count >500 (n=10)
B*8101, CD4 count <100 (n=10)
B*5801, CD4 count >500 (n=8)
B*5801, CD4 count <100 (n=15)
B*5802, CD4 count >500 (n=15)
B*5802, CD4 count <100 (n=54)
Others, CD4 count >500 (n=10)
Others, CD4 count <100 (n=14)
HLA alleles and CD4 count (cells/μl) strata

Figure 4. Viral replication capacity according to HLA Class I and CD4 cell count. (a) Viral replication capacity (VRC) of chimeric NL4-3 viruses containing patient autologous gag-protease stratified by HLA Class I. The VRC is expressed as percentage growth rate compared to control strains. The p-value is calculated using Mann-Whitney test. Panel (b.) shows further stratification according to CD4 cell count. The y-axis depicts VRC, expressed as percentage growth and the x-axis depicts HLA class I, with each HLA class I category divided further into "low" CD4 T cell count strata (<100 cells/mm3) and "high" CD4 T cell count (>500 cells/mm3). The p-value is calculated using Mann-Whitney test.
doi:10.1371/journal.pone.0019018.g004

Compensatory mutations accrue in advanced HIV infection in patients with HLA B*57 and B*5801

As the T242N mutation has been previously linked with compensatory mutations we analysed our cohort to see if these accrued in late stage disease to explain the rise in viral fitness despite the maintenance of T242N. For the T242N mutation in TW10, compensatory mutations have previously been published (H219Q, I223V and M228) [27], and so we examined the frequency of these mutations in the different CD4 T cell count strata (Figure 6). We found a cumulative increase in the number of compensatory mutations in patients with T242N in individuals with lower CD4 cell counts, consistent with the increase in fitness in the previous analyses ($P = 0.013$).

Together these data suggest that advanced HIV disease is associated with an increase in viral fitness, which may contribute to the rise in plasma viral load frequently seen in AIDS. However, this increase in fitness is multifactorial and we report two processes specific to different HLA Class I alleles – reversion due to immune relaxation and compensatory mutations – to potentially explain these findings.

Discussion

In this study, we have explored the hypothesis that progression to AIDS is associated with a rise in viral replicative capacity, and the mechanisms associated with this. We proposed that a rise in viral fitness in AIDS could be due either to a relaxation of CTL-imposed selection pressure resulting in the reversion of costly mutations or, alternatively, if immune pressure is maintained, compensatory mutations might restore the fitness costs of persisting escape mutations.

We initially tested the hypothesis by looking for evidence of changes in T cell-imposed selection pressure measured by gamma interferon ELISPOT assays. We found that the breadth of the response narrowed as the CD4 count declined and that, specifically, the contribution of Gag responses to the overall response decreased. It has previously been reported that acute HIV infection is associated with narrow, high affinity T cell responses which broaden as patients progress into chronic infection [37]. Here, we see the reverse of this process with a return to a much narrower range of T cell responses. It would also be interesting to measure whether there was a loss of high affinity T cells in the advanced stage patients in this cohort as this might be compatible with a weakening in selection pressure.

Although we identified a narrowing of the T cell responses in AIDS, the ELISPOT data alone are not sufficient to infer changes in selection pressure. We therefore looked for evidence of differential selection in patients with high and low CD4 cell counts using a statistical HLA Class I association study. By combining sequence and HLA data from the Bloemfontein cohort and the Durban cohort, we were able to define two groups of patients with CD4 counts less than 100 cells/μl ($n = 196$) and greater than 500 cells/μl ($n = 299$). These CD4 count strata were defined to represent extremes of HIV-related disease without reducing the power of the analysis. The WHO have defined that antiretroviral therapy should be commenced at a CD4 count of 350 cells/μl. Below 200 cells/μl, opportunistic infections become more common. A count of 100 cells/μl or below, therefore, represents advanced HIV disease and progression to AIDS. Above 500 CD4 cells/μl it would be highly unusual for there to be opportunistic infections and the most recent NIH guidelines do not recommend treating above this level [38].

To identify mutations in the HIV genome associated with individual HLA Class I alleles we used an established technique using phylogenetic dependency networks [31], which has been widely implemented in the HIV literature [12,14,35,39]. In this approach, the risk of error due to viral founder effects and misidentification of co-variant sites is corrected for by integrating the phylogeny of the sequences into the analysis. Identified associations are hypothesised to be the result of cytotoxic T cell imposed selection pressure and in other analyses such associations have been proven experimentally. Although there have been a number of HLA-association studies reported, none have stratified patients according to disease progression and compared the strength of selection in different strata according to $\log_2$-adjusted odds ratios.

In our cohorts we identified 151 HLA-associated mutations in HIV-1 *gag*, *pol* or *nef* that were significant in patients with either high or low CD4 cell counts. By comparing $\log_2$ adjusted odds ratios in the CD4 strata, we found that there was evidence for both immune relaxation and intensification in patients with low CD4 counts compared with high CD4 counts. This suggests that for some HLA Class I–restricted responses the selection pressure imposed by CTL is maintained at very low CD4 cell counts but that for others there may be reversion of escape mutations associated with a reduction in selection pressure. Interestingly, the majority of events compatible with immune relaxation were in HIV Gag, which was also the protein for which there was a relative narrowing of T cell ELISPOT responses in AIDS patients.

Two key sites potentially reflecting immune 'persistence' and 'relaxation' were the HLA B*57/B*5801 associated T242N mutation and the B*8101 associated T186S mutation, respectively. For the former, the association analysis provided evidence for on-going mutation associated with compensatory mutations, whereas for the latter there was evidence for reversion. A limitation of this study is that it is cross-sectional rather than longitudinal, potentially biasing our data interpretation. An alternative explanation for immune relaxation is that the patients who progress are the ones who do not make an immune response and so do not select mutants, therefore enriching patients with low CD4 counts with wild-type viruses. This would be compatible with the finding that targeting certain key epitopes may be associated with different speeds of clinical progression [40], but does not explain the eventual outcome of those patients enriched for immune escape mutations with high CD4 cell counts. To determine which occurs would require a prospective longitudinal study of untreated progression to AIDS. In our cohort, analyses of HLA B*8101 positive individuals with data on ELISPOT responses and viral sequence, show that for patients with CD4 counts between 200–500 cells/μl, 50% (3/6) who did not recognise TL9 carried the T186S mutant. In contrast, for patients with CD4 counts <100 cells/μl there were no ELISPOT negative patients with T186S (0/2), suggesting that these had reverted

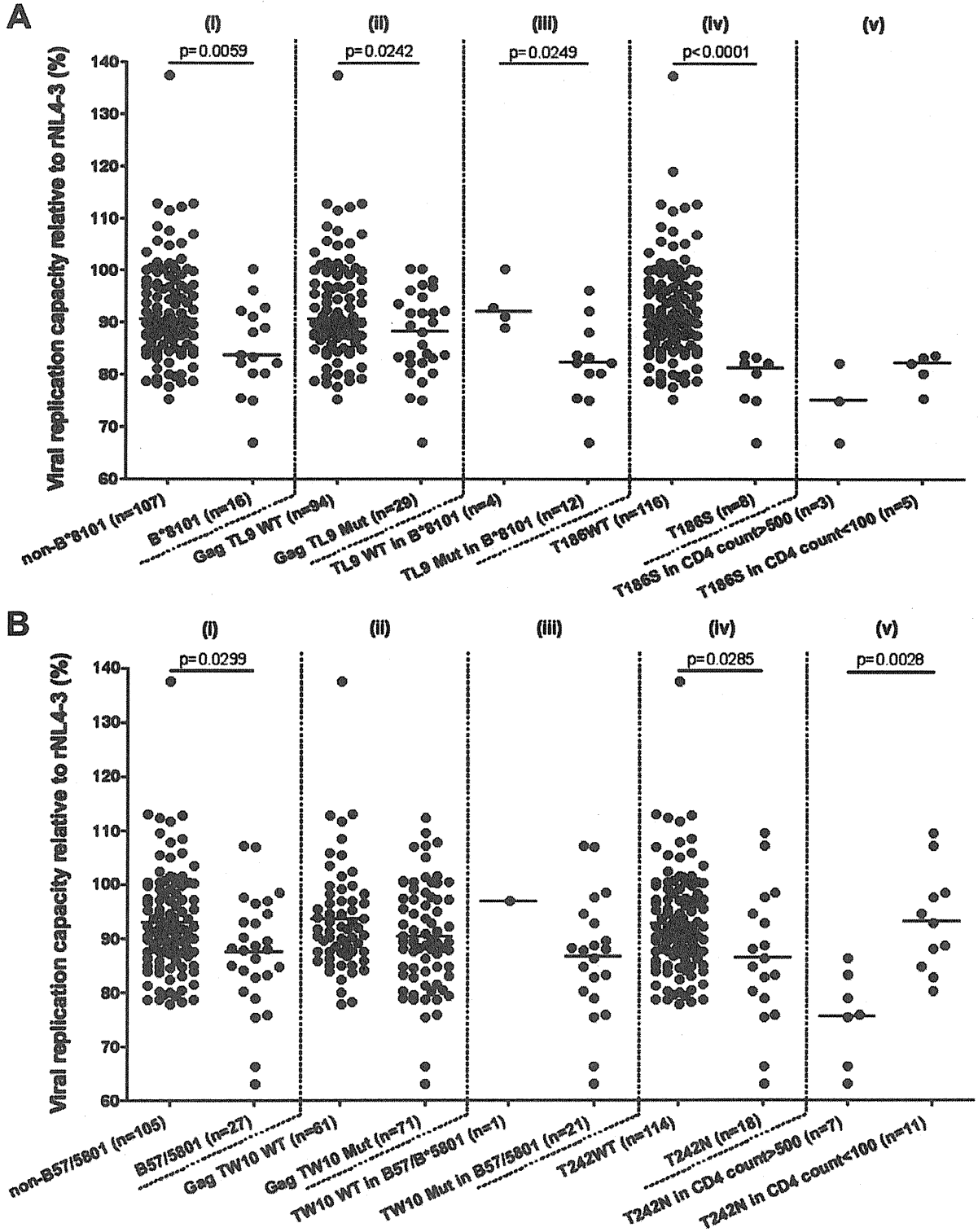


Figure 5. Viral replication capacity (VRC) in viruses with mutants in (a) Gag TL9 and (b) Gag TW10 according to HLA Class I and CD4 cell count. In panel (a), the y-axis depicts viral replication fitness (VRC) of chimeric NL4-3 viruses containing patient autologous *gag-protease*, expressed as percentage growth rate compared to a control recombinant NL4-3 strain. The x-axis depicts the stratification of patients according to (i) HLA-B*8101, (ii) polymorphic status within the *gag* TL9 epitope in all patients, (iii) polymorphic status of *gag* TL9 in patients with HLA B*8101, (iv) presence of T186S in *gag* in Gag TL9, and (v) T186S further stratified according to the CD4 T cell count (<100 cells/μl or >500 cells/μl). The p-value is calculated using Mann-Whitney test. In (b), patients are stratified according to (i) HLA-B*57 or B*5801 alleles, (ii) polymorphic status within the *gag* TW10 epitope, (iii) polymorphic status of TW10 in patients with HLA-B*57 or B*5801 alleles, (iv) polymorphic status of amino acid position 242 (within TW10 epitope – T242N) in all patients and (v) by T242N in patients with HLA-B*57 or B*5801 according to CD4 count. The viral isolates possessing other known costly escape mutations (A163G and T186S) are excluded from this analysis. The p-value is calculated using Mann Whitney test.
doi:10.1371/journal.pone.0019018.g005

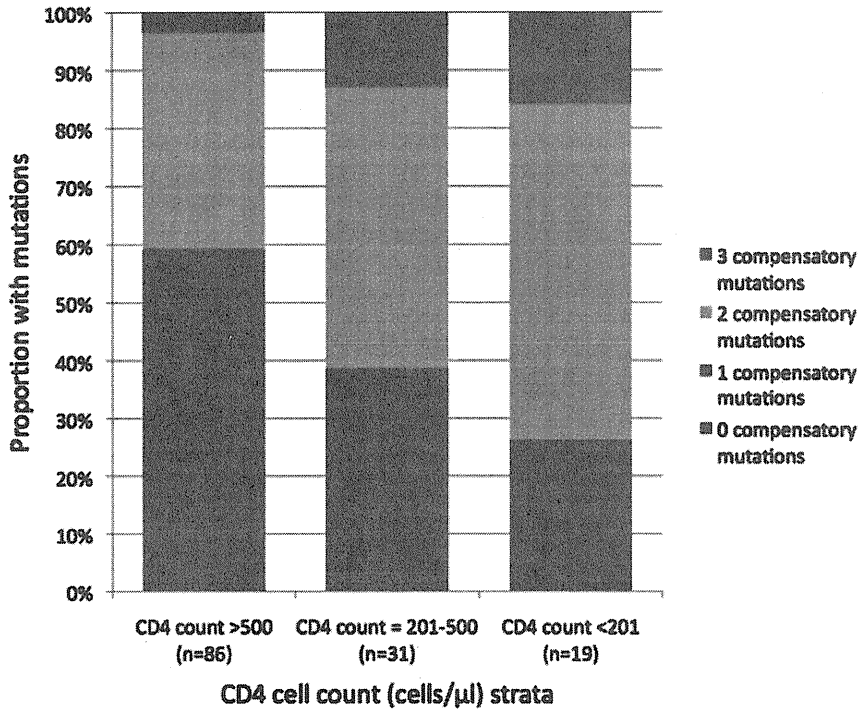


Figure 6. Prevalence of compensatory mutations with the T242N mutation in the HLA-B57 and B*5801 restricted epitope, Gag TW10. In the 100% stacked column, the y-axis shows the number of mutations accumulated at the three Gag protein amino acid sites (H219Q, I223V and M228I) that compensate for the T242N mutation [27]. The x-axis depicts the categorical CD4 T cell count strata and the number of patients within each strata (the "High CD4" group refers to patients with CD4 T cell counts >500 cells/μl, the "Intermediate CD4" group refers to patients with CD4 T cell counts between 200 and 400 cells/μl, and the "Low CD4" group refers to patients with CD4 T cell counts <100 cells/μl).
doi:10.1371/journal.pone.0019018.g006

rather than having been consistently wild type, although in this study numbers are too small to be conclusive.

Viral fitness was measured using a recombinant assay in which a patient-derived RT-PCR amplicon of HIV-1 *gag-protease* was inserted into a *gag-protease*-deleted pNL4-3 plasmid. The advantage of this approach is that it uses the full *gag* gene from the patient sequence as well as the associated *protease* that encodes the enzyme for cleavage of the Gag-Pol protein. The disadvantage of this assay is that it is in a parallel rather than competitive format, and therefore may lack sensitivity for small fitness differences. In addition, the construct is a recombinant of a subtype B backbone with a subtype C insert. Although this approach has been published elsewhere [32], any data should be interpreted relative to the control virus rather than as an absolute measure of viral fitness. Choice of control is important. We used a subtype B wild-type backbone (pNL4-3Δ*gag-protease*) into which we recombined the pNL4-3 *gag-protease*. This controlled for any change in fitness due to the recombination step.

We found that VRC was significantly greater in patients with low CD4 cell counts and there was a significant, although weak, correlation with plasma viral load. These data, which are consistent with a population-based study of HIV subtype B-infected individuals [33], are amongst the first supporting a link between viral fitness and plasma viral load, and show that this *ex vivo* assay reflects viral replication *in vivo*. Analysing our data by HLA Class I revealed that HLA Class I alleles that are associated with clinical advantage (B*8101, B*57, B*5801) were associated with impaired viral fitness, as had been inferred from previous analyses of the reversion of transmitted escape mutations in HLA mismatched hosts [22,23]. Interestingly, however, the increase in viral fitness with lower CD4 cell counts did not appear to be restricted to beneficial HLA alleles and was also present in a selection of viruses from patients with 'neutral/other' alleles.

The HLA-amino acid association analysis showed that for HLA B*8101 there was a decrease in mutations in the Gag TL9 epitope at low CD4 cell counts, whereas for HLA B*57/B*5801 there was persistence of the T242N escape mutation in TW10. The fitness assays showed that the T186S mutation in TL9 was associated with a fitness cost ($P<0.0001$) and it is therefore possible that in patients with B*8101 and low CD4 counts, the decrease in prevalence of the T186S mutation reflects reversion to a fitter strain coincident with the weakening of T cell responses. For B*57/B*5801 we saw a different pattern. Here, possession of HLA B*57/B*5801 was associated with less fit viruses in conjunction with the T242N mutation in the Gag TW10 epitope, but in patients with low CD4 counts fitness was restored even though T242N persisted. We therefore sought compensatory mutations to explain this finding. For T242N there are four documented compensatory mutations at Gag codons 219, 223, 228 and 248 [27]. For this analysis of subtype C HIV-1 we excluded codon 248, as the 248A variant represents the subtype C consensus. When we stratified patients with T242N according to CD4 strata we found a significant ($p = 0.013$) increase in numbers of compensatory mutations at lower CD4 counts, consistent with the rise in VRC.

These data show that viral fitness increases with progression to AIDS. Whether this is cause or effect is difficult to determine, although the fact that we see evidence for both reversion and compensatory mutations in the same cohort indicates that progression may be a mixture of the two. What is clear is that viral fitness is a significant component of progression to AIDS and, as such, should be considered as a target for intervention, for example through vaccines aimed at epitopes which escape with

high fitness costs. Trials such as DART have shown that antiretroviral therapy can be extremely effective in regions such as sub-Saharan Africa [41], however the costs and logistics of provision are complex. Any intervention that could prevent new infections or delay a requirement for therapy could have major economic and health implications and therefore combining such a vaccine with antiretroviral provision could be a productive strategy.

Acknowledgments

We would like to express our gratitude to the physicians and patients involved in the Bloemfontein cohort and the patients and staff at the Cato Manor Clinic and Sinikithemba Clinic, Durban, South Africa without whom this study would not have been possible. Authors acknowledge the generous contribution of the National Health Laboratory Service (NHLS) in the Free State that transported specimens at no cost. Many thanks to Chia-Min Chang for his assistance with data analysis, and to Jennifer Listgarten for helpful discussions regarding the phylogenetic odds ratios.

Author Contributions

Conceived and designed the experiments: K-HGH DG JC MB SM ZB PK DS PG RP CVV JF CS. Performed the experiments: K-HGH DG SM SH CT. Analyzed the data: JC DH JF K-HGH. Contributed reagents/materials/analysis tools: TM MB TN BW. Wrote the paper: K-HGH JF RP.

References

1. Herbeck JT, Gottlieb GS, Winkler CA, Nelson GW, An P, et al. Multistage genomewide association study identifies a locus at 1q41 associated with rate of HIV-1 disease progression to clinical AIDS. J Infect Dis 201: 618–626.
2. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, et al. (2007) A whole-genome association study of major determinants for host control of HIV-1. Science 317: 944–947.
3. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, et al. (2009) Common genetic variation and the control of HIV-1 in humans. PLoS Genet 5: e1000791.
4. Carrington M, O'Brien SJ (2003) The influence of HLA genotype on AIDS. Annu Rev Med 54: 535–551.
5. Gao X, Bashirova A, Iversen AK, Phair J, Goedert JJ, et al. (2005) AIDS restriction HLA allotypes target distinct intervals of HIV-1 pathogenesis. Nat Med 11: 1290–1292.
6. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, et al. (2004) Dominant influence of HLA-B In mediating the potential co-evolution of HIV and HLA. Nature 432: 769–774.
7. Kosmrlj A, Read EL, Qi Y, Allen TM, Altfeld M, et al. (2010) Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. Nature 465: 350–354.
8. Shafer RW, Rhee SY, Bennett DE (2008) Consensus drug resistance mutations for epidemiological surveillance: basic principles and potential controversies. Antivir Ther 13 Suppl 2: 59–68.
9. Shafer RW, Rhee SY, Pillay D, Miller V, Sandstrom P, et al. (2007) HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. Aids 21: 215–223.
10. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, et al. (1991) Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature 354: 453–459.
11. Goulder PJ, Watkins DI (2004) HIV and SIV CTL escape: implications for vaccine design. Nat Rev Immunol 4: 630–640.
12. Brumme ZL, Tao I, Szeto S, Brumme CJ, Carlson JM, et al. (2008) Human leukocyte antigen-specific polymorphisms in HIV-1 Gag and their association with viral load in chronic untreated infection. Aids 22: 1277–1286.
13. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, et al. (2007) Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. PLoS Pathog 3: e94.
14. Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, et al. (2007) Founder effects in the assessment of HIV polymorphisms and HLA allele associations. Science 315: 1583–1586.
15. Feeney ME, Tang Y, Roosevelt KA, Leslie AJ, McIntosh K, et al. (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. J Virol 78: 8927–8930.
16. Kawashima Y, Pfafferott K, Frater J, Matthews P, Payne R, et al. (2009) Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458: 641–645.
17. Frater AJ, Brown H, Oxenius A, Gunthard HF, Hirschel B, et al. (2007) Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. J Virol 81: 6742–6751.
18. Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, et al. (2006) Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. J Virol 80: 3617–3623.
19. Miura T, Brockman MA, Brumme ZL, Brumme CJ, Pereyra F, et al. (2009) HLA-associated alterations in replication capacity of chimeric NL4-3 viruses carrying gag-protease from elite controllers of human immunodeficiency virus type 1. J Virol 83: 140–149.
20. Goepfert PA, Lumm W, Farmer P, Matthews P, Prendergast A, et al. (2008) Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. J Exp Med 205: 1009–1017.
21. Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, et al. (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. J Virol 81: 12382–12393.
22. Duda A, Lee-Turner L, Fox J, Robinson N, Dustan S, et al. (2009) HLA-associated clinical progression correlates with epitope reversion rates in early human immunodeficiency virus infection. J Virol 83: 1228–1239.
23. Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, et al. (2008) Marked epitope and allele-specific differences in rates of mutation in HIV-1 Gag, Pol and Nef CTL epitopes in acute/early HIV-1 infection. J Virol.
24. Addo MM, Draenert R, Rathod A, Verrill CL, Davis BT, et al. (2007) Fully differentiated HIV-1 specific CD8+ T effector cells are more frequently detectable in controlled than in progressive HIV-1 infection. PLoS ONE 2: e321.
25. Zhuang Y, Sun Y, Zhai S, Huang D, Zhao S, et al. (2008) Relative dominance of Env-gp41-specific cytotoxic T lymphocytes responses in HIV-1 advanced infection. Curr HIV Res 6: 239–245.
26. Gandhi RT, Wurcel A, Rosenberg ES, Johnston MN, Hellmann N, et al. (2003) Progressive reversion of human immunodeficiency virus type 1 resistance mutations in vivo after transmission of a multiply drug-resistant virus. Clin Infect Dis 37: 1693–1698.
27. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, et al. (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. J Virol 81: 12608–12618.
28. Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, et al. (2001) Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J Exp Med 193: 375–386.
29. Huang KH, Goedhals D, Fryer H, van Vuuren C, Katzourakis A, et al. (2009) Prevalence of HIV type-1 drug-associated mutations in pre-therapy patients in the Free State, South Africa. Antivir Ther 14: 975–984.
30. Addo MM, Yu XG, Rathod A, Cohen D, Eldridge RL, et al. (2003) 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: 2081–2092.
31. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, et al. (2008) Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol 4: e1000225.
32. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, et al. (2010) Gag-Protease-Mediated Replication Capacity in HIV-1 Subtype C Chronic Infection: Associations with HLA Type and Clinical Parameters. J Virol.
33. Brockman MA, Brumme ZL, Brumme CJ, Miura T, Sela J, et al. (2010) Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated in chronic infection. J Virol.
34. Miura T, Brumme ZL, Brockman MA, Rosato P, Sela J, et al. (2010) Impaired replication capacity of acute/early viruses in persons who become HIV controllers. J Virol 84: 7581–7591.
35. Matthews PC, Prendergast A, Leslie A, Crawford H, Payne R, et al. (2008) Central role of reverting mutations in HLA associations with human immunodeficiency virus set point. J Virol 82: 8548–8559.
36. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, et al. (2004) HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 10: 282–289.
37. Streeck H, Jolin JS, Qi Y, Yassine-Diab B, Johnson RC, et al. (2009) Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. J Virol 83: 7641–7648.
38. Services DoHaH (2009) Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. http://wwwaidsinfonihgov/ContentFiles/AdultandAdolescentGLpdf. pp 1–161.
39. Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, et al. (2008) HLA class I-driven evolution of human immunodeficiency virus type 1 subtype c proteome: immune escape and viral load. J Virol 82: 6434–6446.

40. Dinges WL, Richardt J, Friedrich D, Jalbert E, Liu Y, et al. (2010) Virus-specific CD8+ T-cell responses better define HIV disease progression than HLA genotype. J Virol 84: 4461–4468.

41. Walker S, DART Trial Team. Design of the DART trial and key substudy results; 2009; Cape Town, South Africa.

Coordinate linkage of HIV evolution reveals regions of immunological vulnerability

Vincent Dahirel[a,b,c,1], Karthik Shekhar[a,b,1], Florencia Pereyra[a], Toshiyuki Miura[d], Mikita Artyomov[c,e], Shiv Talsania[b,f], Todd M. Allen[a], Marcus Altfeld[a], Mary Carrington[a,g], Darrell J. Irvine[a,h,i], Bruce D. Walker[a,h,2] and Arup K. Chakraborty[a,b,c,i,2]

[a]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard University, Boston, MA 02129; Departments of [b]Chemical Engineering, [c]Chemistry, and [i]Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Department of Chemistry, Moscow State University, Moscow 119991, Russia; [f]Department of Chemical Engineering, Loughborough University, Leicestershire LE11 3TU, United Kingdom; [g]Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, MD 21702; [h]Howard Hughes Medical Institute, Chevy Chase, MD 20815; and [d]Institute for Medical Sciences, University of Tokyo, Tokyo 108-8639, Japan

Edited* by Laurie H. Glimcher, Harvard University, Boston, MA, and approved May 20, 2011 (received for review April 4, 2011)

Cellular immune control of HIV is mediated, in part, by induction of single amino acid mutations that reduce viral fitness, but compensatory mutations limit this effect. Here, we sought to determine if higher order constraints on viral evolution exist, because some coordinately linked combinations of mutations may hurt viability. Immune targeting of multiple sites in such a multidimensionally conserved region might render the virus particularly vulnerable, because viable escape pathways would be greatly restricted. We analyzed available HIV sequences using a method from physics to reveal distinct groups of amino acids whose mutations are collectively coordinated ("HIV sectors"). From the standpoint of mutations at individual sites, one such group in Gag is as conserved as other collectively coevolving groups of sites in Gag. However, it exhibits higher order conservation indicating constraints on the viability of viral strains with multiple mutations. Mapping amino acids from this group onto protein structures shows that combined mutations likely destabilize multiprotein structural interactions critical for viral function. Persons who durably control HIV without medications preferentially target the sector in Gag predicted to be most vulnerable. By sequencing circulating viruses from these individuals, we find that individual mutations occur with similar frequency in this sector as in other targeted Gag sectors. However, multiple mutations within this sector are very rare, indicating previously unrecognized multidimensional constraints on HIV evolution. Targeting such regions with higher order evolutionary constraints provides a novel approach to immunogen design for a vaccine against HIV and other rapidly mutating viruses.

cytotoxic T-lymphocyte response | elite controllers | random matrix theory

Despite the efficacy of life-extending medications, HIV continues to wreak havoc around the world, particularly in poor nations. An efficient vaccine is urgently needed, and such a vaccine is likely to be one that induces both antibody and cytotoxic T-lymphocyte (CTL) responses (1–3). The extreme variability of the HIV envelope has precluded vaccine-induced generation of broadly neutralizing antibodies (1) that can prevent acquisition, leading some to focus on CTL-based vaccines (4) to prevent disease progression and possibly to limit acquisition (5).

CTLs recognize and respond to viral peptides presented by class I MHC proteins. Single point mutations within and surrounding such HIV epitopes targeted by CTLs can enable escape from immune pressure, leading to a focus on targeting conserved regions of the HIV proteome. Conserved regions are defined to be ones where the frequency of occurrence of mutations at single sites is small, indicating [according to evolutionary theory (6)] that the corresponding mutant viral strains are replicatively less fit. Thus, if such a site is targeted, the outgrowth of a mutant virus that escapes the immune pressure is less likely (7). The emergence of a compensatory mutation that restores fitness is a challenge to this approach (8).

Characterizing the frequency of occurrence of viral strains based on the effects of mutations at single sites results in a unidimensional measure of conservation, which ignores potential couplings between the effects of multiple simultaneous mutations due to structural/functional constraints. A useful multidimensional measure of sequence variation would be obtained if groups of sites in the viral proteome could be identified, such that sites within a group coevolve in a collectively interdependent manner to influence virus viability but each group evolves independent of other groups. If such groups exist, it is possible that a greater proportion of the combinations of mutations involving sites in the group are harmful in some groups compared with other groups. Such a group of sites is multidimensionally more conserved in that multiple mutations are more likely to hurt virus fitness and harmful combinations are less likely to be compensated for by mutations in another group (as it coevolves independently). Such regions should be particularly vulnerable to multiple points of CTL pressure, because escape pathways would be restricted. Targeting multiple points would promote the emergence of multiple mutations to escape the immune pressure, but multiple mutations in such a region would be more likely to result in unfit viruses.

Thus, we sought to determine collectively coevolving groups of residues in HIV proteins. We analyzed publicly available sequences from the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov/) of clade B HIV proteins derived from patients with diverse genotypes using random matrix theory (RMT) (9) to identify groups of sites (not pairs) that evolve collectively, presumably because they act together to perform a function important for the virus. RMT has been applied to diverse realms of physics, to analyze stock price fluctuations (10, 11), and to analyze sequences of an enzyme (12). Although inspired by Halabi et al. (12), we outline our analysis of HIV proteins by analogy with analyses of financial markets.

From price fluctuations of various stocks over time, one can compute the extent of correlation between changes in price of a pair of stocks, averaged over all available time points. This yields a matrix of pair correlations. Each element of the matrix describes correlations between a particular pair of stocks, and mathematical

Author contributions: V.D., K.S., and A.K.C. designed research; V.D., K.S., and A.K.C. performed research; F.P., T.M., M.C., and B.D.W. contributed new reagents/analytic tools; V.D., K.S., M. Artyomov, S.T., B.D.W., and A.K.C. analyzed data; and V.D., K.S., T.M.A., M. Altfeld, D.J.I., B.D.W., and A.K.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]V.D. and K.S. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: bwalker@partners.org or arupc@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105315108/-/DCSupplemental.

properties of the entire matrix (*Results*) reveal groups of stocks whose price fluctuations are collectively interdependent. These properties of the matrix are influenced by "noise" because of a finite sample of times and correlations that reflect changes in overall stock prices attributable to external forces (e.g., recessions). RMT has been used to "clean" the correlation matrix of these effects and obtain groups of companies whose economic activities are intrinsically collectively coupled and essentially independent of others (10, 11). Consistent with intuition, each group was composed of stocks that define known economic "sectors," such as financial service companies and automotive companies.

A similar analysis has enabled us to identify collectively coevolving groups of sites in the HIV proteome (which we term HIV sectors). Further, we have identified a sector in which mutations at multiple sites are collectively more constrained, and thus might be particularly vulnerable to multiple points of immune pressure. We examined immune targeting and sequenced circulating viruses in those rare persons who are able to control HIV without medications to determine whether such mechanisms might contribute to durable immune control. These data support our predictions. We suggest how this new understanding can be harnessed to design immunogens for a vaccine against HIV.

Results

Sequence Analysis Identifies Significant HIV Sectors. Each aligned sequence of an HIV polyprotein is analogous to stock price data at a particular time point. In a specific sequence, one asks if a mutation away from the most frequent amino acid ("wild type") at a particular position i appears simultaneously with a similar mutation at another site j, and this is done for all pairs of amino acids. Average values for the frequency of occurrence of double mutations for each pair of sites in the protein (f_{ij} for sites i and j) are obtained by repeating this procedure for all sequences (*Methods* and *SI Appendix* 1). Similarly, the frequencies with which mutations are observed at individual sites (f_i and f_j for sites i and j) are obtained. A pair correlation matrix, **C**, can be

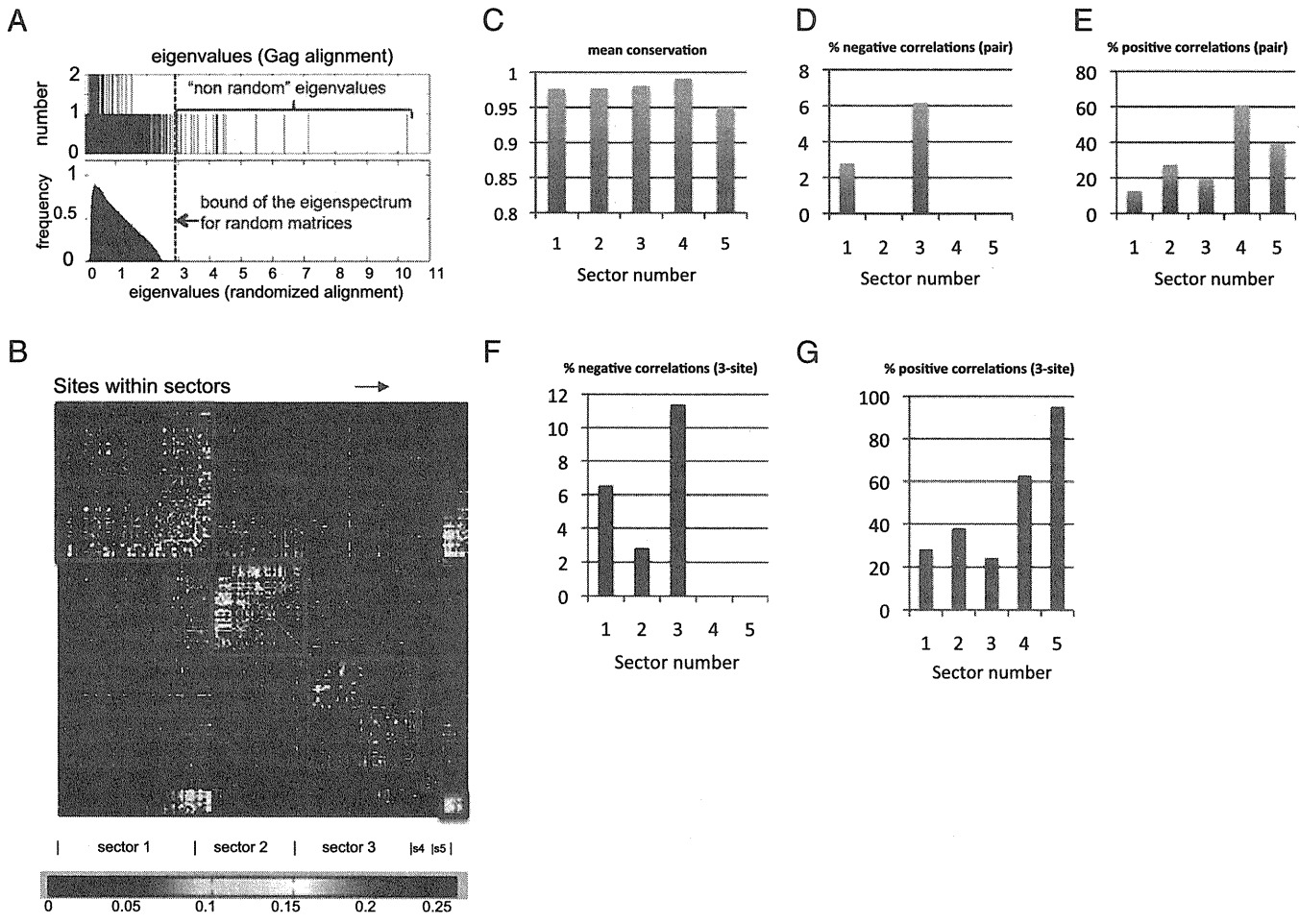


Fig. 1. Defining collectively coevolving groups of sites (sectors) in Gag. (*A, Upper*) Number of eigenvalues (ordinate) of the correlation matrix, **C**, for Gag (defined in the main text) with a given magnitude (abscissa) is shown. In total, 1,600 sequences of Gag polyproteins (500 residues long) were used. Note that a relatively large number of sequences of HIV proteins are available. (*Lower*) Distribution of eigenvalues obtained from 1,000 randomly generated matrices of the same size as **C** for Gag (described in the main text and *SI Appendix* 2). (*B*) Analyzing the eigenvectors corresponding to eigenvalues larger than the highest eigenvalue for random matrices yields collectively coevolving groups of sites or "sectors" (*SI Appendix* 3–7). Sites within a sector are grouped together along the rows and columns so that groups of collectively coevolving sites are vivid as squares along the diagonal of a heat map representing the values of the correlations (obtained from the "cleaned" correlation matrix as described in *SI Appendix* 2). (*C*) Mean frequency of the dominant amino acid at single sites within each of the five Gag sectors. (*D–F*) Threshold value defining a significant correlation was chosen to be such that correlations with magnitudes greater than this value arise with vanishing probability in the randomized matrices ($P < 0.02$). Changing this threshold value does not change qualitative results (*SI Appendix*, Fig. S11). (*D*) Percentage of significant negative ($C_{ij} < -0.03$) pair correlations within each sector. (*E*) Percentage of significant positive ($C_{ij} > 0.1$) pair correlations within each sector. (*F*) Percentage of significant three-site negative ($C_{ijk} < -0.01$) correlations within each sector (method in *SI Appendix* 17). (*G*) Percentage of significant three-site positive ($C_{ijk} > 0.1$) correlations within each sector.

defined, each element of which, C_{ij} (for sites i and j), reflects the interdependence of mutations at these two sites. $C_{ij} = (f_{ij} - f_i f_j)/\sqrt{V_i V_j}$ and measures the difference between the frequency of occurrence of a double mutant and that which would be observed if the two mutations occurred independently; the variances of the distributions of mutations (V_i and V_j) at sites i and j, respectively, normalize the values of C_{ij} for different pairs of sites. Another method (12) for computing C_{ij} yielded qualitatively similar results (*SI Appendix*, Fig. S6).

Properties of this correlation matrix, termed eigenvectors, contain information about collective coevolution of sites. Each eigenvector represents a specific combination of sites, whose mutations occur in a coupled way but are essentially independent of mutations in a combination of sites represented by another eigenvector. The data on HIV evolution can be represented as dependent on these combinations of sites (eigenvectors), rather than individual sites. The association of each site with a particular eigenvector is specified by a number that can be positive or negative. An eigenvector is also associated with a number, called an eigenvalue, whose magnitude reflects the contribution of this eigenvector to the correlations between sites (*SI Appendix*, Eq. S4).

Eigenvalues of the correlation matrix, C, for HIV Gag polyproteins are shown ordered according to their magnitudes (Fig. 1*A*, *Upper*). The corresponding eigenvectors describe groups of sites whose mutations are collectively linked, presumably by the requirements of maintaining viral fitness. However, the information in some eigenvectors is not significant because of noise (attributable to a finite sample of sequences) or because it reflects phylogeny (13).

RMT theorems can help to determine which eigenvectors reflect the influence of noise because they describe the properties of correlation matrices derived from independent random variables. For example, given the length of the Gag polyprotein and the number of available sequences, RMT states (*SI Appendix*, Eq. **S3**) that the largest eigenvalue would equal 2.4 if the correlation matrix reflected noise only. However, this theorem assumes that the length of the polyprotein and the number of sequences are very large. If we randomly shuffle the identity of amino acids at each site over all aligned sequences of a particular HIV protein and recalculate C, we obtain a random correlation matrix of the same size as our sample (*SI Appendix* 2). The distribution of eigenvalues obtained by randomizing the sequences 1,000 times is shown in Fig. 1*A*, *Lower*. We see that the RMT theorem is rather accurate even for finite samples, because very few eigenvalues exceed 2.4 and none exceed 3. The continuous part of the eigenvalue spectrum for the real-sequence alignment for HIV proteins is bounded in the same way as that for the randomized sequences (compare panels in Fig. 1*A* for Gag). The eigenvectors corresponding to eigenvalues <3 correspond to noise.

As noted before (12), and derived more completely in *SI Appendix* 3, the contribution of phylogeny alone to a representation of the data in terms of eigenvectors should result in an eigenvector where the contributions from each individual site in the polyprotein have the same sign. For all HIV polyproteins, the largest eigenvalue corresponds to such an eigenvector. Because each eigenvector is independent, we conclude that the eigenvector corresponding to the largest eigenvalue describes correlations attributable to phylogeny (13). Eigenvectors corresponding to the next few largest eigenvalues contain information on collective correlations between groups of sites that originate largely from relationships between evolutionary constraints and sequence.

By parsing these eigenvectors, groups of sites (sectors) in HIV proteins that coevolve together but essentially independent of others were identified. Many sites do not collectively coevolve with any other sites, and thus do not belong to any sector. For Gag, we found five strongly collectively coupled groups of sites (*SI Appendix* 6 and Fig. S4), which we term sectors 1–5. As expected, we also found an additional group of residues that is weakly coupled collectively, as a result of mutations arising in the context of specific HLA class I molecules that drive mutations at multiple sites (*SI Appendix* 8 and Figs. S4 and S5). The weakly collective linkage between this group of sites is attributable to HLA-associated "footprints" [i.e., these mutations arise in persons with the same HLA type (14)]; thus, we do not consider this quasi-sector further.

Gag sectors 1–5 can be visualized as a heat map reflecting the correlations between sites (Fig. 1*B*). As expected, the sites that comprise a sector are not contiguous along the linear protein sequence. The positions of the sites along the rows and columns in Fig. 1*B* group the sites in a sector together. For example, if a sector was composed of sites 27, 45, and 53 along the linear protein sequence, these sites would be adjacent to each other along a row or column. Each group of sites that comprises a sector is outlined in red. We observed similar groupings for other HIV proteins (*SI Appendix*, Figs. S7–S10). We focus further analysis on Gag because of data indicating correlations between Gag-specific responses and viral control (15) and a strong impact of Gag mutations on viral fitness (16).

Identification of an Immunologically Vulnerable Sector in Gag. From the standpoint of identifying regions with single sites that are more conserved, all five Gag sectors are equivalent (Fig. 1*C*). However, if one examines the relative occurrence of positive and negative correlations between multiple mutations at sites that comprise each collectively evolving sector, differences between the sectors become apparent.

Negative correlation between sites implies that the multiple mutant is observed less frequently than if the individual mutations were to arise independently. For example, consider a simple case where a pair of sites (i and j) is negatively correlated, with each site being 90% conserved [i.e., the frequency with which single mutations at sites i and j are observed is 10%, ($f_i = f_j = 0.1$)]. For C_{ij} (proportional to $f_{ij} - f_i f_j$) to be negative, the frequency with which the double mutant is observed (f_{ij}) must be less than 1%. This implies that when multiple sites are negatively correlated, the multiple mutant is considerably less fit compared with the single mutants. If immune pressure is applied to multiple sites in a sector, escaping the immune pressure is likely to require more than one mutation. The greater the proportion of negative correlations in the sector, the more difficult it is to both escape the CTL pressure and maintain virus fitness, because multiple mutations are much less tolerated, likely because of fitness limitations.

Positive correlation between sites implies that the corresponding multiple mutants are observed more frequently than if the mutations were to occur independently. Thus, positive correlations can be associated with compensatory mutations; indeed, known compensatory mutations do appear as positive correlations (*SI Appendix*, Table S9). Thus, sectors characterized by larger numbers of positive correlations would be less effective targets, because escape from CTLs without substantial loss in virus fitness is more likely.

For pairs of sites within a sector, one finds that Gag sector 3 contains the largest proportion of negative correlations compared with other sectors and a relatively small number of positive correlations (Fig. 1 *D* and *E*). This is also true for three-site (Fig. 1 *F* and *G*) and higher order correlations. The magnitudes of the negative pair correlations between sites in sector 3 are as large as they can be, given the level of single site conservation in this sector (details in *SI Appendix* 11), meaning that the double mutants are expected to be observed very rarely. The eigenmaps in *SI Appendix*, Fig. S4 *B* and *C* also show that several sites in sector 3 are collectively (i.e., at higher order) negatively correlated to several others, which is not true for other sectors. Note that we do not identify important negative pair correlations by screening all pair correlations for those that exceed a cutoff (17). Rather, we first identified groups of sites (sectors) that are significantly collectively coupled (because RMT theorems help to eliminate noise-induced

correlations). We then examined the signs of multibody correlations only within these meaningful collectively coupled sectors.

Our results suggest that sector 3 is the most immunologically vulnerable multidimensionally constrained region in Gag, because multiple mutations are most constrained in this sector. Because sequence analysis methods are not exact, we tested this prediction, and its consequences for HIV infection and vaccination, against existing and new experimental data.

Sector 3 Is Immunologically Vulnerable Because of the Importance of Assembling Multiprotein Structures Critical for Viral Capsid Formation. In sector 3, 52 of the 57 sites are contained in the p24 protein but their locations within an individual p24 protein appear random (Fig. 2*A*; individual amino acids in *SI Appendix*, Table S1). However, hexamers of the p24 protein form the viral capsid (18), and superimposing sector 3 sites on the structure of the p24 hexamer (19) shows that the preponderance of these sites (~68%) is at interfaces between p24 proteins in a hexamer or at interfaces between the hexamers that form the capsid (Fig. 2*B*). Coevolution of this group of sites originates from constraints important for assembly of functional multiprotein structures, highlighting the importance of determining collective correlations.

The structural locations of sector 3 sites suggest the origin of negative correlations. Whereas one mutation in interface residues may still allow formation of p24 hexamers and assembly of the hexamers to form the viral capsid, multiple simultaneous mutations would likely destabilize these protein-protein interfaces. Fitness cost predictions of multiple mutations in sector 3, and comparisons with available data (20), are included in *SI Appendix*, Table S9–S11.

Sector 1 sites also reflect structural constraints associated with supramolecular assembly because they largely comprise the core of the hexamer (Fig. 2*C*). However, our analyses suggest that sector 1 is not as immunologically vulnerable as sector 3. This is because of the following:

i) The ratio of negative to positive correlations for pairs of mutations is a factor of 3 greater for sector 3.
ii) At the level of three-site correlations, the ratio of negative to positive correlations is more than threefold greater for sector 3.
iii) Sector 3 is characterized by collective higher order negative correlations between sites that are absent in sector 1 (*SI Appendix*, Fig. S4).

These results may be because multiple mutations in sector 3 residues result in disruption of intra- and interhexamer interfaces, whereas mutations in the core of the hexamer are less likely to disrupt virion assembly (21).

Elite Controllers of HIV Preferentially Target Multiple Sites in Sector 3. Our results suggest that applying CTL pressure at multiple sites contained in sector 3 is likely to facilitate durable control of viral load to low levels; multiple mutations that allow immune escape are more likely to hurt viral fitness significantly, because multidimensional escape pathways are restricted. Can HLA (human MHC class I) molecules present peptides containing multiple sites within the identified vulnerable region? Because elite controllers achieve viral control (16), we first analyzed the locations of sites contained in peptides targeted in individuals with HLA alleles associated with spontaneous HIV control (HLA-A*25, B*57, B*27, B*14, Cw*08) (22). Remarkably, the sector with the largest proportion of sites contained in the dominant peptides targeted by controllers is sector 3 ($P = 3{\cdot}10^{-7}$; Fig. 3*A* and *SI Appendix*, Table S3). CTL pressure imposed by individuals with these HLA alleles, however, is not the driver of this collectively coevolving group of sites for reasons that include:

i) There are clear structural explanations (Fig. 2) for the collective correlations that characterize sector 3 sites, and structural features are independent of immune pressure.
ii) Negative correlations cannot be induced by immune pressure, because mutations induced in persons with the same HLA would be more likely to arise than if they occurred independently.
iii) As noted, a quasi-sector describes weak correlations associated with HLA-associated mutations (*SI Appendix*, Fig. S5), and these sites are not a part of sector 3.

Thus, the sector we identify to be most vulnerable by analyses of viral sequences and supramolecular structures is the most targeted by controllers.

Examining targeted epitopes lends further insights. For example, eight amino acids in the dominant epitopes targeted by B57-restricted CTLs are in sector 3; this multiplicity is further augmented by the enhanced cross-reactivity of these CTLs (23). $B14^+$ individuals target three amino acids in this sector. Population studies indicate that HLA-$B57^+$/$B14^+$ persons are partic-

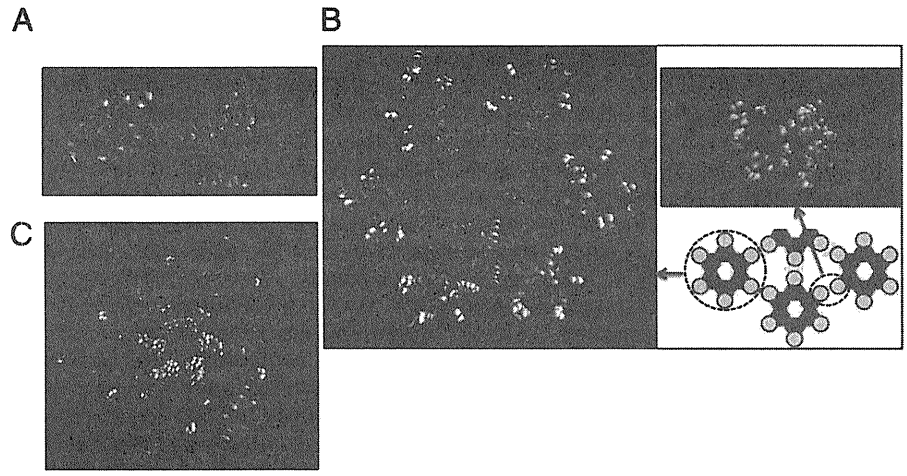


Fig. 2. Protein structures reveal the origin of collective correlations. (*A*) Sector 3 sites represented on the structure of the p24 monomer (PDB ID code 3GV2) are shown as purple spheres. (*B*) Sector 3 sites represented on the structure of the p24 hexamer (PDB code 3GV2) and the structure of the interface between hexamers (PDB code 2KOD). Sites at interfaces between two p24 proteins belonging to two adjacent hexamers are shown in green, and sites at interfaces between two p24 molecules within a hexamer are shown in red. The few remaining sites in sector 3 that are not part of these interfaces are shown in purple. (*C*) Sector 1 sites are shown in cyan on the structure of the p24 hexamer.

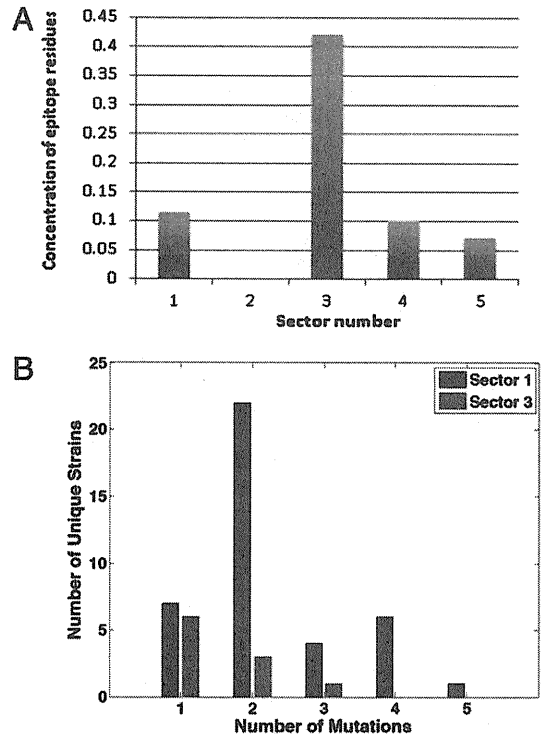


Fig. 3. Concentration of sites from the dominant epitopes presented by elite controllers in each sector and comparison of mutation patterns in sectors 1 and 3 in viruses derived from elite controllers. (*A*) Epitope is defined as dominant if it is the most targeted (in Gag) by individuals with the corresponding HLA allele. Concentration is the number of epitope sites in the sector divided by the total number of residues in the sector. The *P* value of association with each sector is in *SI Appendix* 10. Only sector 3 is significantly enriched with sites from the dominant epitopes of HLA molecules associated with control. Sectors 4 and 5 are not targeted at multiple points because they contain only 10 and 14 sites, respectively. Sectors 1 and 3 are targeted at multiple points. (*B*) Comparison of the number of unique viral strains obtained from a cohort of elite controllers that contains different numbers of mutations in sectors 1 and 3.

ularly efficient controllers of HIV infection. Our results suggest that this is because viral strains that can escape the multiple points of immune pressure in this sector are likely to have very low fitness. Thus, point mutants with lower viral fitness that only partially escape immune pressure or strains with multiple mutations that are not negatively correlated are likely to be the only options for the virus. This is consistent with the observed lower viral load and lower average viral fitness in controllers (16). Note also that CTLs restricted by HLA molecules associated with progression do not target sector 3 sites significantly ($P = 0.37$; *SI Appendix*, Table S3).

Virus Sequences from Patients Show That Multiple Mutations in Sector 3 Are Unlikely. Roughly 10% of sites in sectors 1, 4, and 5 are targeted by controllers, and sector 2 is not targeted (Fig. 3*A*). However, because sectors 4 and 5 are composed of only 10 and 14 sites, respectively, only 1 site (and epitope) is targeted by controllers in these sectors. Because sector 1 comprises 79 sites, many sites in sector 1 are targeted by controllers. If two regions of the proteome are equally targeted, all things being equal, the number and types of mutations observed in these regions should be similar. Sectors 1 and 3, both in Gag, exhibit similar levels of single site conservation, but we find that sector 3 is more multidimensionally constrained. Thus, we predict that even though sector 3 is targeted more than sector 1, fewer HIV strains with multiple mutations in sector 3 sites would be viable compared with those with similar mutations in sector 1.

To test this prediction, we sequenced viruses obtained from elite controllers because they target both sectors 1 and 3. We obtained 72 sequences from plasma and 103 including both plasma and peripheral blood mononuclear cells (PBMCs). The qualitative results obtained from analyzing either set of sequences are the same (Fig. 3*B* and *SI Appendix* 18). As per our predictions, the frequency of single mutations in sectors 1 and 3 is similar but multiple mutations are significantly rarer in sites that comprise sector 3. Our results suggest that this is because viral strains that can escape multiple points of immune pressure in sector 3 are more likely to be associated with negative correlations, and hence have particularly low replicative fitness because of defective capsid assembly. Thus, they are unlikely to be viable, and we do not observe them. These results show that the calculated differences between collective correlations among sites that comprise sector 3 and other sectors (Fig. 1) result in qualitative differences in viral mutations observed in humans that target sector 3 sites. Note also that almost all the few viral strains that we observe with multiple mutations in sector 3 are associated with positive correlations (*SI Appendix*, Fig. S13). Although only positive correlations can correspond to compensatory mutations, positive correlations can also correspond to multiple mutants that are only moderately less fit compared with single mutants, unlike negative correlations, which are only associated with multiple mutants that are very deleterious.

Our results strongly support our hypothesis that applying CTL pressure at multiple points in a multidimensionally constrained group of sites can trap the virus, because we find that controllers target multiple sites in sector 3, but strains with multiple mutations in sector 3 are not viable. Thus, viable strains are still likely to be subject to immune pressure.

Immunogens That May Induce CTL Responses in a Population That Hurt HIV. Can such potent responses be induced by vaccination in persons who are not blessed with HLA alleles that mount the correct dominant responses? Individuals with other HLA alleles present subdominant peptides containing sites in the vulnerable regions we have identified (using the Epitope Tables of the Los Alamos HIV Molecular Immunology Database, http://www.hiv.lanl.gov/content/immunology/tables/tables.html). This suggests the following strategy for design of immunogens that could induce memory CTL responses targeting the most vulnerable regions of HIV in a population: Select protein segments that simultaneously maximize sites in sector 3 (and some sites in sector 1), which are characterized by negative correlations, and those contained in subdominant and dominant epitopes presented by HLA molecules that span a broad section of a population. Such immunogens should elicit memory CTLs that only target the multidimensionally conserved regions of HIV, because epitopes that are dominantly targeted ineffectually are excluded. During natural infection, these memory CTLs are expected to mount robust responses that hit HIV where it hurts early, before naive CTLs mount ineffective dominant responses. Mounting the right responses early is important for HIV infections (2), because there is a short window of time before the immune system is compromised.

To illustrate this strategy, we considered a target population, white Americans with the 25 most frequent haplotypes (~39% of this population) (24). We create all groups of 10 known epitopes (as an example) presented by HLA-A/B molecules that comprise these haplotypes and select the top 10 groups according to the two criteria noted above (*SI Appendix* 12). One top-scoring immunogen (p24 residues 160–188 and 240–277) contains at least 1 or 2 targeted epitopes in 97% and 56% of the target population, respectively (a lower bound, because many epitopes are unknown). Previous empirical data show that controllers target p24 sites 240–272 (25). Sites 160–188 contain only 1 epitope tar-

geted by known controllers but represent an equally vulnerable region that provides broad coverage.

Discussion

Immunogens designed in this way to include only the multidimensionally constrained regions of the HIV proteome, which also contain epitopes presented by diverse HLAs in a population, are candidates for peptide vaccines using synthetic vectors (26), or by linking each segment together, they can be analogs of mosaic immunogens (27) delivered by traditional vectors. Such an immunogen would target the most vulnerable regions of HIV rather than the whole proteome. Targeting the latter is more likely to elicit responses from which HIV can escape via mutations (28), while hindering a focused response directed at regions of immunological vulnerability we have identified. Our goal of identifying such regions, and the methods we have used toward this end, are different from previous efforts to study coevolution of HIV proteins (e.g., 14, 29–31). Further analysis following the logic described by us should reveal additional regions of HIV proteins that are multidimensionally constrained, thereby enhancing the list of regions that are candidates for inclusion in a potent vaccine. A practical vaccine may also require inclusion of CD4 epitopes and flanking residues needed for antigen processing which do not contain protein segments that are likely to elicit ineffectual memory CTL responses. In vitro experiments and studies with animal models are required to develop this new concept, which might also be applied to design efficacious immunogens against other viruses. Our results also suggest the design of new small-molecule inhibitors of HIV replication.

Methods

HIV Sequences and Similarity Analysis. Multiple sequence alignments of nucleotide sequences of HIV-1 Gag, RT, and Nef were downloaded from the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov/) (clade B, one sequence per patient). We checked the phylogenetic homogeneity of each set by performing a principal component analysis of the similarity matrix of sequences (*SI Appendix* 5).

Identification of HIV Sectors. The correlation matrix for a given set of sequences is cleaned from phylogeny and noise using RMT (as described in the main text and *SI Appendix* 1–4). The definition of sectors follows the strategy proposed by Halabi et al. (12), grouping positions with a particular spectral signature into a sector (*SI Appendix* 6).

Targeting by Elite Controllers. We defined a set of epitopes dominantly targeted by HLA alleles associated with control (22), using a published dataset of the frequency of recognition of epitopes in an HLA-specific population (32).

Sequencing Viruses from Elite Controllers. HIV-1 Gag was amplified using nested RT-PCR from previously frozen PBMCs or plasma from elite controllers as previously described (33). PCR products were purified, and cycle-sequencing reactions were performed using 60 HIV-1–specific sequencing primers. Population sequences were obtained using an ABI 3730 PRISM (Applied Biosystems) automated sequencer. Gag mutations in the sequences derived from elite controllers were defined with reference to the clade B consensus sequence.

ACKNOWLEDGMENTS. We thank D. Barouch and H. Eisen for valuable discussions. Financial support was provided by the Ragon Institute, a National Institutes of Health Director's Pioneer Award (to A.K.C.), National Institutes of Health Grants RO130914 (to B.D.W.) and PO1 AI074415 (to M. Altfeld and T.M.A.), The Howard Hughes Medical Institute (B.D.W.) and the Mark and Lisa Schwartz Foundation (B.D.W.) This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract HHSN261200800001E.

1. Walker BD, Burton DR (2008) Toward an AIDS vaccine. *Science* 320:760–764.
2. McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF (2010) The immune response during acute HIV-1 infection: Clues for vaccine development. *Nat Rev Immunol* 10:11–23.
3. Amanna IJ, Slifka MK (2011) Contributions of humoral and cellular immunity to vaccine-induced protection in humans. *Virology* 411:206–215.
4. Barouch DH, Korber B (2010) HIV-1 vaccine development after STEP. *Annu Rev Med* 61:153–167.
5. Hansen SG, et al. (2009) Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. *Nat Med* 15:293–299.
6. Hartl DL, Clark AG (2007) *Principles of Population Genetics* (Sinauer, Sunderland, MA), 4th Ed.
7. Altfeld M, Allen TM (2006) Hitting HIV where it hurts: An alternative approach to HIV vaccine design. *Trends Immunol* 27:504–510.
8. Goulder PJ, Watkins DI (2004) HIV and SIV CTL escape: Implications for vaccine design. *Nat Rev Immunol* 4:630–640.
9. Wigner EP (1967) Random matrices in physics. *SIAM Rev* 9:1–23.
10. Plerou V, et al. (2002) Random matrix approach to cross correlations in financial data. *Phys Rev E Stat Nonlin Soft Matter Phys* 65:066126.
11. Laloux L, Cizeau P, Bouchaud JP, Potters M (1999) Noise dressing of financial correlation matrices. *Phys Rev Lett* 83:1467–1470.
12. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: Evolutionary units of three-dimensional structure. *Cell* 138:774–786.
13. Bhattacharya T, et al. (2007) Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* 315:1583–1586.
14. Brumme ZL, et al. (2009) HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLoS ONE* 4:e6687.
15. Kiepiela P, et al. (2007) CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* 13:46–53.
16. Miura T, et al. (2009) HLA-associated viral mutations are common in human immunodeficiency virus type 1 elite controllers. *J Virol* 83:3407–3412.
17. Hoffman NG, Schiffer CA, Swanstrom R (2003) Covariation of amino acid positions in HIV-1 protease. *Virology* 314:536–548.
18. Ganser-Pornillos BK, Yeager M, Sundquist WI (2008) The structural biology of HIV assembly. *Curr Opin Struct Biol* 18:203–217.
19. Pornillos O, et al. (2009) X-ray structures of the hexameric building block of the HIV capsid. *Cell* 137:1282–1292.
20. Troyer RM, et al. (2009) Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. *PLoS Pathog* 5:e1000365.
21. Adamson CS, Jones IM (2004) The molecular basis of HIV capsid assembly—Five years of progress. *Rev Med Virol* 14:107–121.
22. Pereyra F, et al.; International HIV Controllers Study (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330: 1551–1557.
23. Kosmrlj A, et al. (2010) Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* 465:350–354.
24. Maiers M, Gragert L, Klitz W (2007) High-resolution HLA alleles and haplotypes in the United States population. *Hum Immunol* 68:779–788.
25. Streeck H, et al. (2007) Recognition of a defined region within p24 gag by CD8+ T cells during primary human immunodeficiency virus type 1 infection in individuals expressing protective HLA class I alleles. *J Virol* 81:7725–7731.
26. Melief CJ, van der Burg SH (2008) Immunotherapy of established (pre)malignant disease by synthetic long peptide vaccines. *Nat Rev Cancer* 8:351–360.
27. Barouch DH, et al. (2010) Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nat Med* 16:319–323.
28. Altfeld M, Goulder PJ (2011) The STEP study provides a hint that vaccine induction of the right CD8+ T cell responses can facilitate immune control of HIV. *J Infect Dis* 203: 753–755.
29. Liu Y, Eyal E, Bahar I (2008) Analysis of correlated mutations in HIV-1 protease using spectral clustering. *Bioinformatics* 24:1243–1250.
30. Fares MA, Travers SAA (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* 173:9–23.
31. Bickel PJ, et al. (1996) Covariability of V3 loop amino acids. *AIDS Research and Human Retroviruses* 12:1401–1411.
32. Streeck H, et al. (2009) Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. *J Virol* 83:7641–7648.
33. Miura T, et al. (2008) Genetic characterization of human immunodeficiency virus type 1 in elite controllers: Lack of gross genetic defects or common amino acid changes. *J Virol* 82:8422–8430.

AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 27, Number 00, 2011

DOI: 10.1089/aid.2011.0073

Imbalanced Production of Cytokines by T Cells Associates with the Activation/Exhaustion Status of Memory T Cells in Chronic HIV Type 1 Infection

Kaori Nakayama,[1] Hitomi Nakamura,[2] Michiko Koga,[2] Tomohiko Koibuchi,[2] Takeshi Fujii,[2] Toshiyuki Miura,[1] Aikichi Iwamoto,[1,2] and Ai Kawana-Tachikawa[1]

Abstract

Chronic HIV-1 infection is characterized by immune cell dysfunctions driven by chronic immune activation. Plasma HIV-1 viral load (VL) is closely correlated with disease progression and the level of immune activation. However, the mechanism by which the persistent presence of HIV-1 damages immune cells is still not fully understood. To evaluate how HIV-1 affects disruption of T cell-mediated immune responses during chronic HIV-1 infection we determined the functional profiles of T cells from subjects with chronic HIV-1 infection. We measured the capacity of peripheral blood mononuclear cells (PBMCs) to produce 25 specific cytokines in response to nonspecific T cell stimulation, and found that the capacity to produce Th-1-related cytokines (MIP-1α, MIP-1β, RANTES, IFN-γ, and MIG), sIL-2R, and IL-17, but not Th-2-related cytokines, was inversely correlated with plasma VL. The capacities to produce these cytokines were interrelated; notably, IL-17 production had a strong direct correlation with production of MIP-1α, MIP-1β, RANTES, and IFN-γ. In both $CD4^+$ and $CD8^+$ T cells, dysfunctional production of cytokines was associated with T cell activation (CD38 expression) and exhaustion (PD-1 and/or CTLA-4 expression) status of memory subsets. Although the capacity to produce these cytokines was recovered soon after multiple $\log_{10}$ reduction of plasma viral levels by antiretroviral therapy, memory $CD8^+$ T cells remained activated and exhausted after prolonged virus suppression. Our data suggest that HIV-1 levels directly affect the ability of memory T cells to produce specifically Th1- and Th17-related cytokines during chronic HIV-1 infection.

Introduction

PLASMA VIRAL LOAD (VL) and CD4-positive T cell count are two surrogate markers of HIV-1 disease progression.[1] Throughout the course of infection both innate and adaptive immune systems are highly activated, and disease progression is strongly correlated with immune activation status.[2,3] Notably, immune activation is observed in both HIV-1 and pathogenic SIV infection, but not in nonpathogenic SIV infection in a natural host.[4,5] Moreover, studies have shown that T cells in patients with high VL and progressive disease are less functional, have less proliferative capacity, and are more exhausted than T cells in patients with low VL and slow disease progression.[6–12] In those patients, exhaustion is seen not only in HIV-1-specific T cells, but also in nonspecific T cells.[11,13] These data suggest that immune cells have lost their original function due to persistent hyperactivation, which depends on VL, during chronic HIV-1 infection.

The immune system is highly coordinated: the cytokine network regulates interactions between cells, and cytokine balance dictates how the immune system responds. Cytokine production determines the specific helper functions of $CD4^+$ T cells and allows balance in immune responses *in vivo*.[14–16] A possible explanation of the impaired immune response in chronic HIV-1 infection is that the ability of T cells to balance cytokine production has been altered, just as alteration of balance between Th1- and Th2-type immune response affects the clinical course of certain infectious diseases and autoimmune syndromes.[17,18]

To evaluate T cell impairment resulting from persistent immune activation during chronic HIV-1 infection, we compared the cytokine expression spectra of peripheral blood mononuclear cells (PBMCs) in response to nonspecific T cell stimulation in treatment-naive HIV-1-infected subjects with low or high VL. We also examined the differentiation states and activation levels of $CD4^+$ and $CD8^+$ T cells from

[1]Division of Infectious Diseases, Advanced Clinical Research Center, and [2]Department of Infectious Diseases and Applied Immunology, Research Hospital, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

HIV-1-infected subjects to elucidate relationships between cytokine expression capacity and T cell phenotypic status.

Materials and Methods

Study design

HIV-1-infected individuals who were under medical supervision at our clinic were asked to provide blood samples for this study. Blood samples were taken from selected patients in the chronic phase of HIV-1 infection, with CD4 counts >200 cells/ml. We requested blood samples from antiretroviral therapy (ART)-naive patients with either low plasma viral load (VL) values (<5000 copies/ml; LVL group) or high VL values (>25,000 copies/ml; HVL group), and from treatment-experienced patients who had received ART >2 years (Tx group). Blood samples were also obtained from a small number of HIV-1-infected patients who had first initiated ART within the previous 1–2 months. As controls, blood samples were obtained from HIV-1-seronegative individuals (healthy controls; HC).

All participants gave written informed consent, and the study was approved by the institutional review boards of the Institute of the Medical Science of the University of Tokyo (No. 11-2-0329 and 20-47-210521).

PBMC cultures and PHA stimulation

PBMCs were isolated from heparinized whole blood by Ficoll-Paque PLUS density gradients (GE Healthcare, Piscataway, NJ) and cryopreserved in liquid nitrogen until use. The frozen cells were thawed 1 day before stimulation and cultured in R10 medium [RPMI 1640 medium (Sigma, St. Louis, MO) supplemented with 10% heat-inactivated fetal calf serum (FCS; Sigma), 100 U penicillin/ml, 100 μg/ml streptomycin (Sigma), 2 mmol/liter L-glutamine (Sigma), and 10 mmol/liter 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES; Sigma)] at 37°C, 5% CO_2. The following day 5×10^5 cells/well were cultured in 250 μl/well of R10 medium with or without 2 μg/ml phytohemagglutinin L (PHA; Roche Applied Science, Mannheim, Germany). Culture supernatants were harvested after 48 h and stored at −80°C until use for multiple cytokine assays.

Quantification of cytokines

The human cytokine 25-plex antibody kit (Invitrogen Corporation, Carlsbad, CA) was used to measure the levels of 25 cytokines in culture supernatants: interleukin (IL)-1 receptor antagonist protein (IL-1RA), IL-1β, IL-2, soluble IL-2R (sIL-2R), IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12p40/70, IL-13, IL-15, IL-17, eotaxin, interferon gamma (IFN-γ)-induced protein 10 kDa (IP-10), monocyte chemoattractant protein-1 (MCP-1), monokine induced by IFN-γ (MIG), macrophage inflammatory protein 1alpha (MIP-1α), MIP-1β, regulated on activation normal T cell expressed and secreted (RANTES), tumor necrosis factor-α (TNF-α), granulocyte-macrophage colony-stimulating factor (GM-CSF), IFN-α, and IFN-γ. The detection limits for the cytokines measured by the kit were as follows: IL-5, IL-6, IL-8, 3 pg/ml; MIG, 4 pg/ml; IL-4, IL-10, IFN-γ, IP-10, eotaxin, 5 pg/ml; IL-2, 6 pg/ml; IL-7, IL-13, IL-15, IL-17, TNF-α, MIP-1α, MIP-1β, MCP-1, 10 pg/ml; IL-1β, IL-12p40/70, IFN-α, GM-CSF, RANTES, 15 pg/ml; IL-1RA, sIL-2R, 30 pg/ml. As the amounts of IL-6, IL-8, TNF-α, MIP-1α, MIP-1β, IP-10, MIG, and MCP-1 produced from PHA-stimulated PBMCs were beyond the range of the assay, we diluted the samples 10-fold prior to measurement of these cytokines. However, IL-8 levels were out of range in most patient samples and could not be measured accurately.

Samples were loaded onto the Luminex100 system (Luminex Corporation, Austin, YX), and samples were quantified by analysis of the median fluorescence intensity of the beads using MasterPlex QT version 2.5 (Luminex Corporation). The assays were performed according to the manufacturer's instructions, and all samples were run in duplicate.

Identification of cytokine-producing cells in PBMCs

$CD14^+$ cells (monocytes), $CD8^+$ T cells, $CD4^+$ T cells, and $CD56^+CD16^+$ (NK) cells were isolated sequentially from PBMCs of each healthy subject. Magnetic cell separation (MACS) selection was performed using anti-CD14, anti-CD8, and anti-CD4 antibody-conjugated microbeads or using the $CD56^+CD16^+$ NK cell isolation kit (Miltenyi Biotic, Bergisch Gladbach, Germany). The purity of each cell fraction was >95% as determined by flow cytometry.

Fractionated cells were cultured separately or were co-cultured in the presence of 2 μg/ml PHA at 37°C, 5% CO_2, for 48 h. Levels of MIP-1α, MIP-1β, RANTES, IL-2R, IFN-γ, and IL-17 in culture supernatants were measured with DuoSet ELISA Development Systems (R&D Systems). The absolute numbers of each cell fraction used in the experiments were calculated from the average proportion of each subset in PBMCs.

Antibodies

The fluorochrome-conjugated monoclonal antibodies (mAb) used in the study were as follows: fluorescein isothiocyanate (FITC)-labeled anti-MIP-1α, anti-MIP-1β, and anti-RANTES (R&D Systems, Minneapolis, MN); FITC-labeled anti-PD-1 and anti-Ki67, phycoerythrin (PE)-labeled anti-Bcl-2, peridinin cholorophyll protein/cyanin5.5 (PerCP Cy5.5)-labeled anti-CD38 and anti-CD3, PE Cy7-labeled anti-CCR7, allophycocyanin (APC)-labeled anti-CD45RA, and pacific blue-labeled anti-CD4 (BD Biosciences, San Jose, CA); APC AlexaFluor 750-labeled anti-CD4, pacific blue-labeled anti-IFN-γ, and AlexaFluor 647-labeled anti-IL-17A (eBioscience, San Diego, CA); PE-labeled anti-IL-4 (Becton Dickinson, Franklin Lakes, NJ); APC Cy7-labeled anti-CD3 (BioLegend, San Diego, CA); and Pacific Orange-labeled anti-CD8 (Invitrogen).

Surface phenotypic and intracellular cytokine staining

For intracellular cytokine staining, cryopreserved PBMCs were thawed and cultured in R10 overnight. The following day cells were stimulated with phorbol ester (PMA)/calcium ionophore (ionomycin) in the presence of Golgi inhibitor (brefeldin A) for 5 h. Cells were stained with a panel of fluorescently labeled antibodies against cell-surface markers. For detection of dead cells, the cells were also stained with 5 μg/ml ethidium monoazide bromide (EMA; Sigma). Cells were washed twice and exposed to fluorescent light for 10 min on ice to allow the EMA to bind to DNA in dead cells. Cells were then fixed in 2% paraformaldehyde and permeabilized in BD FACS Permeabilizing Solution 2 (BD Biosciences) prior to antibody staining for intracellular molecules.

Dead or dying cells were detected by surface phenotypic staining with propidium iodide (PI; Sigma).

Flow cytometric analysis

Samples were analyzed on a FACSAria multilaser cytometer (Beckton Dickinson) running FACSDiva software, with collections of 60,000–100,000 lymphocyte-gated events. Data were analyzed with FlowJo software (Tree Star, Ashland, OR).

Statistical analysis

GraphPad Prism5 software (San Diego, CA) was used for all statistical analysis. Differences between groups were tested for statistical significance using the nonparametric Mann–Whitney *U* test. Since previous studies revealed that production of multiple cytokines by HIV-specific T cells was limited in progressors compared to nonprogressors,[6,19] the production levels of cytokines were expected to differ among LVL, HVL, and healthy control subjects. For this reason, we did not consider multiple comparison correction for Mann–Whitney *U* tests to avoid false-negative results. Correlation analysis was performed using Spearman's rank correlation. The level of significance for all analyses was set at $p < 0.05$.

Results

Study population

Most analyses were performed using blood samples collected from 35 HIV-1-infected, ART-naïve patients, 15 HIV-1-infected, treatment-experienced patients, and 16 HIV-1-seronegative individuals. Demographic characteristics of these 50 HIV-1-infected patients are presented in Table 1. The 35 HIV-1-infected, ART-naive patients included 19 patients with low VL (LVL group; median VL: 1200, range: 53 to 3600) and 16 patients with high VL (HVL group; median VL: 62,000; range: 25,000 to 500,000). The median CD4 counts in the LVL and HVL groups were 449 (range: 316 to 749) and 407 (range: 228 to 520), respectively; the difference was not statistically significant. The groups also showed no significant difference in age, another factor that influences immune status.

The 15 HIV-1-infected individuals recruited into the study to represent treatment-experienced patients had received ART and successfully controlled their disease over a long period of time (median: 66 months; range: 22 to 149 months). To examine the impact of actively decreasing VL on the functional profile of PBMCs, blood samples were also collected from six HIV-1-infected patients who had initiated treatment only in the previous 1–2 months.

Cytokine production in PHA-stimulated PBMCs

Cytokine measurements from cells cultured for 48 h in an unstimulated state were at the limit of detection (data not shown). We initially compared anti-CD3-antibody and PHA as a nonspecific stimulus of PBMCs to induce cytokine production, and found that the production levels of most cytokines were much higher in PHA-stimulated PBMCs than in anti-CD3-antibody-stimulated PBMCs (data not shown). When cells were stimulated with PHA and cultured for 48 h, production of most cytokines increased dramatically (Fig. 1A). There were no significant differences between any groups in IL-2, IL-13, IL-15, IL-1β, IFN-α, TNF-α, eotaxin, or IP-10 production (data not shown).

Cytokine production in PBMCs from treatment-naive HIV-1 subjects was compared to cytokine production in PBMCs from healthy control subjects. Median levels of many cytokines in the HVL group were significantly different from those in the healthy control group: MIP-1α [6.33 (range 0.99–21.01) vs. 16.92 (10.36–23.87) ng/ml; $p = 0.0005$], MIP-1β [8.51 (1.37–26.42) vs. 21.44 (10.26–34.11) ng/ml; $p = 0.0036$], IFN-γ [1.50 (0.30–5.75) vs. 2.64 (0.79–5.78) ng/ml; $p = 0.0402$], IL-7 [<0.01 (<0.01–0.67) vs. 0.04 (<0.01–0.07) ng/ml; $p = 0.0077$], IL-1Ra [18.93 (0.59–27.61) vs. 1.50 (0.55–14.95) ng/ml; $p = 0.0184$], IL-6 [0.63 (0.11–12.23) vs. 1.77 (0.68–4.93) ng/ml; $p = 0.0254$], and IL-10 [0.08 (<0.005–0.40) vs. 0.55 (0.13–0.83) ng/ml; $p = 0.0031$]. In contrast, significant differences between the LVL group and the healthy control group were seen only in levels of IL-10 [0.12 (<0.005–0.72) vs. 0.55 (0.13–0.83) ng/ml; $p = 0.0050$] and IL-1Ra [17.05 (0.49–28.31) vs. 1.50 (0.55–14.95) ng/ml; $p = 0.0282$] (Fig. 1A). These data suggest that although PBMCs from HVL subjects are abnormal in some way, PBMCs from LVL subjects are almost normal in terms of cytokine production.

As shown in Fig. 1A, mean cytokine levels were significantly lower in HVL subjects compared to LVL subjects, as follows: MIP-1α [6.33 (0.99-21.01) vs. 14.36 (2.29–29.16) ng/ml; $p = 0.0077$], MIP-1β [8.51 (1.37–26.42) vs. 20.14 (4.31–48.75) ng/ml; $p = 0.0034$], RANTES [2.01 (<0.015–4.57) vs. 3.40 (1.33–6.90) ng/ml; $p = 0.0014$], sIL-2R [2.30 (0.02–4.96) vs. 3.72 (1.72–7.38) ng/ml; $p = 0.0136$], IL-17 [0.04 (<0.01–0.12) vs. 0.08 (<0.01–0.17) pg/ml; $p = 0.0256$], and IL-7 [<0.01 (<0.01–0.67) vs. 0.05 (<0.01–1.05) pg/ml; $p = 0.0029$]. Notably, there was an inverse correlation between VL and production of these cytokines, and of IFN-γ (Fig. 1B). No relationship was observed between cytokine levels and CD4 cell count (data not shown). These data suggest that VL directly affects the capacity of PBMCs to produce certain cytokines during chronic infection.

Th1- and Th17-type T cells have impaired cytokine production in HVL subjects

Although PHA is considered a T cell mitogen, other cell populations also produce cytokines in response to PHA stimulation.[20–22] The next step was to determine which cells were responsible for the alterations in cytokine production observed under our experimental conditions. The cytokines whose production was inversely correlated with VL can be produced by several cell populations in PBMCs. To identify the major cell population producing these cytokines, we fractionated PBMCs in healthy donors by positive selection and determined the cell population producing these cytokines. $CD4^+$ T cells, $CD8^+$ T cells, monocytes ($CD14^+$), and NK cells ($CD56^+CD16^+$) were isolated from PBMCs, cultured separately or cocultured, and stimulated with PHA. We then measured levels of MIP-1α, MIP-1β, RANTES, IFN-γ, sIL-2R, and IL-17. Little or no production of these cytokines was detected in any of the single cell fractions (Fig. 2A). Production of cytokines MIP-1α, MIP-1β, RANTES, IFN-γ, and sIL-2R was observed in cocultures of $CD4^+$ and $CD14^+$ cells and in cocultures of $CD8^+$ and $CD14^+$ cells (Fig. 2A, and data not shown). IL-17 production was detected only in cocultures of $CD4^+$ and $CD14^+$ cells (Fig. 2A). As T cell stimulation by