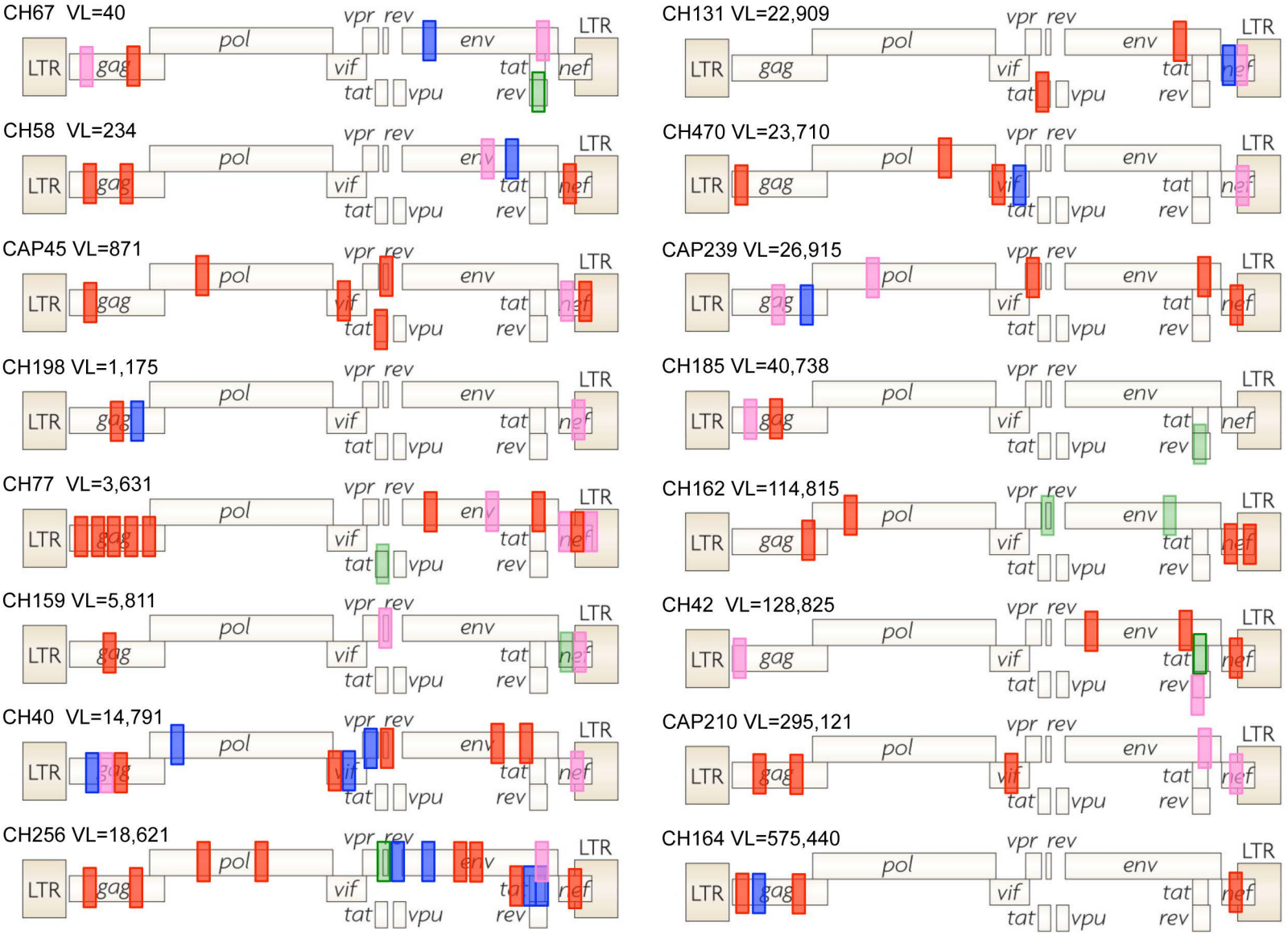
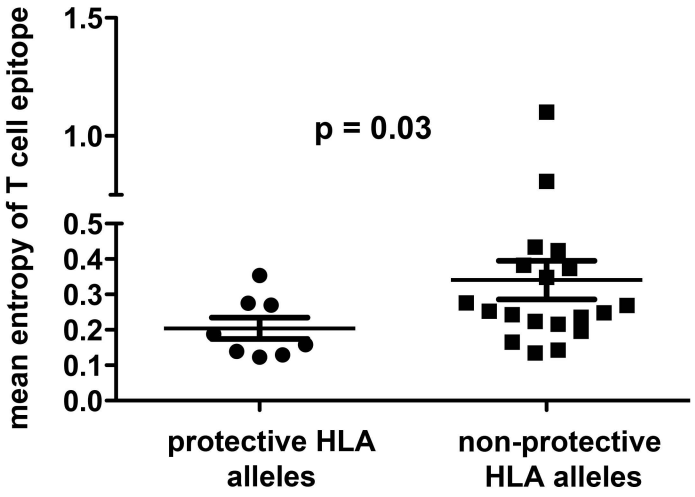
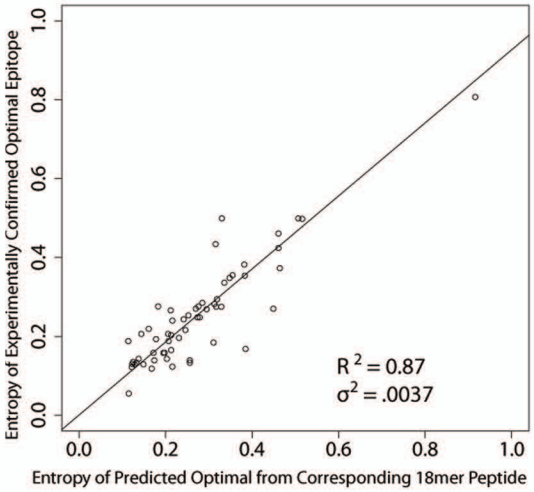


Supplementary Figure 1. The majority of rapid and early non-synonymous (NS) mutations across the HIV-1 proteome are selected by T cells. A schematic of the HIV-1 proteome is displayed for ART naive individuals with virus setpoint shown, units = copies/ml. SGA sequencing was performed in acute infection to deduce each individual's T/F virus. Serial SGA sequencing was then performed on subsequent samples over the next 6-24 months to identify ns changes that emerged across the proteomes as infection progressed. T cell mapping using ex-vivo IFN-γ ELISpots and autologous peptides was performed on PBMCs collected from HIV-1 acutely infected individuals within 50 days of Fiebig stage I/II to define which mutational clusters represented virus escape from T cell selection. T cell epitopes, putative reversions and unknown selections are represented by pink, green and blue stripes, respectively. Reversions were defined as a change back to the clade consensus sequence (>50% all SGA sequences) in the absence of a detectable immune response. Non-classified mutations could have resulted from other immune selection pressures such as NK cells or antibodies. Equally, some non-classified mutational stripes may have emerged in linkage with true immune escape mutations. The epitopes targeted by T cell responses that were detected during the first 50 days from Fiebig I/II but remained invariant in this 50 day window are represented by red stripes. NS mutations are displayed only if >or= 50% change occurred in SGA sequences compared to the T/F sequence in the first 50 days of Fiebig I-II.



Supplementary Figure 2. HIV-1 specific T cell epitopes restricted by HLA alleles associated with lower plasma viral loads and/or delayed progression to disease have lower mean entropies. Entropy scores were calculated for all experimentally confirmed optimal epitopes detected in the first 50 days from Fiebig I-II (n=34). Data were divided into 2 groups: epitopes restricted by protective HLA alleles B*5701/03, B*5801, B*8101 and all other epitopes. Epitopes were considered only once reducing the overall epitopes analysed to n=27. Individual data points are plotted overlaid with mean+/- SEM. A Mann-Witney 1-tailed test were performed and $p < 0.05$ considered significant.



Supplementary Figure 3. Correlation between Shannon entropy calculations for experimentally confirmed T cell epitopes and predicted epitopes defined from mutation patterns. Entropy scores were calculated from experimentally confirmed optimal T cell epitopes (n=60) mapped in the first year following HIV-1 infection across the patient data set (data not shown). Note, that because of broadening of the HIV-1 specific T cell response over time, this was a much larger dataset than T cell responses mapped in the first 50 days from Fiebig I-II. Rules were defined to predict the location of 9mer T cell epitopes in reactive 18mer peptides. Predicted 9mers were centered on residues where ns changes from the T/F emerged over time and mean entropy calculated. Where no mutations emerged, the peptide was split into non-overlapping 9mers and the average of entropies for each 9mer calculated.

Entropy scores from experimentally confirmed and predicted optimals were then compared using linear regression to assess the degree to which entropy of the optimal epitope could be predicted for the corresponding 18mer peptide. Nearly 90% of the variation in the entropy of the optimal epitope was predictable for the corresponding 18mer and the relationship between the two was linear. The equation of the line of best fit for this comparison was used to compute an entropy score for reactive 18mer peptides (detected in the first 50 days from Fiebig I-II) for which the optimal epitope had not been defined experimentally, increasing the study 'n' to 59.