

**Supplemental Figure 1.** HIV-1 subtype A sequence coverage by variant inclusive peptide sets. A, Percent coverage of reported HIV-1 subtype A sequences in the Los Alamos National Library (LANL) HIV Sequence Database for the Gag p55 protein. The x-axis represents 491 sequential 10-mer windows and the variant inclusive peptide set coverage per window is plotted as the area under the curve in blue and represents the percent of coverage for all known sequence diversity. B, Percent coverage of reported HIV-1 subtype A sequences in LANL HIV Sequence Database for the gp41 protein. The x-axis represents 344 sequential 10-mer windows and the variant inclusive peptide set is plotted as the area under the curve in red. C, Percent coverage of reported HIV-1 subtype A sequences in LANL HIV Sequence Database for the Nef protein. The x-axis represents 196 sequential 10-mer windows and the variant inclusive peptide set is plotted as the area under the area under the curve in green.



**Supplemental Figure 2.** Variation in CD4 T cell IFN- $\gamma$  response to HIV-1 subtype A gag or CMVpp65 peptides. PBMC from HIV-1 subtype A infected individuals were thawed and assessed in a standard five-function polychromatic flow cytometry assay. PBMC were stimulated with peptides or in the absence of stimulation. A, Shows the variation of 3 CMV-responders in the right column with the corresponding unstimulated control in the left column. High, medium, and low IFN- $\gamma$  responses were presented in the top, middle, and bottom rows respectively. B, Shows the variation of 3 HIV-gag-responders in the right column with the corresponding unstimulated control in the left column. High, medium, and low IFN- $\gamma$  responses were presented in the top, middle, and bottom rows respectively. B, Shows the variation of a HIV-gag-responders in the right column with the corresponding unstimulated control in the left column. High, medium, and low IFN- $\gamma$  responses were presented in the top, middle, and bottom rows respectively.



**Supplemental Figure 3.** Memory phenotype definitions and trends in chronic HIV-1 subtype A infection. HIV-1 negative (n = 39) and HIV-1 positive (n = 48) individuals enrolled in the cohort, were selected to assess and compare the phenotype in chronic untreated infection compared to uninfected community matched controls. PBMC were thawed and stained to characterize phenotype (CD45RO, CCR7, and CD62L). A, Representative polychromatic histograms showing the CD45RO distribution on CD4 and CD8 T cells in HIV-1 uninfected participants. CD45RO positive and negative populations are further characterized based on CCR7 and CD62L expression presented in zebra plots. CD45RO<sup>+</sup>CCR7<sup>+</sup>CD62L<sup>+</sup> populations are defined as central memory (T<sub>CM</sub>). CD45RO<sup>+</sup>CCR7<sup>+</sup>CD62L<sup>+</sup> populations are defined as naïve (T<sub>Naïve</sub>). CD45RO<sup>-</sup>CCR7<sup>+</sup>CD62L<sup>+</sup> populations are defined as naïve (T<sub>Naïve</sub>). CD45RO<sup>-</sup>CCR7<sup>-</sup>CD62L<sup>-</sup> populations are defined as naïve (T<sub>Naïve</sub>).

different memory phenotypes in HIV-infected (gray box) and HIV-uninfected (white box) study participants. C, Box and whisker plots showing the median and 10th–90th percentiles of CD8 T cell absolute counts for different memory phenotypes in HIV-infected (gray box) and HIV-uninfected (white box) study participants. Statistically significant (p < 0.0001) differences are labeled as \*\*\*. CD8 T<sub>EM</sub> (P = 0.046, r = 0.261) cell numbers correlated directly to viral load whereas CD8 T<sub>CM</sub> (P = 0.003, r = 0.382) and CD8 T<sub>Naïve</sub> (P < 0.001, r = 0.565) correlated directly with CD4 T cell counts (data not shown). Memory CD4 T cell subset frequency correlated inversely to viral load and directly to CD4 absolute counts (data not shown).