

Table S1. HLA-I allele-associated HIV-1 polymorphisms not occurring within or near known CTL epitopes in chronically clade C-infected South Africans

| HIV-1 subregion | Common mutations ^a | Amino acid residue ^b | Specific HLA-I association ^c |
|-------------------------|-----------------------------------|--|--|
| Gag-p17 | RSILRGEKLDKWEKI | Gag 12 | A*74, P = 2.7×10^{-14} ; q = 0 |
| | LDKWEKIRLRPGGKK | Gag20 | A*74, P = 6.5×10^{-10} ; q = 0 |
| | YCVHEKIEVRDTKEA | Gag 93 | B*41, P = 9.4×10^{-5} ; q = 0.05 |
| Gag-p24 | NYPVQNQLQGQMVHQ | Gag 138 | B*5703, P = 3×10^{-3} ; q = 0.15 |
| | LNTMLNTVGGHQAAAM | Gag 191 | A*3601, P = 2.8×10^{-3} ; q = 0.15 |
| | GHQAAMQMLKDTINE | Gag 200 | A*01, P = 2.8×10^{-3} ; q = 0.19 |
| | TINEEAAEWDRLLHPV | Gag 211 | Cw*1601, P = 9.6×10^{-4} ; q = 0.09 |
| | INEEAAEWDRLLHPVH | Gag 212 | Cw*0501, P = 1.6×10^{-3} ; q = 0.11 |
| | EAAEWDRLLHPVHAGP | Gag 215 | B*13, Cw*02, P = 2.1×10^{-3} ; q = 0.14 |
| | EWDRLLHPVHAGPIAP | Gag 218 | A*6602, P = 5.9×10^{-4} ; q = 0.07 |
| | WDRLLHPVHAGPIAPG | Gag 219 | Cw0704, P = 3.5×10^{-4} ; q = 0.04 |
| | STLQEQIAWMTSNPP | Gag 248 | B*41, P = 1.6×10^{-3} ; q = 0.12 |
| | DIYKRWIILGLNKIV | Gag 267 | B*5703, P = 1.8×10^{-3} ; q = 0.11 |
| | SPVSILDIRQGPKEP | Gag 285 | A*6802, P = 1.4×10^{-3} ; q = 0.11 |
| | WMTDTLLVQANPDC | Gag 323 | A*2902, P = 3×10^{-3} ; q = 0.15 |
| | GPGHKARVLAEAMSQ | Gag 362 | A*6601, P = 1.1×10^{-3} ; q = 0.1 |
| Nef | SSIVGWPA VR ERIRR | Nef 15 | A*6801, P = 1.5×10^{-4} ; q = 0.07 |
| | VPLRPM TY KAAFDLS | Nef 81 | Cw*0702, P = 7×10^{-4} ; q = 0.18 |
| | LKEKGGLE I YSKKRQ | Nef 98 | A*2301, P = 5.6×10^{-4} ; q = 0.12 |
| | KLVPVDP RE VEEANK | Nef 151 | A*2301, P = 3.1×10^{-4} ; q = 0.13 |
| | D PRE VEEANKGENNC | Nef 156 | Cw*0802, P = 4.1×10^{-4} ; q = 0.13 |
| | PREVEE AN KGENNCL | Nef 157 | B*13, P = 1.4×10^{-3} ; q = 0.18 |
| | NNCLL HP MSQHG MED | Nef 168 | B*4201, P = 3.8×10^{-4} ; q = 0.09 |
| | HPMSQHG ME DEEREV | Nef 173 | B*4201, P = 7.1×10^{-4} ; q = 0.16 |
| | SQHG ME DEEREVLKW | Nef 176 | B*4403, P = 1.7×10^{-6} ; q = 0 |
| | DEEREVL K WKF D SSL | Nef 182 | Cw*0804, P = 5.8×10^{-4} ; q = 0.16 |
| | F D SSLARRHLARELH | Nef 192 | A*74, P = 4.7×10^{-5} ; q = 0.03 |
| RHLAREL HP EYKDC | Nef 199 | Cw*1601, P = 3.7×10^{-7} ; q = 0.09 | |

These data are taken from 82 HLA associations (q < 0.2) defined in a clade C South African cohort (unpublished data), in which analysis of gag sequences was performed using data from 672 study subjects and analysis of nef sequences was performed using data from 443 study subjects. The Los Alamos National Laboratory HIV Molecular Immunology Database (available at <http://www.hiv.lanl.gov/content/index>) was used to define known CTL epitopes.

^aHLA-associated mutations (bolded) are shown.

^bProtein and amino acid numbering of the bolded polymorphism are based on the HXB2 sequence

^cMethods for calculating p- and q- values have been previously described (Bhattacharya, T., M. Daniels, D. Heckerman, B. Foley, N. Frahm, C. Kadie, J. Carlson, K. Yusim, B. McMahon, B. Gaschen, et al. 2007. *Science*. 315:1583–1586). The HLA-I restriction of the Gag 215 mutation could not be resolved between B*13 or Cw*02.