bated at 4°C for 20 min, washed once (using 1 × PBS containing 2% human AB serum and 1% pen-strep) and then incubated with a second step of Goat-anti-mouse IgG FITC at 1:200 for 20 min at 4°C. The cells were washed twice and then analyzed on the FACS machine. Infected 721.221 cells that showed HLA surface expression (varying from 5% to 60%) were subsequently expanded, stained and sorted for W6/32 positive cells. Sorted cells were expanded in DMEM complete containing 400 μg/ml G418, and aliquots were frozen. The following 721.221 cell lines are available: A0101, A0201, A0202, A0203, A0204, A0205, A0206, A0207, A0301, A1101, A2301, A2402, A2403, A2902, A3101, A3301, A6601, A6801, A6802, B0702, B0801, B1501, B1502, B1510, B2702, B2703, B2705. B2706, B3501, B3801, B3909, B39011, B4002, B4402, B4403, B5101, B5301, B5401, B5502, B5701, C0102, C0304, C0702.

Results and discussion

Strategy for selection of HIV-1 specific CTL cell epitopes and generation of peptide pools

In this study, we approached the formulation of CTL targets from a collection of 199 unique HIV-1 CTL epitopes retrieved from the Los Alamos HIV database [14]. The peptides were all 9mers, which is the optimal peptide length for MHCI binding, and are known to elicit CTL responses in one or more HIV-1-infected individuals, indicating that they are naturally processed during the course of infection. Identification of processed epitopes eliminates one hurdle in CTL epitope-based vaccine design. Nonetheless, a successful CTL-based strategy must confront the sequence variability of HIV-1 strains and polymorphism of the human MHCI molecules (HLA I). HIV-1 sequence variation largely contributes to immune evasion. In fact, it has been suggested that at a population level, HIV-1 variants have evolved to adapt to CTL responses [23,24]. Certainly, CTL escape variants can arise from mutations within and/or flanking HIV-1 CTL epitopes [12].

To select target epitopes less likely to escape from CTL, we first subjected representative viruses from all HIV-1 clades (Table 1) to a Shannon Entropy (H) variability analysis [25], thereby removing or "masking out" any residue with H >1. See Additional file 1 for the resulting HIV-1 consensus proteome. All residues in the selected CTL epitopes are therefore quasi-invariant. The most proximal residue flanking the C-terminus of the CTL epitope is a determinant for cleavage by the proteasome, and mutations in that residue can abrogate T cell recognition [26]. Accordingly, we also excluded those CTL epitopes with a C-terminal flanking residue with H >1. As a result, only 37 of the 199 HIV-1 CTL peptides were chosen.

HLA I polymorphisms complicate development of a broadly protective CTL epitope vaccine by limiting population coverage. For a given selection of CTL epitopes, the population protection coverage (PPC) is given by the proportion of cumulative phenotypic frequency (CPF) of the HLA alleles restricting these epitopes. CPF can be computed from the HLA gene and haplotype frequency in the population. Using the HLA I gene and haplotype frequency reported by Cao et al. [17] for five major American ethnicities (Black, Caucasian, Hispanic, Native American, and Asian), we developed an algorithm that first computed the binding of each epitope to 55 HLA I alleles and then selected epitope combinations providing a PPC ≥ 95% for all ethnic groups considered (details in supporting methods). From this analysis, we predicted that as few as five epitopes from the 37 conserved HIV-1 specific CTL epitopes should be recognized by = 95% of the population, regardless of ethnicity. Furthermore, we identified 5-epitope combinations using only 25 of the 37 conserved HIV-1 specific CTL epitopes (Table 2). These 25 CTL epitopes were distributed as follows: POL (reverse transcriptase, integrase and protease), 14 epitopes; GAG, 5 epitopes; ENV, 3 epitopes; and NEF, 3 epitopes. None of the epitopes were found in the five remaining protein open reading frames (ORFs) encoded by the HIV-1 genome (VIF, TAT, REV, VPU/VPX, and VPR). This distribution reflects, in large part, the overall size as well as degree of conservation of the different HIV-1 ORFs. Furthermore, visual inspection of the relevant 3D structures reveals that these epitopes encompass residues that are important either for structural integrity and/or catalytic activity. For example, epitopes TLVDVGDAY and VIYQYMDDL from HIV-1 reverse transcriptase (POL) bear 6 and 4 residues, respectively, involved in substrate binding and catalysis [27]. The penalty for mutations at any of these sites vis-a-vis viral fitness presumably precludes ready development of escape variants.

A minority of HIV-1 patients responds to HIV-1 specific CTL peptide pools

That as few as 5 peptides may provide the requisite PPC largely stems from their ability to bind to many different HLA I molecules. The peptide SPRTLNAWV itself is anticipated to bind to eight distinct HLA I allelic variants, offering a PPC ranging from 35% in Asians to 52% in North American Indians. Such promiscuous MHCI-peptide binding often resides in previously defined HLA I supertypes (i.e. alleles with similar peptide binding specificity [28]), e.g., the peptide SPRTLNAWV binds to the B7 supertype and the peptide AVFIHNFKR to the A3 supertype. Nevertheless, binding promiscuity is not exclusively confined to HLA I supertypes. For instance, the peptide EKEGKISKI is predicted to bind to the alleles B2701, B3801 B39011, B3909, B4402, B5101 and B0801, which do not conform to any known supertype, and peptide