tat Exon 1 Exhibits Functional Diversity during HIV-1 Subtype C Primary Infection

Raabya Rossenkhan,[a,b,c] Iain J. MacLeod,[b,c] Theresa K. Sebunya,[a] Eduardo Castro-Nallar,[d,e] Mary Fran McLane,[b] Rosemary Musonda,[b,c] Berhanu A. Gashe,[a] Vlad Novitsky,[b,c] M. Essex[b,c]

Department of Biological Sciences, University of Botswana, Gaborone, Botswana[a]; Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA[b]; Botswana Harvard AIDS Institute, Gaborone, Botswana[c]; Department of Biology, Brigham Young University, Provo, Utah, USA[d]; Computational Biology Institute, George Washington University, Ashburn, Virginia, USA[e]

Human immunodeficiency virus type 1 (HIV-1) Tat is a mediator of viral transcription and is involved in the control of virus replication. However, associations between HIV-1 Tat diversity and functional effects during primary HIV-1 infection are still unclear. We estimated selection pressures in *tat* exon 1 using the mixed-effects model of evolution with 672 viral sequences generated from 20 patients infected with HIV-1 subtype C (HIV-1C) over 500 days postseroconversion. *tat* exon 1 residues 3, 4, 21, 24, 29, 39, and 68 were under positive selection, and we established that specific amino acid signature patterns were apparent in primary HIV-1C infection compared with chronic infection. We assessed the impact of these mutations on long terminal repeat (LTR) activity and found that Tat activity was negatively affected by the Ala[21] substitution identified in 13/20 (65%) of patients, which reduced LTR activity by 88% ($\pm$1%) ($P < 0.001$). The greatest increase in Tat activity was seen with the Gln[35]/Lys[39] double mutant that resulted in an additional 49% ($\pm$14%) production of LTR-driven luciferase ($P = 0.012$). There was a moderate positive correlation between Tat-mediated LTR activity and HIV-1 RNA in plasma ($P = 0.026$; $r = 0.400$) after 180 days postseroconversion that was reduced by 500 days postseroconversion ($P = 0.043$; $r = 0.266$). Although Tat activation of the LTR is not a strong predictor of these clinical variables, there are significant linear relationships between Tat transactivation and patients' plasma viral loads and CD4 counts, highlighting the complex interplay between Tat mutations in early HIV-1C infection.

Global incidence data from 2010 estimate that ~2.7 million people were newly infected with human immunodeficiency virus (HIV), contributing to a prevalence of ~34 million, with subtype C accounting for the majority of worldwide infections (1). Transmission rates are highest during primary infection (2, 3), and as a transactivator of HIV transcription, the viral protein Tat is a large contributor to the rate of viral replication. It is widely accepted that Tat is an important mediator of HIV type 1 (HIV-1) disease progression: Tat plays a central role in the regulation of HIV-1 gene expression by transactivating viral transcription (reviewed in reference 4), in addition to altering cellular gene expression (5–7). Tat interacts with RNA polymerase II (RNAPII) solely to mediate elongation of viral RNA; in the absence of Tat, full-length transcripts are rarely produced, as RNAPII prematurely dissociates from the template during early transcription, leading to the accumulation of abortive short viral mRNA.

Without external stimuli, latently infected CD4$^+$ T cells and monocytes restrict HIV expression via cellular components that negatively regulate the proviral long terminal repeat (LTR), suppressing the production of complete viral transcripts (8–10). This blockade can be relieved by Tat, which is produced from rare full-length transcripts and binds to the *trans*-activation response element (TAR) (11, 12), an RNA secondary structure at the 5′ end of nascent transcripts. The early, inefficient transcription of viral mRNA is further enhanced by Tat recruiting positively regulating cellular transcription factors to the 5′ LTR, ultimately facilitating the accumulation of full-length viral mRNA. Transcription is initially promoted by the modification of chromatin and the acetylation of histones by the CREB-binding protein (CBP)/p300 complex (13), recruited to the HIV promoter by Tat (14–16). The affinity of the CBP/p300 complex for components of the basal transcription machinery, such as the TATA-binding protein and

transcription factor IIB (17), is subsequently increased through conformational alteration by interacting directly with Tat. Corresponding acetylation of key Tat residues by the CBP/p300 complex enhances the subsequent engagement of Tat with the positive transcription elongation factor P-TEFb complex (12, 18, 19), composed of cyclin T1 and CDK9 subunits, permitting crucial phosphorylation of serine residues in the C-terminal domain (CTD) of RNAPII, resulting in complete viral gene expression (20). Additional phosphorylation events occur between Tat and cellular transcription factors, such as NF-κB and Sp-1 (21, 22), which play critical roles in the regulation of viral gene transcription. Subtype differences exist in the configuration of these and other transcription factor binding sites within the LTR, which have been reported to alter the responsiveness of HIV transcription to external stimuli: a single NF-κB binding site is present in the promoter of HIV-1 subtype E (HIV-1E), yet subtype B contains two with subtype C containing at least three, but as many as four, NF-κB consensus sequences (23). The variation in the number of transcription factor-responsive elements in the LTR has been shown to differentially regulate viral transcription (24) and replication in a dose-dependent manner upon exposure of HIV-infected cells to cytokines, such as tumor necrosis factor alpha (TNF-α) (25–27). Furthermore, the transactivation potential of Tat from different subtypes is not uniform: subtype E and C Tat

are strong mediators of LTR transcription compared to subtype B (28), and this is thought to be due to a higher affinity for the TAR hairpin (28). The half-life of Tat from subtype E is almost twice as long as that from subtypes B and C, which may be a compensatory mechanism for the reduction in NF-κB binding (28). Intrasubtype differences also exist, with a recent signature pattern analysis identifying unique residues in Tat, with higher frequencies of Ala[21], Asn[24], Lys[29], Lys[40], and Gln[60] from HIV-1 subtype C (HIV-1C) circulating in southern India compared to southern Africa (29, 30).

The restoration of complete viral gene expression by Tat reinforces its role as a major contributor to the establishment and enhancement of HIV replication during primary infection. During this time, plasma viral levels of HIV-1 are between 10^5 and 10^8 RNA copies/ml (31, 32), yet the levels decrease by 2 to 3 orders of magnitude within 4 to 6 months (33, 34). However, prolonged high viremia has been reported in a subset of patients during primary HIV-1 subtype C infection, through a yet undefined mechanism (35, 36). It is during this primary phase of infection that the reservoir of latently infected, resting $CD4^+$ T cells is established and that Tat, particularly variants with impaired activity, may contribute to the formation of a pool of quiescently infected cells (37).

The critical nature of Tat in controlling virus replication through augmented gene expression questions the extent to which Tat can endure sequence diversity while maintaining function, particularly during early infection when the establishment of a productive infection is paramount. It has been reported for HIV subtype B and simian immunodeficiency virus (SIV) that *tat* is one of the earliest viral genes to be under host selection pressure (38, 39), leading to sequence diversity early in infection, yet Tat can tolerate up to 40% sequence variation and still conserve its function (40). Such rapid mutation of viral cytotoxic T lymphocyte (CTL) epitopes in Tat, and other viral proteins, constrains successful HIV vaccine development. Data on differences between genes of regulatory and virion-associated HIV proteins suggest that HIV may actually evolve to permit an immune response against viral proteins that are not as essential as Tat and Rev during early replication, such as Gag, as a decoy (41). Nevertheless, levels of anti-Tat antibodies negatively correlate with diseases progression, suggesting that Tat plays a major role in determining AIDS progression (42, 43).

The conserved nature of Tat is compatible with its role as an important target for possible HIV treatments, as well as a logical vaccine candidate (44, 45). Current phase II trials in South Africa are evaluating a therapeutic Tat-based vaccine in anti-Tat antibody-negative individuals on antiretroviral therapy (ART) with high $CD4^+$ T cell counts (Istituto Superiore di Sanità [ISS] T-003 trial; ClinicalTrials.gov registration no. NCT01513135) (46). However, the expansion of Tat vaccine trials to incorporate HIV prevention will not be possible without a thorough understanding of the genetic and associated immunogenic characteristics of Tat in transmitted virus and during primary infection. We hypothesized that Tat diversity would be evident in patients with HIV-1 subtype C primary infection and that subsequent disparities in the transactivation potential of these Tat variants would be associated with plasma viral load. To our knowledge, this is the first study to genetically characterize HIV-1 *tat* during primary HIV-1C infection using single-genome amplification. To address this gap in

HIV research, we asked whether HIV-1 subtype C *tat* exon 1 was different when comparing primary infection to chronic infection.

TABLE 1 Characteristics of study subjects over the first 500 days postseroconversion[a]

Characteristic	Median value (IQR) unless specified otherwise for the following patients:		
	Acutely infected (patients A to H)	Early phase of infection (patients OC to QU)	P value
Viral load ($\log_{10}$ no. of RNA copies/ml)	4.95 (4.00–5.52)	4.40 (3.84–4.75)	0.11
CD4 count (no. of cells/μl)	337 (258–426)	374 (258–490)	0.72
Age (yr)	31.50 (26.75–34.25)	25.50 (24.75–27.00)	0.11
Sex, no. of females/total no. of patients (%)	6/8 (75)	9/12 (75)	1.00
Fiebig stage	II (II–II)	V (V–VI)	<0.01
No. of time points sequenced	3.50 (2.75–4.25)	2.00 (1.75–3.00)	0.06
Duration of follow-up (days)	362 (203–381)	279 (179–374)	0.59
No. of sequences	48 (33–73)	23 (16–27)	<0.01

[a] IQR, interquartile range. P values compare the values for acutely infected patients and patients in the early phase of infection.

MATERIALS AND METHODS

Study population. Patients were enrolled in an HIV-1C primary infection cohort in Botswana (47, 48), and a subset of 20 subjects was selected based on the stage of HIV infection: eight acutely infected individuals (patient code A to H) and 12 randomly selected seroconverters (patient code OC to QU) (Table 1). Acutely infected individuals were identified before seroconversion by a positive HIV-1 reverse transcription-PCR (RT-PCR) test with negative HIV-1 serology (Fiebig stage II) (49). Seroconverters were identified in the early stage of HIV-1 infection (Fiebig stages IV to VI). The time of seroconversion (day 0) was estimated as the midpoint between the last seronegative test and the first seropositive test for the acutely infected subjects and by the midpoint of the corresponding Fiebig stage for the recently infected subjects (47). Table 1 provides baseline clinical parameters, demographic information, and sampling characteristics summarized for the patients enrolled in the acute and early phases of primary infection. Written informed consent was obtained from all study participants, and ethical approval for this research was obtained from the Human Research Development Committee of the Botswana Ministry of Health and the Office of Human Research Administration at the Harvard School of Public Health.

Viral RNA extraction and single-genome amplification. Viral RNA extraction from plasma samples was carried out using the QIAamp viral RNA minikit (Qiagen, Valencia, CA) according to the manufacturer's instructions followed by single-genome amplification as described previously (50). Briefly, reverse transcription of viral RNA was performed using SuperScript III (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The single-genome amplification was based on the method of limiting dilutions (51) and was used with minor modifications. HIV-1C primary infection *tat* exon 1 sequences are available publically through GenBank accession numbers JQ895561 to JQ896230.

Phylogenetic relationships of HIV-1C *tat* exon 1 sequences. To assess the relationships between newly generated quasispecies from the 20 subjects in this study, JModelTest (52) was used to select the optimal evolutionary model. A maximum likelihood (ML) phylogenetic tree was constructed using PhyML (53). A SIV *tat* exon 1 sequence (Ref.CPZ.US.85.US_Marilyn.AF103) was used to root the tree. The tree was visualized in FigTree v1.1.2 (http://tree.bio.ed.ac.uk/software/figtree/) and MEGA5 (54).

MEME. Positive-selection analyses were performed under a likelihood framework by investigating for selection in a population-wise manner. Accordingly, signatures of selection were examined using the newly available mixed-effects model of evolution (MEME) available in HyPhy 2.11

(55–57). We used the Hasegawa-Kishino-Yano (HKY) nucleotide substitution model (58) crossed with the Muse and Gaut 1994 (MG94) codon model (59). The significance level was set at 0.1 for a site to be considered under selection. Those sites under positive selection were mapped to the Tat crystal structure (Protein Data Bank [PDB] accession no. 3MI9) and visualized using the Visual Molecular Dynamics program (VMD) (60). The sequence logo was generated with the online web logo tool at the University of California at Berkeley (UC Berkeley) (http://weblogo.berkeley.edu).

LANL sequences used for comparisons. Single-patient subtype C sequences available in the Los Alamos National Laboratory (LANL) HIV database (http://www.hiv.lanl.gov/), included 694 subtype C *tat* exon 1 sequences of the 1,544 total. Chronic infection HIV-1 subtype C sequence criteria included the following: (i) stage of infection at the time of sampling that was >1,000 days postinfection/seroconversion; (ii) nonrecombinant HIV-1 subtype C; and (iii) full-length *tat* exon 1. This included 183 HIV-1 subtype C chronic infection *tat* exon 1 sequences retrieved from the LANL HIV Database (accessed 28 September 2012). The majority of the available sequences were from southern Africa ($n = 179$); four sequences were from India. We assessed phylogenetic clustering of these sequences in relation to our newly characterized sequences and found that all study sequences were scattered across the phylogenetic tree with reference sequences, and no distinct clades or clusters were detected.

Amino acid frequency variation. Single-genome sequences generated from the earliest available time point postseroconversion were used to construct a primary infection consensus sequence for each patient. These 20 sequences, in addition to full-length chronic infection (>1,000 days postseroconversion) LANL *tat* exon 1 subtype C sequences ($n = 183$), were used in MargFreq (http://sray.med.som.jhmi.edu/SCRoftware/margfreq/) to generate amino acid frequencies at each codon. Differences in the primary infection virus were assessed statistically using the two-tailed Fisher's exact test with an α of 0.05.

HLA typing of major histocompatibility complex (MHC) class I and epitope analysis. High-resolution HLA typing was performed for all study subjects, except subjects OK, ON, and OP, as described previously (61). All ambiguous positions were resolved by resequencing. Polymorphisms outside the targeted exons that could not resolve heterozygote combinations were refined using a statistical method that calculates an HLA haplotype given the probability of association with fully resolved patient HLA class I genotypes at other positions (62). Previously defined CTL/CD8$^+$ epitopes to HIV-1C Tat were obtained from the LANL HIV Molecular Immunology Database. Putative HLA-restricted epitopes within subtype C Tat were identified by scanning the patient consensus sequences for HLA-restricted epitope anchor residue motifs using two independent bioinformatic tools: Motif Scan for HIV HLA anchor residue motifs (http://www.hiv.lanl.gov/content/immunology/motif_scan/motif_scan) and Epipred (63).

Generation of Tat expression clones. A *tat* exon 1 consensus sequence was generated for each patient from viral quasispecies sequenced during primary infection. *tat* exon 2, which is not required for transactivation of the HIV-1 LTR sequence, was derived from the HIV-1 subtype C consensus. *tat* sequences were cloned into the pGen2.1 vector, allowing for Tat expression via the cytomegalovirus (CMV) promoter. Introducing the relevant nucleotide substitutions into *tat* exon 1 of the subtype C consensus full-length Tat vector generated pGen2.1-ΔTat clones containing a single amino acid mutation. All cloning and site-directed mutagenesis were performed by Genscript (Piscataway, NJ, USA).

Cell culture and transfections. TZM-bl cells, obtained from the NIH AIDS Reagents Program, are HeLa cell derivatives that stably express β-galactosidase and firefly luciferase under the control of the HIV-1 LTR (64–68). The cells were cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, 100 U of penicillin/ml, 100 mg of streptomycin/ml, 2 mM L-glutamine, and 0.1 mM nonessential amino acids. No selective antibiotics were required for the maintenance of firefly luciferase expression. TZM-bl cells were transfected by electroporation with 1 μg of pGen2.1-Tat per 10^6 cells using the Neon system (Invitrogen, Carlsbad, CA) with two pulses of 1,200 V for 20 ms. For a negative control, a pGen2.1 vector lacking both *tat* exons (pGen2.1-empty) was similarly transfected into TZM-bl cells. The cells were aliquoted at 2×10^4 per well in 100 μl of supplemented DMEM without antibiotics in 96-well plates. To normalize for potential differences in transfection efficiency and variation in the number of cells plated, 4 μg of pRL-SV40 (Promega), a *Renilla* luciferase-expressing construct under the control of the simian virus 40 (SV40) promoter, which is minimally transactivated by Tat, was added to a master batch of TZM-bl cells before they were aliquoted for transfection with a pGen2.1-Tat construct.

Luciferase assay. After 48 h, the cells were lysed in Dual-Glo firefly luciferase buffer (Promega) and incubated at room temperature for 10 min, and firefly luciferase activity was determined using a PerkinElmer luminometer. Firefly luciferase activity was then simultaneously quenched by the addition of the Dual-Glo *Renilla* luciferase substrate, which was also incubated at room temperature for 10 min, and activity was similarly determined by the luminometer.

Data and statistical analyses. Firefly luciferase values were normalized to *Renilla* luciferase activity. The subtype C consensus Tat construct was assigned the value 1, and differences in LTR transactivation by patient Tat constructs were calculated relative to the consensus and are given as the median difference in expression ± interquartile range. A normal distribution of the luciferase data was assessed by the D'Agostino and Pearson omnibus normality test (69), and normalized luciferase data were $\log_{10}$ transformed for use in linear regression and parametric analyses. A two-tailed Student's *t* test was performed on the transformed, normalized luciferase data between the Tat activity for consensus HIV-1C compared to each one of the Tat constructs from the patients. The Pearson product-moment correlation coefficient was used to determine the strength of association or correlation between Tat transactivation and patients' plasma RNA viral loads, CD4 counts, and proviral loads. *P* values were corrected for multiple comparisons using the Benjamini-Hochberg method to control the false discovery rate (70). Adequate quasispecies sampling from each patient was determined by estimating the likelihood of missing infrequent viral variants present in patient plasma: with *n* single genomes sequenced, the probability (*P*) of missing a quasispecies after screening *n* genomes is calculated as $f = 1 - (1 - P^{1/n})$ when the variant comprises a fraction *f* (or less) of the virus population (71).

RESULTS

Minimal Tat diversity during primary infection. The phylogenetic relationships between HIV-1C *tat* exon 1 viral quasispecies from 20 subjects during primary infection are shown in Fig. 1. The maximum likelihood tree highlights distinct patient-specific clusters (except the subject pair OK/OI, as described previously [50]), and a high level of homogeneity between quasispecies over 500 days postseroconversion.

Positive selection of Tat residues during primary infection. We employed an evolutionary framework to determine whether any specific sites in *tat* exon 1 are under positive selection during primary infection. The mixed-effects model of evolution (MEME) detects positively selected sites despite the majority of lineages at a given site evolving under purifying selection, which may mask potential episodic selection acting on a small proportion of lineages. Our analysis found that residues 3, 4, 21, 24, 29, 39, and 68 of *tat* exon 1 were undergoing positive selection during primary infection (Table 2). We observed positive selection affecting a small subset of branches at sites 3 and 39 and, to some extent, at sites 24 and 29. In contrast, positive selection appears to be pervasive, in a rather large proportion of branches, at sites 4, 21, and 68 and with various levels at sites 24 and 29 (Table 2). Furthermore, our results support recently published data revealing an epistatic
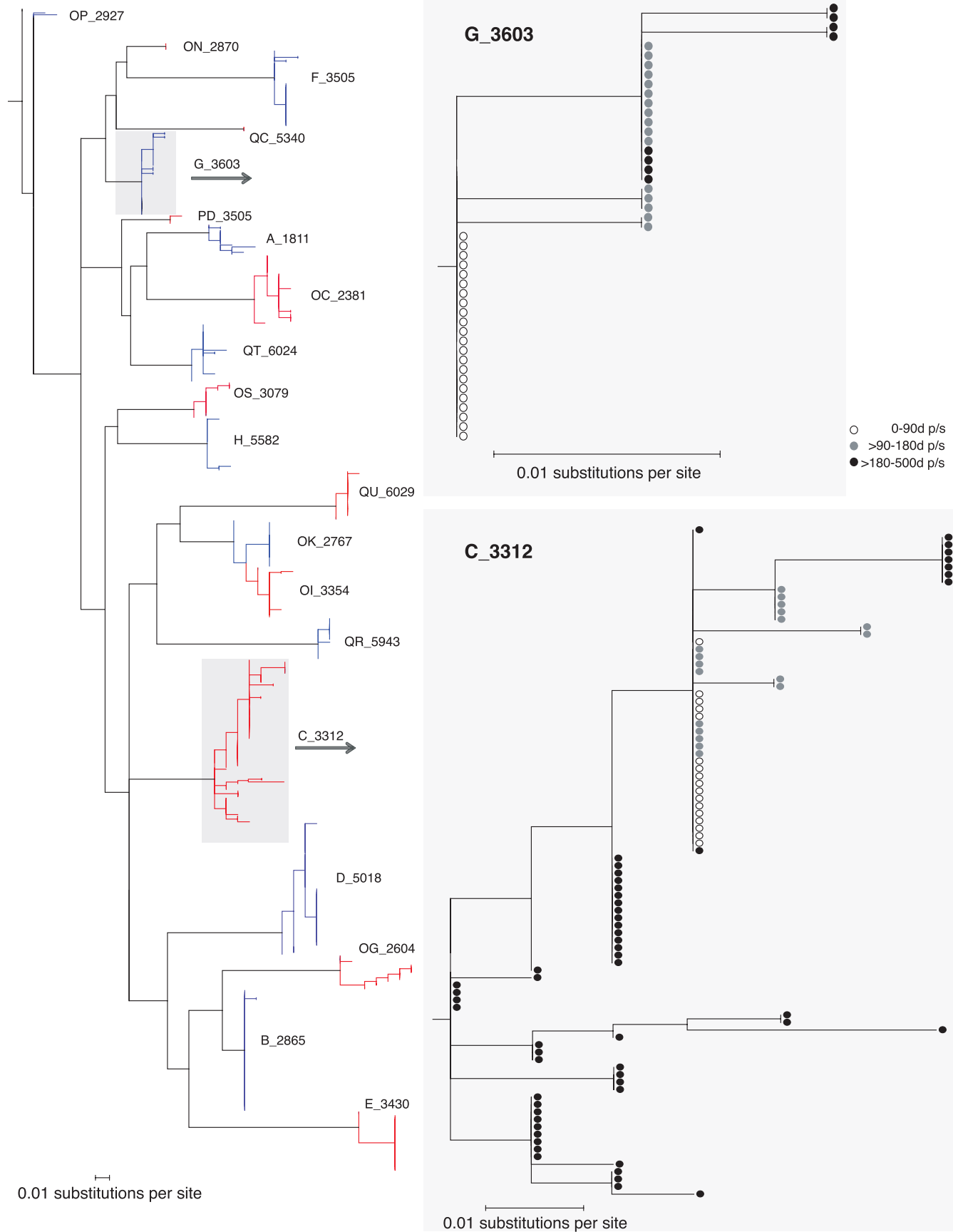
FIG 1 Maximum likelihood phylogenetic analysis of primary infection HIV-1 subtype C *tat* exon 1 viral quasispecies. A phylogenetic tree was constructed using PhyML (16) and the TVM+I+G model (TVM stands for transversion model, I is the proportion of invariable sites, and G is the gamma distribution across the branches) for nucleotide substitution. The tree was visualized in FigTree and MEGA5 (45). SIV sequence (Ref.CPZ.US.85.US_Marilyn.AF103) was used to root the tree. The data are shown for 0 to 90, >90 to 180, and >180 to 500 days postseroconversion (d p/s).

TABLE 2 MEME selection for *tat* exon 1 from 20 patients over the first 500 days postseroconversion

Codon	Wild-type residue	Replacement residue(s)	MLE[a] of the following:					P value[g]
			α[b]	β^{-c}	$Pr\beta^{-d}$	β^{+e}	$Pr\beta^{+f}$	
3	Glu	Asp	5.00×10^{-9}	1.00×10^{-15}	9.13×10^{-1}	9.51	0.087	0.036
4	Pro	Gln/Ile	5.00×10^{-9}	54.99×10^{-9}	1.00×10^{-9}	1.50	1.00	0.070
21	Pro	Ala	4.00×10^{-7}	4.00×10^{-7}	1.26×10^{-1}	2.34	0.874	0.033
24	Lys	His/Asn/Pro/Ser/Thr	9.92×10^{-1}	3.60×10^{-2}	5.75×10^{-1}	26.12	0.425	0.003
29	His	Ala/Glu/Ile/Lys/Asn/Pro/Gln/Arg	8.23×10^{-1}	4.36×10^{-6}	6.76×10^{-1}	49.51	0.324	<0.001
39	Gln	Leu	5.00×10^{-9}	3.75×10^{-16}	9.80×10^{-1}	30.76	0.020	0.031
68	Leu	Pro/Ser	5.00×10^{-9}	4.98×10^{-9}	1.00×10^{-9}	2.34	1.00	0.021

[a] MLE, maximum likelihood estimate.
[b] α is the synonymous substitution rate.
[c] β^{-} is the nonsynonymous substitution rate for the branch class with $\beta \leq \alpha$.
[d] $Pr\beta^{-}$ is the proportion of sites evolving at β^{-}.
[e] β^{+} is the nonsynonymous rate for the branch class with β^{+}.
[f] $Pr\beta^{+}$ is the proportion of sites evolving at β^{+}.
[g] P values comparing whether $\beta > \alpha$.

interaction between residues in Tat (72). We did not observe individual changes at either Leu[35] or Gln[39], but we determined that all sequenced patient quasispecies from patient F carried genes encoding both the L35Q and Q39L coevolving mutations. Although MEME identified only residue 39 as positively selected, given the conservative approach of this model, a possible explanation is that two-nucleotide changes were necessary for the Q39L transition (CAG to TTG) in patient F, while only one-nucleotide change was necessary for L35Q (CAG to CAA).

To appreciate the spatial interactions of positively selected residues 3, 4, 21, 24, 29, and 39 in the context of the Tat quaternary structure, we mapped these residues to the only high-resolution (2.1-Å) crystal structure of Tat available (Fig. 2A) (73). We were unable to map residue 68, because reliable crystal structures for that region are not defined. The majority of selected amino acid residues lie on the Tat activation domain (amino acids 1 to 48) that interacts with cellular proteins such as the positive transcription elongation factor P-TEFb; therefore, we might expect changes in these residues to have functional consequences. In particular, residues 3 and 4 reside on the critical acidic/proline-rich region, and residues 21, 24, 29, and 39 reside on the cysteine-rich/ Zn^{2+} finger region of the activation domain, which mediates the coordination of Zn^{2+} ions. We also observed high diversity within positively selected sites, probably related to the fact that those sites are found to be under diversifying selection (Fig. 2B). Two of the more diverse codon positions (sites 24 and 29) exhibiting a balanced proportion of branches under β^{+} and β^{-}, which suggests that a substantial amount of both positive and negative selection is occurring on the branches, though over 50% are under negative selection (Table 2 and Fig. 2B).

Frequency of Tat amino acid residues differs between primary and chronic HIV infection at specific sites. Signature pattern differences were apparent in Tat from the primary infection cohort compared with chronic HIV-1C infection by assessment of amino acid frequencies at each codon position of *tat* exon 1. Glu[2] was less frequent during primary infection (75%) compared to chronic infection (93%) ($P = 0.020$), which was complemented by a significantly greater prevalence of Asp[2] in primary infection (25%) compared to in chronic infection (7%). Only two of seven sites that were found to be under positive selection during primary infection showed significant differences compared with chronic

infection sequences: 15% Lys[24] in primary infection versus 44% in chronic infection ($P = 0.015$) and 30% Ser[24] in primary infection versus 3% in chronic infection ($P < 0.001$). Significant differences were also identified at Ala[29], present in 10% of primary infection clinical isolates yet only 1% in those chronically infected with HIV-1C ($P = 0.048$).

Additionally, we found a statistically significant difference between primary infection and chronic infection *tat* exon 1 amino acids at position 59 in terms of the amino acid charge in the two stages of HIV infection. The effect of the amino acid in side chain hydrophobicity/hydrophilicity from the chronic infection sequences are not well understood, but this change was observed in the auxiliary region of *tat* exon 1 (amino acids 58 to 71). During primary infection, *tat* exon 1 had significantly more neutral residues at position 59 than in chronic infection, with the hydrophilic Pro[59] residue being the most prevalent. Only 6 (3%) sequences from chronic infection had a neutral amino acid at position 59 compared with 15% of our newly characterized *tat* exon 1 sequences ($P = 0.046$ by Fisher's exact test).

Amino acid substitutions in Tat associate with HLA. CTL responses have been mapped to only a small number of HIV-1 subtype C Tat epitopes. To systematically address the role of HLA in influencing intrapatient amino acid diversity during primary infection, we analyzed single-genome sequencing data to assess substitutions in Tat, compared to the subtype C consensus sequence (Fig. 3). In addition to the functional data available from LANL, we employed two predictive analyses, MotifScan and Epipred, to discern regions that might be under immune pressure based on the HLA genotypes of the patients. There were only two instances where the transmitted virus had the wild-type Tat residue and mutated over the first 500 days postseroconversion, both of which were potentially HLA restricted: in patient OC, the switch from His[29] to Pro[29] or Arg[29] by 208 days postseroconversion was restricted by both HLA-A*03:01 and HLA-A*68:02; by 411 days postseroconversion, patient OS had switched to Asp[24], completely replacing the wild-type Lys[24] residue, which was restricted by HLA-B*35:01. Both of these substitutions occurred in a region that has been identified as immunodominant in subtype C Tat (74). The conserved region, comprising the protein transduction domain, 47-YGRKKRRQRR-56 and the nuclear localization signal, 48-GRKKR-52, was highly conserved during primary in-
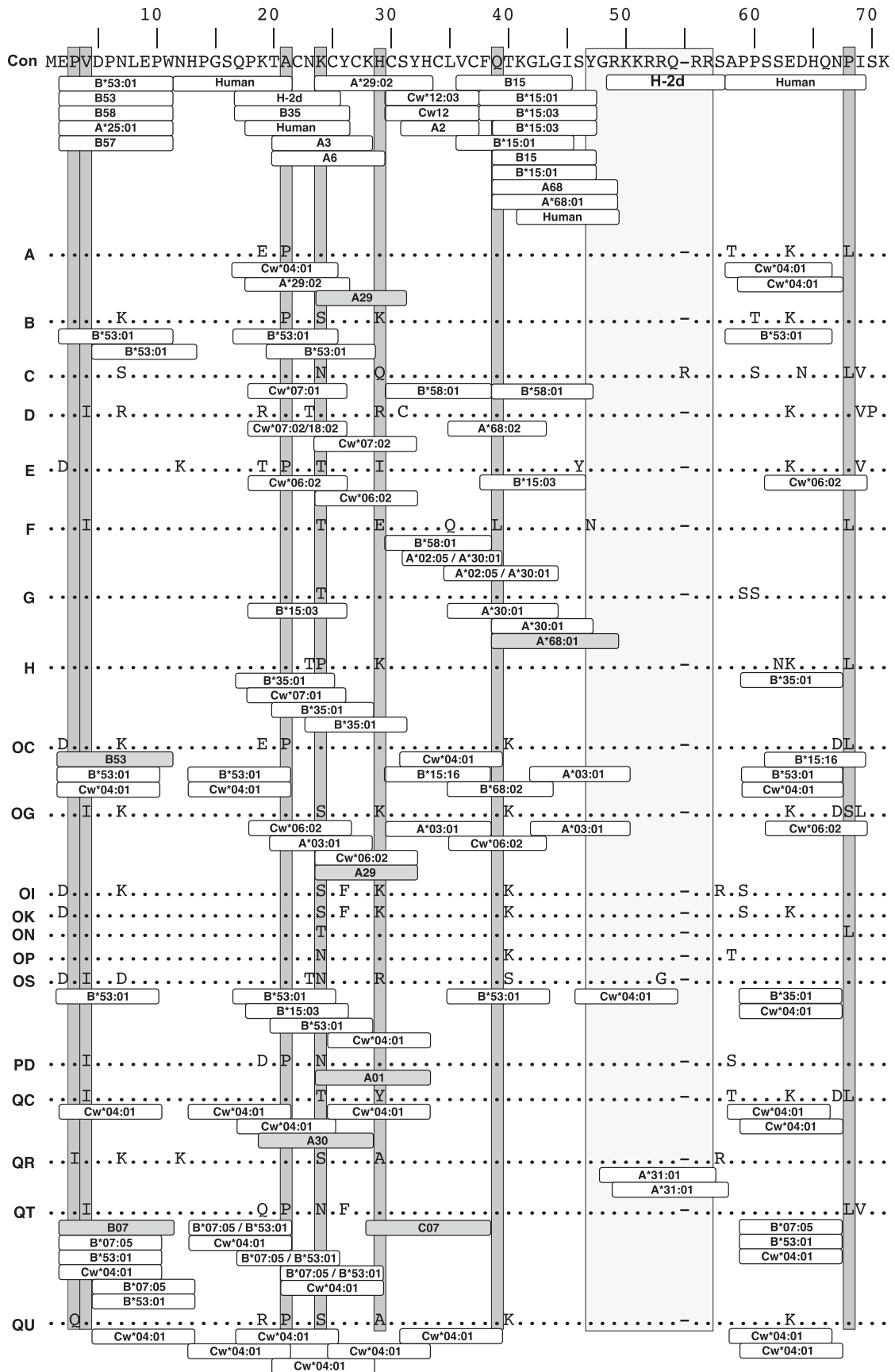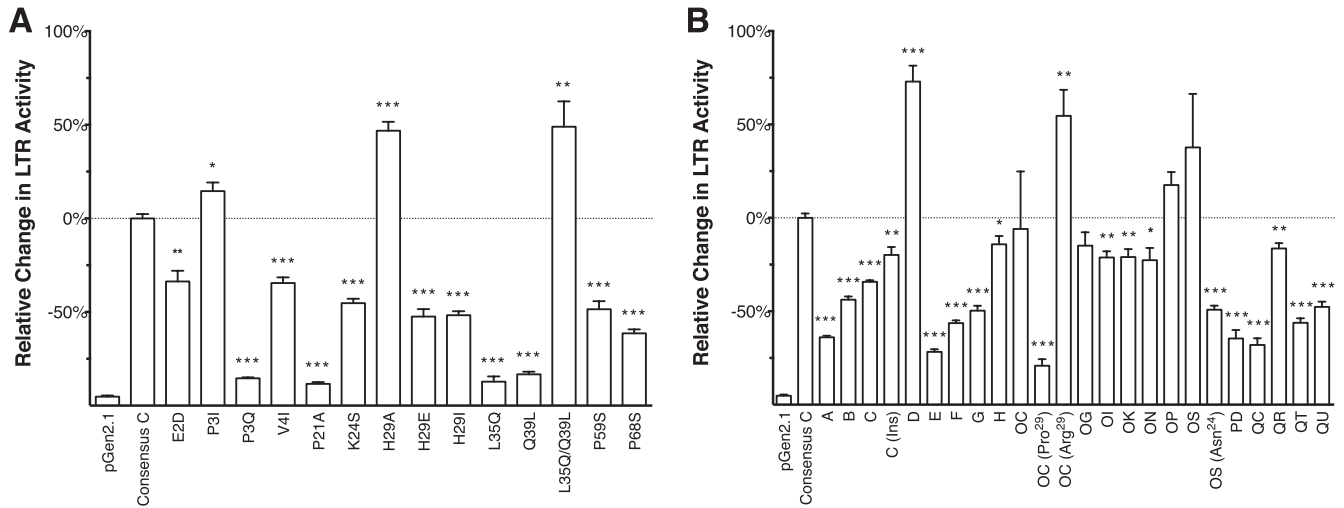
A
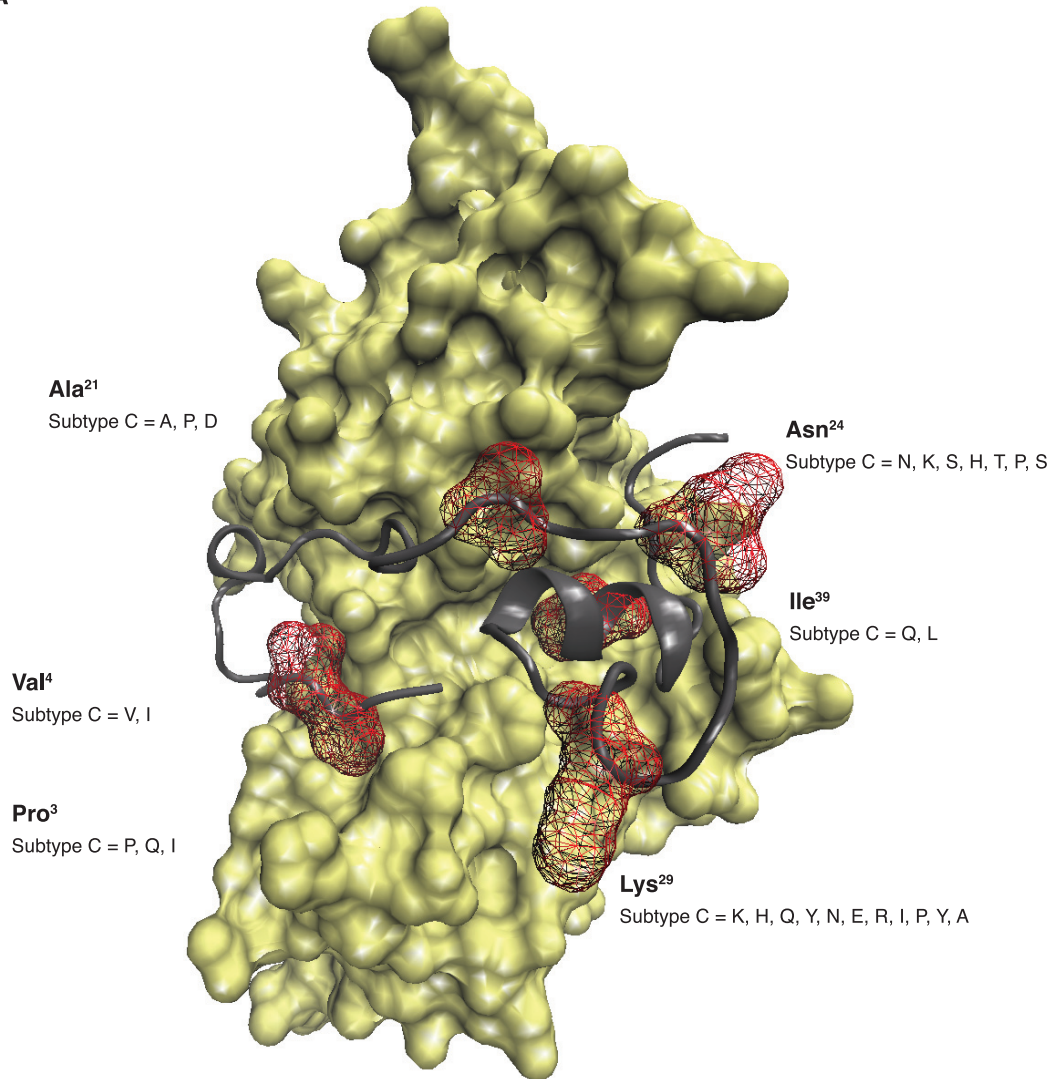
B
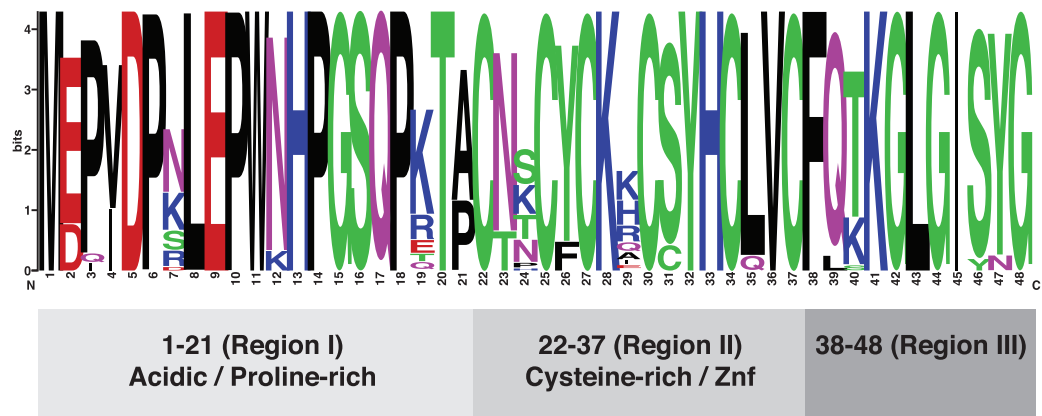
FIG 2 Positive selected Tat residues in context. (A) HIV-1 subtype B (HXB2) crystal structure with positively selected subtype C Tat sites mapped in relation to cyclin T1. Six out of seven sites were mapped (red), and variation in subtype C during primary infection is shown compared to the residue present in the HIV-1 HXB2 strain (bold type). (B) Sequence logo showing the Tat amino acid diversity observed at positions 1 to 48. The three regions of *tat* exon 1 are highlighted: region I (amino acids 1 to 21), an acidic/proline-rich region; region II (amino acids 22 to 37), a cysteine-rich/Zn^{2+} F region; and region III (amino acids 38 to 48), containing conserved phenylalanine (F).

```
                    10        20        30        40        50        60        70
                    |         |         |         |         |         |         |
Con  MEPVDPNLEPWNHPGSQPKTACNKCYCKHCSYHCLVCFQTKGLGISYGRKKRRQ-RRSAPPSSEDHQNPISK
```


Sequence alignment figure with HLA epitope annotations.

FIG 4 Relative changes in LTR activity. (A) Changes in LTR-luciferase activity by individual Tat mutations were determined relative to the HIV-1 subtype C consensus sequence at 48 h posttransfection. The pGen2.1 empty vector was included as a negative control. Values are the medians from four independent experiments, and the error bars indicate the interquartile ranges. (B) Changes in LTR-luciferase activity by patient-derived Tat were determined relative to the subtype C consensus sequence at 48 h posttransfection. The pGen2.1 empty vector was included as a negative control. Values are the medians from four independent experiments, and the error bars indicate the interquartile ranges. Statistical significance is indicated as follows: ∗, $P < 0.05$; ∗∗, $P < 0.01$; ∗∗∗, $P < 0.001$.

fection, with only two instances of mutation: R53G in patient OS, and S57R, flanking the conserved region, in patient QR. Both changes, present at the earliest time point and maintained for at least the first 500 days postseroconversion, were predicted to be under HLA restriction by Cw*04:01 and A31:01, respectively.

HIV-1C interpatient polymorphisms significantly impact LTR transactivation. A number of patients in the primary infection cohort contained amino acid residues that differed from the HIV-1 subtype C consensus at sites identified by MEME as being under positive selection. On the basis of these data, we introduced the corresponding single-amino-acid changes into *tat* exon 1 of the subtype C consensus to determine the functional effects of these positively selected sites. The majority of these single polymorphisms had a detrimental impact on LTR activation by Tat (Fig. 4A). The greatest impact on Tat was seen with the Ala[21] that was identified in 13/25 (65%) patients, which reduced LTR activity by 88% ($\pm$1%) ($P < 0.001$), which was less than a 2.5-fold increase above background luciferase levels seen with the pGen2.1-empty vector control. Conversely, the greatest increase in Tat activity was seen with the Gln[35]/Lys[39] double mutant that resulted in an additional 49% ($\pm$14%) production of LTR-driven luciferase ($P = 0.012$). It has previously been reported that Gln[35] and Lys[39] coevolve *in vivo* and that mutation of either residue results in a defective virus (72). Our data recapitulate this by showing that Gln[35] reduces LTR activation by 87% ($\pm$3%) ($P < 0.001$) and Lys[39] by 83% ($\pm$1%) ($P < 0.001$) compared to the wild-type Tat construct.

Patient-derived Tat exhibits functional diversity during primary infection. Compensatory mutations may maintain protein structure and function when deleterious changes occur at other residues, so we chose to determine the LTR transactivation capacity in the context of the full-length *tat* exon 1 from our cohort of patients during HIV-1C primary infection relative to the activity of the HIV-1 subtype C consensus Tat (Fig. 4B). Tat from patients OG, OP, and OS and Tat expressed by patient OC at 27 days postseroconversion demonstrated no significant change in luciferase expression compared to the HIV-1C consensus Tat. At 208 days postseroconversion, patient OC expressed three Tat quasispecies that differed at a single amino acid, encoding either His[29], which was dominant at earlier time points and had no significant effect on LTR activity, or Arg[32] or Pro[32]. The largest decrease in Tat activity across all patients, relative to the subtype C consensus, was seen with the OC patient Tat that contained the Pro[29] mutation, resulting in a 79% ($\pm$4%) ($P < 0.001$) decrease in LTR activation. Conversely, the Arg[29] substitution significantly upregulated LTR-luciferase expression by 55% ($\pm$14%) ($P = 0.008$). The only other significant increase in Tat transactivation was seen in patient D, resulting in a 73% ($\pm$9%) ($P < 0.001$) upregulation of expression.

The Tat sequence for patient C contained a three-nucleotide insertion that introduced an additional arginine residue into the transactivation domain, subsequently increasing LTR activity by 22% ($\pm$6%) ($P = 0.034$) compared to the patient C Tat without the insertion, yet remained significantly less active than the consensus C Tat. Of the two patients that demonstrated sequence diversity over time, patient OS expressed the wild-type Lys[24] residue at day 203 days postseroconversion, which demonstrated no difference in Tat transactivation activity compared to the consen-

FIG 3 *tat* exon 1 during primary HIV-1 subtype C infection and putative HLA epitope determined by Epipred (shaded) and MotifScan (white). MEME positively selected sites are highlighted in gray. The conserved region, comprising the nuclear localization signal (48-GRKKR-52) and protein transduction domain (47-YGRKKRRQRR-56) is shown at the top of the figure. Amino acid residues identical to those in the conserved region are indicated by periods for the sequences from the patients. Due to the arginine (R) insertion in patient C, gaps were introduced to maximize alignment and are indicated by dashes. No HLA typing data were available for patients OK, ON, and OP.
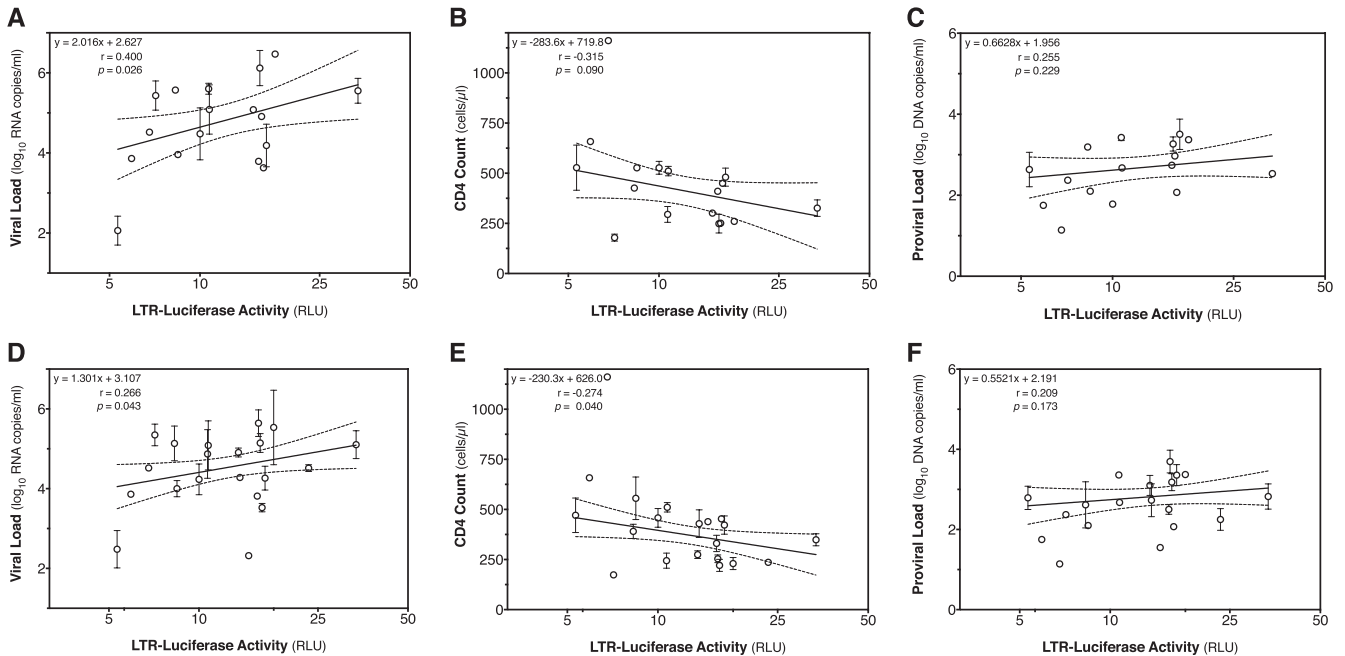
FIG 5 Correlation of Tat activity with clinical variables. A Pearson product-moment correlation coefficient and linear regression (solid line) were performed to determine a predictive association between the median LTR-luciferase activity (in relative light units [RLU]) and the patient plasma viral loads (A), CD4 counts (B), and proviral loads (C) over the first 180 days postseroconversion and the patient plasma viral loads (D), CD4 counts (E), and proviral loads (F) over the first 500 days postseroconversion. The 95% confidence intervals are shown as dotted lines.

sus. Yet by 411 days postseroconversion, complete replacement with Asn[24] (A-to-T change at the third nucleotide of codon 24) reduced Tat activity by 63% ($\pm$2%) ($P = 0.021$) compared to activity at 203 days postseroconversion. Similarly, patient OC expressed a Tat variant at 27 days postseroconversion that transactivated the LTR at a level similar to that of the HIV-1 consensus C. However, by 208 days postseroconversion, 11 of 35 sequences (31.4%) had Arg[29] and 5 sequences (14.3%) had Pro[29], with the remaining 19 sequences (54.3%) retaining the wild-type His[29] residue. The Arg[29] and Pro[29] mutations had opposing effects on Tat activity: a switch to proline at position 29 reduced Tat activity by 78% ($\pm$4%) ($P = 0.027$), whereas an arginine residue increased Tat activity by 64% ($\pm$15%) ($P = 0.164$) compared to the His[29] Tat variant that predominated in patient OC at 27 days postseroconversion. The nonsignificant P value for the large change in LTR activity by the Arg[29] mutation is likely due to the high standard deviation seen in the luciferase assay data performed using the Tat construct from patient OC 27 days postseroconversion.

Association of LTR transactivation by Tat with clinical characteristics. The relationships between the Tat-mediated LTR activity and patient plasma viral loads, CD4 counts, and proviral loads over the first 180 days postseroconversion were examined by the Pearson product-moment correlation coefficient and linear regression. There was a moderate positive correlation with plasma viral load ($P = 0.026$; $r = 0.400$) (Fig. 5A) and a weak negative correlation with CD4 count ($P = 0.090$; $r = -0.315$) (Fig. 5B), but not with proviral DNA load ($P = 0.229$; $r = 0.255$) (Fig. 5C). Three patients (patients ON, OP, and OS) did not have clinical data available before 180 days, so they were excluded from the analysis. Later set point viral loads, up to 500 days postseroconversion, were more weakly correlated with Tat-mediated LTR activity ($P = 0.043$; $r = 0.266$) (Fig. 5D) and showed a weaker neg-

ative correlation with CD4 count ($P = 0.040$; $r = -0.274$) (Fig. 5E). There remained no correlation with proviral load ($P = 0.173$; $r = -0.209$) (Fig. 5F). While the independent linear relationships between the Tat-mediated luciferase expression and patients' plasma viral loads and CD4 counts are significant, Tat activation of the LTR may not be a strong predictor of these clinical variables.

DISCUSSION

To our knowledge, this is the first in-depth functional characterization of *tat* exon 1 during primary HIV-1 subtype C infection. We employed the recently developed MEME method for detecting selection acting on HIV-1 subtype C Tat sequences during primary infection. The MEME method has advantages over previous methods, as it is able to detect sites under diversifying selection on most lineages of the underlying phylogeny (pervasive selection), as well as those sites under selection only on a small proportion of lineages (episodic selection). Usually, traditional selection detection methods mask sites that are subjected to selection in a small proportion of lineages, which hampers the power to observe early dynamic processes. Seven sites were found to be under positive selection with various levels of pervasiveness on the underlying phylogeny. For instance, sites 3 and 39 are involved in Tat regions that are essential for its catalytic and regulatory activities, respectively. The finding that these sites are being selected on a small proportion of sites might indicate a change in Tat performance and thus have functional consequences. Similarly, we identified sites (sites 24 and 29) under positive selection in about 50% of lineages, and the proportions of Lys[24], Ser[24], and Ala[29] differed in primary and chronic infection. This finding suggests different selective pressure in primary and chronic HIV-1C infection and is supported by functional analysis showing that these amino acid

substitutions have significant functional consequences for Tat transactivation.

We found that site 39 is under episodic positive selection on a small proportion of branches ($Pr\beta^+ = 0.020$) and is also engaged in bidirectional epistatic interactions with site 35 (posterior probability = 1), a phenomenon that has already been reported *in vivo* (72). This reinforces the results of our selection analysis, since a mutation at site 39 results in reduced LTR activation in our subsequent functional assays; naturally, most of the lineages show site 39 to be under negative or purifying selection, suggesting that this position is constrained by its epistatic interaction with site 35. Given that the majority of HIV-1 Tat data available are from population sequencing, to our knowledge, this is the first report demonstrating that the Glu[35] and Lys[39] substitutions are epistatically linked in the same RNA molecule.

On the basis of the MEME data, we performed functional analysis of individual Tat amino acid substitutions at sites identified as being under positive selection during primary infection. The majority of these changes had a significant negative impact on the transactivation potential of Tat to activate the LTR. Only two single alterations, P3I and H29A significantly increased LTR transcription compared to the wild-type HIV-1 subtype C Tat, and these substitutions were present in the same patient, patient QR. However, expression of this patient's full-length *tat* exon 1 resulted in a significant 16% reduction in LTR activity, highlighting the complex interplay between mutations in Tat. This is exemplified by the epistatic substitutions at Lys[35] and Glu[39] that individually abrogate important interactions between Tat and P-TEFb (72) but when coevolving, restore the transactivation activity to beyond wild-type levels. A number of Tat residues in exon 1 have been identified as crucial for their ability to recruit CBP/p300 and P-TEFb and to transactivate the LTR. Numerous sites have been reported to undergo posttranslational modification by cellular proteins. (i) The arginine-rich motif, involved in RNA binding, is modified by methylation and acetylation at a string of residues. (ii) Lys[50] is acetylated by p300 after coupling with Tat (75). (iii) Methylation occurs at Lys[51] by the lysine methyltransferase KMT7 (76, 77). (iv) Methylation occurs at Arg[52] and Arg[53] by the protein arginine N-methyltransferase PRMT6 (78). The arginine-rich motif lies within the transduction domain, which was completely conserved in all patients except for the R53G substitution observed in patient OS, which was potentially restricted by Cw*04:01 and maintained longitudinally. Tat from this patient containing the Arg[53] substitution had no discernible effect on LTR transcription in functional assays. Autoacetylation by Tat occurs specifically on Lys[41] and Lys[71] (79), two residues that are completely conserved in our primary HIV-1C infection data. The absence of Lys[41] reduces virus replication by preventing the enhancement of histone acetyltransferase activity by p300 (79).

Phosphorylation occurs at Ser[16] and Ser[46] by cyclin-dependent kinase 2 (CDK2) (80) and at Ser[62], Thr[64], and Ser[68] by protein kinase R (PKR) in subtype B (81), yet variability in the prevalence of these serine residues in primary HIV-1C infection suggests that subtype differences may exist. Amino acid substitutions at residues 16 and 46 are reported to significantly reduce virus production *in vitro*, and these serine residues were conserved in our cohort except for a S46Y substitution observed in patient E, a switch that was previously identified to be restricted by the same HLA found in patient E, B*15:03 (80). Of those Tat residues phosphorylated by PKR, only Ser[62] is conserved in our primary infection

cohort; Thr[64] is present as Lys[64] or Asp[64], and Ser[68] is present as Leu[68] or Pro[68]. In subtype B, the importance of retaining these phosphorylation sites has been demonstrated: a 42% reduction in LTR activity was reported with either an A64T or A68S substitution, with a Tat double mutant reducing transactivation by 59% (81). However, these deleterious mutations were in the context of an HIV-1B Tat; despite the high prevalence of Thr[64] and Ser[68] substitutions in our primary infection cohort, subtype C Tat has been shown to be a more potent transactivator than its subtype B counterpart. It may be that compensatory mutations exist in HIV-1C, as our functional data demonstrated that substituting Ser[68] with Pro[68] actually significantly increased LTR transcription by >250%.

The established coevolution of sites Glu[35] and Lys[39] was observed in one of our patients, and functional analysis in the context of a subtype C Tat background reinforced the negative impact of mutating either residue individually, with rescue of the wild-type phenotype achieved by the introduction of the corresponding compensatory substitution. Glu[35] and Lys[39] independently contribute to distinct aspects of Tat-mediated transcription: binding to P-TEFb and promoting phosphorylation of RNAPII by P-TEFb, respectively (72). The single L39Q mutant will retain the ability to interact with P-TEFb yet will fail to phosphorylate the CTD of RNAPII, whereas the Q35L substitution abrogates any binding with P-TEFb (72). The importance of invariant cysteine residues in Tat, which are highly conserved between subtypes, has been extensively studied (82). Of the seven cysteines, only Cys[31] has been shown to be dispensable for Tat transactivation, while the substitution of any other cysteine residue between sites 22 and 37 abolishes Tat function. DNA binding and transcriptional activity of the NF-κB p65 subunit is enhanced through direct binding to Tat via this cysteine-rich region and is complemented by a corresponding association between the arginine-rich motif of Tat, which sequesters the NF-κB inhibitor, IκBα, allowing the transcription factor to relocalize to the nucleus. All of the crucial cysteines are entirely conserved during primary HIV-1C infection, with the variable Cys[31] present only in patient D, with all other patients expressing the characteristic subtype C Ser[31] residue. The adjacent, invariable Cys[30] forms a CC chemokine motif in other HIV-1 subtypes that retain the Cys[31] (83), which is reported to be important for monocyte chemotaxis and subsequent interleukin 10 (IL-10) induction (84). Abrogated binding of the CS motif in subtype C to chemokine (CC motif) receptor 2b (CCR2b) prevents downstream signaling that culminates in cytokine induction and the upregulation of chemokine (CXC motif) receptor 4 (CXCR4) on CD4+ T cells (83). Although patient D harbored the CC motif and displayed significantly higher levels of LTR induction in comparison to the wild-type subtype C consensus Tat, it has previously been shown that substituting the Ser[31] residue for Cys[31] does not alter transactivation potential in subtype C (83). The retention of serine at this site may have yet unidentified functional importance and could potentially act as an alternative phosphorylation site.

Other subtype differences exist in sites that have a less established role in mediating LTR transactivation. Selection for Arg[23] in HIV-1 subtype B Tat confers a replicative advantage, yet Arg[23] was the predominant residue in primary subtype C infection and showed no significant difference in prevalence with chronic infection ($P = 0.778$ by Fisher's exact test). We observed low frequencies of Ala[21], Asn[24], Lys[29], Lys[40], and Gln[60], suggesting our data are

in agreement with a recent report identifying unique signature residues for HIV-1C in southern India compared to southern Africa. However, we also identified amino acid frequency variations at sites 24 and 29 during primary infection compared to chronic HIV-1C, with a significant increase in Ser^{24} and Ala^{29} in our cohort. Removal of the four *tat* exon 1 sequences from chronic infection viruses originating in India did not alter the significance of our results. Heterogeneity within HIV-1C *tat* exon 1 coupled with the variation seen at different stages of infection suggest deviating evolutionary paths for two geographically distinct clusters of subtype C, reinforcing the need for comparative characterization of these subepidemics.

We recently found that *tat* exon 1 intrapatient diversity was associated with plasma viral loads at baseline ($r^2 = 0.505$; $P = 0.044$) and at later stages of infection, between 181 and 500 days postseroconversion ($r^2 = 0.516$; $P = 0.039$) (50). With the need for Tat activity to be determined in the context of potentially unknown compensatory substitutions, we chose to evaluate the association between patient *tat* exon 1 and plasma viral load, CD4 count, and proviral load during the first 500 days postseroconversion. Although the correlation between Tat-mediated LTR activity and viral load was not strong, our results show that Tat transactivation is a significant contributor to plasma viral load and CD4 count in primary infection, but not to proviral load. However, a limitation to our approach prevented us from matching a patient's HIV-1 LTR sequence with the associated *tat* exon 1, instead relying on an integrated HIV-1B LTR-driven luciferase reporter in TZM-bl cells. The original study (50) targeted diversity and diversification of accessory genes and was designed to produce an amplicon that extended from the start of the *vif* open reading frame to the end of *vpu*, which encompassed *tat* exon 1 but does not span *tat* exon 2. The spatial separation of *tat* exon 2 also precluded its inclusion in this analysis. Although we may have identified a stronger correlation between Tat transactivation of its autologous LTR with viral and proviral loads and CD4 count during primary infection, our use of a standardized LTR allowed us to evaluate the impact of subtype C Tat mutations without the additional variability in promoter transactivation arising from potential LTR substitutions.

An additional limitation of this study was the difference in the median number of sequences generated for those with acute HIV-1C infection compared to those in the later stages of primary infection (47 versus 23 sequences, respectively; $P < 0.01$). However, we estimated that there is 90% likelihood that Tat quasispecies would have been missed if it comprised <10% of the virus population when sequencing 23 single genomes and <5% of the population when sequencing 47 genomes. Conversely, we are confident that we were able to isolate quasispecies from an average of >90% of the circulating virus in the 20 patients with primary HIV-1C infection. It should be noted that a constraint exists due to the overlapping nature of nonstructural genes, making it difficult to assign specific mutational events to specific proteins. The overlap between the first exon of *tat* and *vpu* or *rev* was assessed for evidence of positive selection: although there were instances of an amino acid change observed in the *vpu* or *rev* reading frames, no evidence of positive selection was found in these genes using the same MEME analysis as described for *tat*. Furthermore, the second exon of *tat*, which inhibits splicing (85–87), has been shown to be involved in the cellular uptake of exogenous Tat protein (88),

and may contribute to viral infectivity and to binding to cell surface integrins and have a critical role in activating NF-κB (89–92).

Other than associations with previous reported immunodominant regions in HIV-1 subtype C Tat (74, 93–98), the HLA data presented here were predicted based on searching patients' Tat sequences for all known HLA anchor residue motifs and associating the findings with patient HLA genotypes. Within five unique epitopes identified for subtype C *tat* exon 1 and associated with a particular HLA genotype, there is a clear disparity with the extent of data available for subtype B, which has to be addressed.

Subtype C Tat has been established as a more potent transactivator than its subtype B counterpart, yet it is functionally deficient in mediating chemotaxis and cytokine dysregulation in monocytes, a key aspect of HIV pathogenesis maintained by other subtypes (28, 83). Our results suggest that genetic and functional differences are also apparent between primary and chronic infection, with specific residues under evolutionary pressure. The significant, albeit nonlinear, relationship between Tat transactivation and plasma viral load during the first 500 days postseroconversion emphasize its role as a major mediator of virus replication in early infection. Increased understanding of the evolution and dynamics of HIV-1C Tat is crucial for Tat-based vaccine advancement. The functional effects highlight a complex interplay between individual viral mutations in early HIV-1C infection.

ACKNOWLEDGMENTS

REFERENCES

1. **United Nations Programme on HIV/AIDS (UNAIDS).** 2011. Report on the global AIDS epidemic. Joint United Nations Programme on HIV/AIDS (UNAIDS), Geneva, Switzerland.
2. **Lingappa JR, Hughes JP, Wang RS, Baeten JM, Celum C, Gray GE, Stevens WS, Donnell D, Campbell MS, Farquhar C, Essex M, Mullins JI, Coombs RW, Rees H, Corey L, Wald A.** 2010. Estimating the impact of plasma HIV-1 RNA reductions on heterosexual HIV-1 transmission risk. PLoS One **5**:e12598. doi:10.1371/journal.pone.0012598.
3. **Modjarrad K, Chamot E, Vermund SH.** 2008. Impact of small reductions in plasma HIV RNA levels on the risk of heterosexual transmission and disease progression. AIDS **22**:2179–2185.
4. **Romani B, Engelbrecht S, Glashoff RH.** 2010. Functions of Tat: the

versatile protein of human immunodeficiency virus type 1. J. Gen. Virol. **91**:1–12.

5. **Li JC, Lee DC, Cheung BK, Lau AS.** 2005. Mechanisms for HIV Tat upregulation of IL-10 and other cytokine expression: kinase signaling and PKR-mediated immune response. FEBS Lett. **579**:3055–3062.

6. **Matsui M, Warburton RJ, Cogswell PC, Baldwin AS, Jr, Frelinger JA.** 1996. Effects of HIV-1 Tat on expression of HLA class I molecules. J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. **11**:233–240.

7. **Zheng L, Yang YD, Lu GC, Salvato MS.** 2005. Extracellular HIV Tat and Tat cysteine rich peptide increase CCR5 expression in monocytes. J. Zhejiang Univ. Sci. B **6**:668–672.

8. **Della Chiara G, Crotti A, Liboi E, Giacca M, Poli G, Lusic M.** 2011. Negative regulation of HIV-1 transcription by a heterodimeric NF-kappaB1/p50 and C-terminally truncated STAT5 complex. J. Mol. Biol. **410**:933–943.

9. **Duverger A, Jones J, May J, Bibollet-Ruche F, Wagner FA, Cron RQ, Kutsch O.** 2009. Determinants of the establishment of human immunodeficiency virus type 1 latency. J. Virol. **83**:3078–3093.

10. **Williams SA, Chen LF, Kwon H, Ruiz-Jarabo CM, Verdin E, Greene WC.** 2006. NF-kappaB p50 promotes HIV latency through HDAC recruitment and repression of transcriptional initiation. EMBO J. **25**:139–149.

11. **Feng S, Holland EC.** 1988. HIV-1 tat trans-activation requires the loop sequence within tar. Nature **334**:165–167.

12. **Wei P, Garber ME, Fang SM, Fischer WH, Jones KA.** 1998. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. Cell **92**:451–462.

13. **Deng L, de la Fuente C, Fu P, Wang L, Donnelly R, Wade JD, Lambert P, Li H, Lee CG, Kashanchi F.** 2000. Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones. Virology **277**:278–295.

14. **Benkirane M, Chun RF, Xiao H, Ogryzko VV, Howard BH, Nakatani Y, Jeang KT.** 1998. Activation of integrated provirus requires histone acetyltransferase. p300 and P/CAF are coactivators for HIV-1 Tat. J. Biol. Chem. **273**:24898–24905.

15. **Hottiger MO, Nabel GJ.** 1998. Interaction of human immunodeficiency virus type 1 Tat with the transcriptional coactivators p300 and CREB binding protein. J. Virol. **72**:8252–8256.

16. **Marzio G, Tyagi M, Gutierrez MI, Giacca M.** 1998. HIV-1 tat transactivator recruits p300 and CREB-binding protein histone acetyltransferases to the viral promoter. Proc. Natl. Acad. Sci. U. S. A. **95**:13519–13524.

17. **Furia B, Deng L, Wu K, Baylor S, Kehn K, Li H, Donnelly R, Coleman T, Kashanchi F.** 2002. Enhancement of nuclear factor-kappa B acetylation by coactivator p300 and HIV-1 Tat proteins. J. Biol. Chem. **277**:4973–4980.

18. **Ott M, Schnolzer M, Garnica J, Fischle W, Emiliani S, Rackwitz HR, Verdin E.** 1999. Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity. Curr. Biol. **9**:1489–1492.

19. **Zhu Y, Pe'ery T, Peng J, Ramanathan Y, Marshall N, Marshall T, Amendt B, Mathews MB, Price DH.** 1997. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. Genes Dev. **11**:2622–2632.

20. **Zhou Q, Yik JH.** 2006. The yin and yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation. Microbiol. Mol. Biol. Rev. **70**:646–659.

21. **Demarchi F, Gutierrez MI, Giacca M.** 1999. Human immunodeficiency virus type 1 tat protein activates transcription factor NF-kappaB through the cellular interferon-inducible, double-stranded RNA-dependent protein kinase, PKR. J. Virol. **73**:7080–7086.

22. **Rossi A, Mukerjee R, Ferrante P, Khalili K, Amini S, Sawaya BE.** 2006. Human immunodeficiency virus type 1 Tat prevents dephosphorylation of Sp1 by TCF-4 in astrocytes. J. Gen. Virol. **87**:1613–1623.

23. **Roof P, Ricci M, Genin P, Montano MA, Essex M, Wainberg MA, Gatignol A, Hiscott J.** 2002. Differential regulation of HIV-1 clade-specific B, C, and E long terminal repeats by NF-kappaB and the Tat transactivator. Virology **296**:77–83.

24. **Bachu M, Yalla S, Asokan M, Verma A, Neogi U, Sharma S, Murali RV, Mukthey AB, Bhatt R, Chatterjee S, Rajan RE, Cheedarla N, Yadavalli VS, Mahadevan A, Shankar SK, Rajagopalan N, Shet A, Saravanan S, Balakrishnan P, Solomon S, Vajpayee M, Satish KS, Kundu TK, Jeang KT, Ranga U.** 2012. Multiple NF-kappaB sites in HIV-1 subtype C long

terminal repeat confer superior magnitude of transcription and thereby the enhanced viral predominance. J. Biol. Chem. **287**:44714–44735.

25. **Montano MA, Nixon CP, Ndung'u T, Bussmann H, Novitsky VA, Dickman D, Essex M.** 2000. Elevated tumor necrosis factor-alpha activation of human immunodeficiency virus type 1 subtype C in Southern Africa is associated with an NF-kappaB enhancer gain-of-function. J. Infect. Dis. **181**:76–81.

26. **Montano MA, Novitsky VA, Blackard JT, Cho NL, Katzenstein DA, Essex M.** 1997. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J. Virol. **71**:8657–8665.

27. **van Opijnen T, Jeeninga RE, Boerlijst MC, Pollakis GP, Zetterberg V, Salminen M, Berkhout B.** 2004. Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. J. Virol. **78**:3675–3683.

28. **Desfosses Y, Solis M, Sun Q, Grandvaux N, Van Lint C, Burny A, Gatignol A, Wainberg MA, Lin R, Hiscott J.** 2005. Regulation of human immunodeficiency virus type 1 gene expression by clade-specific Tat proteins. AIDS Res. Hum. Retroviruses **79**:9180–9191.

29. **Kandathil AJ, Kannangai R, Abraham OC, Pulimood SA, Sridharan G.** 2009. Amino acid sequence divergence of Tat protein (exon1) of subtype B and C HIV-1 strains: does it have implications for vaccine development? Bioinformation **4**:237–241.

30. **Neogi U, Gupta S, Sahoo PN, Shet A, Rao SD, Ranga U, Prasad VR.** 2012. Genetic characterization of HIV type 1 Tat exon 1 from a Southern Indian clinical cohort: identification of unique epidemiological signature residues. AIDS Res. Hum. Retroviruses **28**:1152–1156.

31. **Clark SJ, Saag MS, Decker WD, Campbell-Hill S, Roberson JL, Veldkamp PJ, Kappes JC, Hahn BH, Shaw GM.** 1991. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. N. Engl. J. Med. **324**:954–960.

32. **Daar ES, Moudgil T, Meyer RD, Ho DD.** 1991. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. N. Engl. J. Med. **324**:961–964.

33. **Kaufmann GR, Cunningham P, Kelleher AD, Zaunders J, Carr A, Vizzard J, Law M, Cooper DA.** 1998. Patterns of viral dynamics during primary human immunodeficiency virus type 1 infection. The Sydney Primary HIV Infection Study Group. J. Infect. Dis. **178**:1812–1815.

34. **Mellors JW, Kingsley LA, Rinaldo CR, Jr, Todd JA, Hoo BS, Kokka RP, Gupta P.** 1995. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. Ann. Intern. Med. **122**:573–579.

35. **Novitsky V, Ndung'u T, Wang R, Bussmann H, Chonco F, Makhema J, De Gruttola V, Walker BD, Essex M.** 2011. Extended high viremics: a substantial fraction of individuals maintain high plasma viral RNA levels after acute HIV-1 subtype C infection. AIDS **25**:1515–1522.

36. **Novitsky V, Wang R, Bussmann H, Lockman S, Baum M, Shapiro R, Thior I, Wester C, Wester CW, Ogwu A, Asmelash A, Musonda R, Campa A, Moyo S, van Widenfelt E, Mine M, Moffat C, Mmalane M, Makhema J, Marlink R, Gilbert P, Seage GR, III, DeGruttola V, Essex M.** 2010. HIV-1 subtype C-infected individuals maintaining high viral load as potential targets for the "test-and-treat" approach to reduce HIV transmission. PLoS One **5**:e10148. doi:10.1371/journal.pone.0010148.

37. **Yukl S, Pillai S, Li P, Chang K, Pasutti W, Ahlgren C, Havlir D, Strain M, Gunthard H, Richman D, Rice AP, Daar E, Little S, Wong JK.** 2009. Latently-infected CD4+ T cells are enriched for HIV-1 Tat variants with impaired transactivation activity. Virology **387**:98–108.

38. **Addo MM, Altfeld M, Rosenberg ES, Eldridge RL, Philips MN, Habeeb K, Khatri A, Brander C, Robbins GK, Mazzara GP, Goulder PJ, Walker BD.** 2001. The HIV-1 regulatory proteins Tat and Rev are frequently targeted by cytotoxic T lymphocytes derived from HIV-1-infected individuals. Proc. Natl. Acad. Sci. U. S. A. **98**:1781–1786.

39. **Allen TM, O'Connor DH, Jing P, Dzuris JL, Mothe BR, Vogel TU, Dunphy E, Liebl ME, Emerson C, Wilson N, Kunstman KJ, Wang X, Allison DB, Hughes AL, Desrosiers RC, Altman JD, Wolinsky SM, Sette A, Watkins DI.** 2000. Tat-specific cytotoxic T lymphocytes select for SIV escape variants during resolution of primary viraemia. Nature **407**:386–390.

40. **Campbell GR, Loret EP.** 2009. What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? Retrovirology **6**:50. doi:10.1186/1742-4690-6-50.

41. **Vider-Shalit T, Almani M, Sarid R, Louzoun Y.** 2009. The HIV hide and seek game: an immunogenomic analysis of the HIV epitope repertoire. AIDS **23**:1311–1318.

42. **Rezza G, Fiorelli V, Dorrucci M, Ciccozzi M, Tripiciano A, Scoglio A,**

Collacchi B, Ruiz-Alvarez M, Giannetto C, Caputo A, Tomasoni L, Castelli F, Sciandra M, Sinicco A, Ensoli F, Butto S, Ensoli B. 2005. The presence of anti-Tat antibodies is predictive of long-term nonprogression to AIDS or severe immunodeficiency: findings in a cohort of HIV-1 seroconverters. J. Infect. Dis. 191:1321–1324.

43. Rodriguez SK, Sarr AD, Olorunnipa O, Popper SJ, Gueye-Ndiaye A, Traore I, Dia MC, Mboup S, Kanki PJ. 2006. The absence of anti-Tat antibodies is associated with risk of disease progression in HIV-2 infection. J. Infect. Dis. 194:760–763.

44. Allard SD, De Keersmaecker B, de Goede AL, Verschuren EJ, Koetsveld J, Reedijk ML, Wylock C, De Bel AV, Vandeloo J, Pistoor F, Heirman C, Beyer WE, Eilers PH, Corthals J, Padmos I, Thielemans K, Osterhaus AD, Lacor P, van der Ende ME, Aerts JL, van Baalen CA, Gruters RA. 2012. A phase I/IIa immunotherapy trial of HIV-1-infected patients with Tat, Rev and Nef expressing dendritic cells followed by treatment interruption. Clin. Immunol. 142:252–268.

45. Ensoli B, Bellino S, Tripiciano A, Longo O, Francavilla V, Marcotullio S, Cafaro A, Picconi O, Paniccia G, Scoglio A, Arancio A, Ariola C, Ruiz Alvarez MJ, Campagna M, Scaramuzzi D, Iori C, Esposito R, Mussini C, Ghinelli F, Sighinolfi L, Palamara G, Latini A, Angarano G, Ladisa N, Soscia F, Mercurio VS, Lazzarin A, Tambussi G, Visintini R, Mazzotta F, Di Pietro M, Galli M, Rusconi S, Carosi G, Torti C, Di Perri G, Bonora S, Ensoli F, Garaci E. 2010. Therapeutic immunization with HIV-1 Tat reduces immune activation and loss of regulatory T-cells and improves immune function in subjects on HAART. PLoS One 5:e13540. doi:10.1371/journal.pone.0013540.

46. Ensoli B, Fiorelli V, Ensoli F, Lazzarin A, Visintini R, Narciso P, Di Carlo A, Tripiciano A, Longo O, Bellino S, Francavilla V, Paniccia G, Arancio A, Scoglio A, Collacchi B, Ruiz Alvarez MJ, Tambussi G, Tassan Din C, Palamara G, Latini A, Antinori A, D'Offizi G, Giuliani M, Giulianelli M, Carta M, Monini P, Magnani M, Garaci E. 2009. The preventive phase I trial with the HIV-1 Tat-based vaccine. Vaccine 28:371–378.

47. Novitsky V, Lagakos S, Herzig M, Bonney C, Kebaabetswe L, Rossenkhan R, Nkwe D, Margolin L, Musonda R, Moyo S, Woldegabriel E, van Widenfelt E, Makhema J, Essex M. 2009. Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. Virology 383:47–59.

48. Novitsky V, Woldegabriel E, Wester C, McDonald E, Rossenkhan R, Ketunuti M, Makhema J, Seage GR, III, Essex M. 2008. Identification of primary HIV-1C infection in Botswana. AIDS Care 20:806–811.

49. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH, Busch MP. 2003. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS 17:1871–1879.

50. Rossenkhan R, Novitsky V, Sebunya TK, Musonda R, Gashe BA, Essex M. 2011. Viral diversity and diversification of major non-structural genes vif, vpr, vpu, tat exon 1 and rev exon 1 during primary HIV-1 subtype C infection. PLoS One 7:e35491. doi:10.1371/journal.pone.0035491.

51. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, Davidov O, Zhao LP, Mullins JI. 1996. HIV quasispecies and resampling. Science 273:415–416.

52. Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

53. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

54. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

55. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8:e1002764. doi:10.1371/journal.pgen.1002764.

56. Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.

57. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. 28:3033–3043.

58. Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

59. Muse SV, Gaut BS. 1994. A likelihood approach for comparing synony-

60. mous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11:715–724.

60. Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14:33–38, 27–28.

61. Novitsky V, Wang R, Margolin L, Baca J, Kebaabetswe L, Rossenkhan R, Bonney C, Herzig M, Nkwe D, Moyo S, Musonda R, Woldegabriel E, van Widenfelt E, Makhema J, Lagakos S, Essex M. 2009. Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection. PLoS One 4:e7727. doi:10.1371/journal.pone.0007727.

62. Listgarten J, Brumme Z, Kadie C, Xiaojiang G, Walker B, Carrington M, Goulder P, Heckerman D. 2008. Statistical resolution of ambiguous HLA typing data. PLoS Comput. Biol. 4:e1000016. doi:10.1371/journal.pcbi.1000016.

63. Heckerman D, Kadie C, Listgarten J. 2007. Leveraging information across HLA alleles/supertypes improves epitope prediction. J. Comput. Biol. 14:736–746.

64. Derdeyn CA, Decker JM, Sfakianos JN, Wu X, O'Brien WA, Ratner L, Kappes JC, Shaw GM, Hunter E. 2000. Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120. J. Virol. 74:8358–8367.

65. Platt EJ, Bilska M, Kozak SL, Kabat D, Montefiori DC. 2009. Evidence that ecotropic murine leukemia virus contamination in TZM-bl cells does not affect the outcome of neutralizing antibody assays with human immunodeficiency virus type 1. J. Virol. 83:8289–8292.

66. Platt EJ, Wehrly K, Kuhmann SE, Chesebro B, Kabat D. 1998. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophage-etropic isolates of human immunodeficiency virus type 1. J. Virol. 72:2855–2864.

67. Takeuchi Y, McClure MO, Pizzato M. 2008. Identification of gamma-retroviruses constitutively released from cell lines used for human immunodeficiency virus research. J. Virol. 82:12585–12588.

68. Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, Saag MS, Wu X, Shaw GM, Kappes JC. 2002. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob. Agents Chemother. 46:1896–1905.

69. D'Agostino R, Pearson ES. 1973. Tests for departure from normality. Empirical results for the distributions of b^2 and $\sqrt{b^1}$. Biometrika 60:613–622.

70. Benjamini Y, Hochberg Y. 1995. Controlling false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57:289–300.

71. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JA, Swanstrom R, Haynes BF, Athreya GS, Korber BT, Sharp PM, Shaw GM, Hahn BH. 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J. Virol. 82:3952–3970.

72. Dey SS, Xue Y, Joachimiak MP, Friedland GD, Burnett JC, Zhou Q, Arkin AP, Schaffer DV. 2012. Mutual information analysis reveals co-evolving residues in Tat that compensate for two distinct functions in HIV-1 gene expression. J. Biol. Chem. 287:7945–7955.

73. Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. 2010. Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature 465:747–751.

74. Novitsky V, Cao H, Rybak N, Gilbert P, McLane MF, Gaolekwe S, Peter T, Thior I, Ndung'u T, Marlink R, Lee TH, Essex M. 2002. Magnitude and frequency of cytotoxic T-lymphocyte responses: identification of immunodominant regions of human immunodeficiency virus type 1 subtype C. J. Virol. 76:10155–10168.

75. Dormeyer W, Dorr A, Ott M, Schnolzer M. 2003. Acetylation of the HIV-1 Tat protein: an in vitro study. Anal. Bioanal. Chem. 376:994–1005.

76. Pagans S, Kauder SE, Kaehlcke K, Sakane N, Schroeder S, Dormeyer W, Trievel RC, Verdin E, Schnolzer M, Ott M. 2010. The cellular lysine methyltransferase Set7/9-KMT7 binds HIV-1 TAR RNA, monomethylates the viral transactivator Tat, and enhances HIV transcription. Cell Host Microbe 7:234–244.

77. Van Duyne R, Easley R, Wu W, Berro R, Pedati C, Klase Z, Kehn-Hall K, Flynn EK, Symer DE, Kashanchi F. 2008. Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. Retrovirology 5:40. doi:10.1186/1742-4690-5-40.

78. Xie B, Invernizzi CF, Richard S, Wainberg MA. 2007. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by

PRMT6 negatively affects Tat interactions with both cyclin T1 and the Tat transactivation region. J. Virol. **81**:4226–4234.

79. **Deng L, Wang D, de la Fuente C, Wang L, Li H, Lee CG, Donnelly R, Wade JD, Lambert P, Kashanchi F.** 2001. Enhancement of the p300 HAT activity by HIV-1 Tat on chromatin DNA. Virology **289**:312–326.

80. **Ammosova T, Berro R, Jerebtsova M, Jackson A, Charles S, Klase Z, Southerland W, Gordeuk VR, Kashanchi F, Nekhai S.** 2006. Phosphorylation of HIV-1 Tat by CDK2 in HIV-1 transcription. Retrovirology **3**:78. doi:10.1186/1742-4690-3-78.

81. **Endo-Munoz L, Warby T, Harrich D, McMillan NA.** 2005. Phosphorylation of HIV Tat by PKR increases interaction with TAR RNA and enhances transcription. Virol. J. **2**:17. doi:10.1186/1743-422X-2-17.

82. **Rice AP, Carlotti F.** 1990. Mutational analysis of the conserved cysteine-rich region of the human immunodeficiency virus type 1 Tat protein. J. Virol. **64**:1864–1868.

83. **Ranga U, Shankarappa R, Siddappa NB, Ramakrishna L, Nagendran R, Mahalingam M, Mahadevan A, Jayasuryan N, Satishchandra P, Shankar SK, Prasad VR.** 2004. Tat protein of human immunodeficiency virus type 1 subtype C strains is a defective chemokine. J. Virol. **78**:2586–2590.

84. **Wong JK, Campbell GR, Spector SA.** 2010. Differential induction of interleukin-10 in monocytes by HIV-1 clade B and clade C Tat proteins. J. Biol. Chem. **285**:18319–18325.

85. **Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM.** 1994. Presence of negative and positive cis-acting RNA splicing elements within and flanking the first tat coding exon of human immunodeficiency virus type 1. Mol. Cell. Biol. **14**:3960–3970.

86. **Amendt BA, Si ZH, Stoltzfus CM.** 1995. Presence of exon splicing silencers within human immunodeficiency virus type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. Mol. Cell. Biol. **15**:4606–4615.

87. **Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C.** 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. J. Biol. Chem. **276**:40464–40475.

88. **Ma M, Nath A.** 1997. Molecular determinants for cellular uptake of Tat protein of human immunodeficiency virus type 1 in brain cells. J. Virol. **71**:2495–2499.

89. **Barillari G, Sgadari C, Fiorelli V, Samaniego F, Colombini S, Manzari V, Modesti A, Nair BC, Cafaro A, Sturzl M, Ensoli B.** 1999. The Tat protein of human immunodeficiency virus type-1 promotes vascular cell growth and locomotion by engaging the alpha5beta1 and alphavbeta3 integrins and by mobilizing sequestered basic fibroblast growth factor. Blood **94**:663–672.

90. **Fiorelli V, Barillari G, Toschi E, Sgadari C, Monini P, Sturzl M, Ensoli B.** 1999. IFN-gamma induces endothelial cells to proliferate and to invade the extracellular matrix in response to the HIV-1 Tat protein: implications for AIDS-Kaposi's sarcoma pathogenesis. J. Immunol. **162**:1165–1170.

91. **Mahlknecht U, Dichamp I, Varin A, Van Lint C, Herbein G.** 2008. NF-kappaB-dependent control of HIV-1 transcription by the second coding exon of Tat in T cells. J. Leukoc. Biol. **83**:718–727.

92. **Smith SM, Pentlicky S, Klase Z, Singh M, Neuveut C, Lu CY, Reitz MS, Jr, Yarchoan R, Marx PA, Jeang KT.** 2003. An in vivo replication-important function in the second coding exon of Tat is constrained against mutation despite cytotoxic T lymphocyte selection. J. Biol. Chem. **278**:44816–44825.

93. **Chopera DR, Mlotshwa M, Woodman Z, Mlisana K, de Assis Rosa D, Martin DP, Abdool Karim S, Gray CM, Williamson C.** 2011. Virological and immunological factors associated with HIV-1 differential disease progression in HLA-B 58:01-positive individuals. J. Virol. **85**:7070–7080.

94. **De Groot AS, Rivera DS, McMurry JA, Buus S, Martin W.** 2008. Identification of immunogenic HLA-B7 "Achilles' heel" epitopes within highly conserved regions of HIV. Vaccine **26**:3059–3071.

95. **Frahm N, Kiepiela P, Adams S, Linde CH, Hewitt HS, Sango K, Feeney ME, Addo MM, Lichterfeld M, Lahaie MP, Pae E, Wurcel AG, Roach T, St John MA, Altfeld M, Marincola FM, Moore C, Mallal S, Carrington M, Heckerman D, Allen TM, Mullins JI, Korber BT, Goulder PJ, Walker BD, Brander C.** 2006. Control of human immunodeficiency virus replication by cytotoxic T lymphocytes targeting subdominant epitopes. Nat. Immunol. **7**:173–178.

96. **Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, Rathnavalu P, Moore C, Pfafferott KJ, Hilton L, Zimbwa P, Moore S, Allen T, Brander C, Addo MM, Altfeld M, James I, Mallal S, Bunce M, Barber LD, Szinger J, Day C, Klenerman P, Mullins J, Korber B, Coovadia HM, Walker BD, Goulder PJ.** 2004. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature **432**:769–775.

97. **Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, Reddy S, de Pierres C, Mncube Z, Mkhwanazi N, Bishop K, van der Stok M, Nair K, Khan N, Crawford H, Payne R, Leslie A, Prado J, Prendergast A, Frater J, McCarthy N, Brander C, Learn GH, Nickle D, Rousseau C, Coovadia H, Mullins JI, Heckerman D, Walker BD, Goulder P.** 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nat. Med. **13**:46–53.

98. **Rousseau CM, Daniels MG, Carlson JM, Kadie C, Crawford H, Prendergast A, Matthews P, Payne R, Rolland M, Raugi DN, Maust BS, Learn GH, Nickle DC, Coovadia H, Ndung'u T, Frahm N, Brander C, Walker BD, Goulder PJ, Bhattacharya T, Heckerman DE, Korber BT, Mullins JI.** 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.