

Transmitted/Founder HIV-1 Subtype C Viruses Show Distinctive Signature Patterns in Vif, Vpr, and Vpu That Are Under Subsequent Immune Pressure During Early Infection

Raabya Rossenkhan,¹⁻⁴ Iain J. MacLeod,^{1,2} Zabrina L. Brumme,^{5,6} Craig A. Magaret,⁴ Theresa K. Sebnnya,³ Rosemary Musonda,^{1,2} Berhanu A. Gashe,³ Paul T. Edlefsen,⁴ Vlad Novitsky,^{1,2} and M. Essex^{1,2}

Abstract

Viral variants that predominate during early infection may exhibit constrained diversity compared with those found during chronic infection and could contain amino acid signature patterns that may enhance transmission, establish productive infection, and influence early events that modulate the infection course. We compared amino acid distributions in 17 patients recently infected with HIV-1C with patients with chronic infection. We found significantly lower entropy in inferred transmitted/founder (t/f) compared with chronic viruses and identified signature patterns in Vif and Vpr from inferred t/f viruses. We investigated sequence evolution longitudinally up to 500 days postseroconversion and compared the impact of selected substitutions on predicted human leukocyte antigen (HLA) binding affinities of published and predicted cytotoxic T-lymphocyte epitopes. Polymorphisms in Vif and Vpr during early infection occurred more frequently at epitope-HLA anchor residues and significantly decreased predicted epitope-HLA binding. Transmission-associated sequence signatures may have implications for novel strategies to prevent HIV-1 transmission.

Keywords: HIV transmission, accessory genes, Vif, Vpr, Vpu, CTL epitope, evolution

Introduction

THE SEVERE GENETIC bottleneck during heterosexual HIV acquisition frequently leads to the establishment of productive infection by a single transmitted/founder (t/f) virus.¹⁻⁵ Consistent amino acid patterns or signatures present in t/f viruses and their descendants during early infection support the existence of viral genetic characteristics associated with increased transmission fitness,^{6,7} including increased transmissibility across genital mucosa, enhanced infectivity of target cells, or increased replication in newly infected cells, to allow the preferential establishment by that variant.⁶ Although the establishment of infection may be stochastic to some extent,⁸ the existence of transmission signatures has been demonstrated, for example, in gp120/gp41 where reduced N-linked glycosylation and variable region lengths are more frequently observed in viral variants found in early compared with chronic infection.^{2,9} Isolates

from acutely infected HIV-1 subjects are also more resistant to interferon alpha suppression than isolates found during chronic infection.¹⁰

Following transmission, within-host immune selection pressures drive HIV-1 evolution. The t/f virus initially rapidly diversifies in a characteristic star-like phylogenetic shape. Following this, cytotoxic T lymphocytes (CTLs) play a critical role in the control of HIV infection and mediate another bottleneck event from which escape variants emerge.¹¹ Human leukocyte antigen (HLA) class I-mediated CTL responses are important predictors of disease progression¹² that drive virus escape mutations that are highly adapted to the host environment.^{4,13-18} Recent studies employing ultradeep sequencing^{3,19} confirmed dramatic shifts in the frequencies of epitope variants during the first weeks of HIV infection.

HIV-1 accessory proteins, Vif, Vpr, and Vpu, modify the host cell environment and disable innate antiretroviral

¹Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

²Botswana Harvard AIDS Institute, Gaborone, Botswana.

³Department of Biological Sciences, University of Botswana, Gaborone, Botswana.

⁴Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, Washington.

⁵Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada.

⁶British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada.

defenses. As such, the transmission bottleneck may select for variants that more readily establish productive infection. However, limited data exist on amino acid signature patterns within viral proteins, particularly in HIV-1 subtype C t/f viruses.²⁰ Moreover, limited data exist on the timing of escape mutations in HIV-1 subtype C accessory genes.^{16,18} In this study, we reconstructed the t/f viruses derived from the earliest available sequences from 17 individuals with acute/early HIV-1C infection. To determine if, and to what extent, Vif, Vpr, and Vpu signature patterns are present in HIV-1C, we compared these t/f viruses to an independent dataset of chronic HIV-1C infection sequences. We also examined longitudinal sequence changes within published and predicted HLA-restricted CTL epitopes in Vif, Vpr, and Vpu among 17 early infected patients up to 500 days postseroconversion (p/s) to investigate early immune escape events.

Materials and Methods

Study population

Patients were enrolled in an HIV-1C primary infection cohort in Botswana,^{21,22} and a subset of 17 subjects was selected based on the stage of HIV infection: six acutely infected individuals (patient code A to H) and 11 randomly selected early infections (patient code OC to QR) (Table 1). Acutely infected individuals were identified before seroconversion by a positive HIV-1 reverse transcription polymerase chain reaction test accompanied by negative HIV-1 serology (Fiebig stage II).^{23,24} Seroconverters were identified within early stage of HIV-1 infection (Fiebig stages IV–V). 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 midpoint of the corresponding Fiebig stage for the recently infected subjects.²¹ Written informed consent was obtained from all study participants, and ethics 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 was carried out using the QIAamp viral RNA Mini kit (Qiagen, Valencia, CA) according to the manufacturer's instructions, followed by single-genome amplification as described previously.²⁵ 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 with minor modifications.²⁶

Most recent common ancestor reconstruction and phylogenetic relationship of vif, vpr, and vpu sequences

A consensus sequence was determined from the earliest available sample point p/s for each patient to approximate the most recent common ancestor (MRCA). Both DIVEIN²⁷ (earliest time point sequences using Group M outgroup) and earliest consensus MRCA comparisons were made. To assess intrapatient sequence diversity, maximum likelihood (ML) phylogenetic trees were constructed using PhyML²⁸ and the HKY85 nucleotide substitution model. The tree was visualized in FigTree v1.1.2 (<http://tree.bio.ed.ac.uk/software/figtree/>) and Mega 5.²⁹

Chronic Los Alamos National Laboratory sequences used for comparisons

Single-patient HIV-1 subtype C sequences available in the Los Alamos National Laboratory (LANL) HIV-1 database (www.hiv.lanl.gov/) were used for chronic infection comparisons. Chronic HIV-1 subtype C infection sequence selection criteria included (1) HIV-1 infection at the time of sampling >1,000 days postinfection/seroconversion; (2) nonrecombinant HIV-1 subtype C; and (3) full-length vif, vpr, and vpu. This included 187 nonrecombinant HIV-1 subtype C vif and vpr sequences and 197 vpu sequences retrieved from the LANL HIV Database (August 26, 2013). All available subtype C sequences fitting these criteria were included in our analysis. The majority of sequences included

TABLE 1. CHARACTERISTICS OF N=17 ACUTE/EARLY STUDY SUBJECTS OVER THE FIRST 500 DAYS POSTSEROCONVERSION

	Acute (N=6) (patients A to H)	Early (N=11) (patients OC to QU)	p
Median (range) viral load; log ₁₀ RNA copies/ml	4.95 (4.00–5.52)	4.40 (3.84–4.75)	.11
Median (range) CD4 count; cells/ μ l	337 (258–426)	374 (258–490)	.72
Median (range) age; years	31.50 (26.75–34.25)	25.50 (24.75–27.00)	.11
Proportion of females (%)	5/6 (83)	8/11 (73)	1.00
Median Fiebig stage (range)	2 (2–2)	5 (5–6)	<.01
Median number of time points sequenced (range)			
Vif	4.50 (3.00–5.00)	2.00 (2.00–2.00)	
Vpr	4.50 (3.00–5.00)	2.00 (1.00–2.00)	
Vpu	4.00 (3.00–5.00)	2.00 (2.00–2.00)	
Median (range) of follow-up; days	362 (203–381)	279 (179–374)	.59
Median (range) total number of sequences analyzed			
Vif	46 (32–71)	22 (16–27)	
Vpr	45 (33–70)	21 (10–27)	
Vpu	40 (30–63)	21 (14–27)	

South Africa (~50%), followed by Botswana (~25%) and others (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/aid). A maximum likelihood phylogeny inferred from our combined early and chronic HIV-1 *vif* sequences was created to confirm that the acute/early sequences are generally interspersed among the chronic ones (Supplementary Fig. S1).

Measurement of amino acid variability: Shannon entropy

Shannon entropy is a simple measure of local variation in DNA and protein sequence alignments that reflects in a single value both the number of variants and their distribution.³⁰ To calculate the comparative Shannon entropy between the MRCA for Vif, Vpr, and Vpu for each of the 17 patients in our primary infection cohort, and those with chronic HIV-1C infection, we utilized the online tool at the LANL HIV Sequence Database: www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_two.html

Determination of amino acid frequency

Amino acid frequencies for each codon position of Vif, Vpr, and Vpu were calculated using the Viral Epidemiology Signature Pattern Analysis (VESPA) program from the LANL HIV Sequence Database (www.hiv.lanl.gov/content/sequence/VESPA/vespa.html).

Amino acid frequencies in the *t/f* viruses of our early infection cohort were compared with those in chronically infected HIV-1C patients to identify codons where the consensus (most common) amino acid differed between the early and chronic datasets. In cases where the most common amino acid was the same for both groups, we used an exact binomial test to compare those that differed in frequency by ≥ 1 standard deviation of the mean frequency for the dominant amino acids across the whole protein. To correct for multiple hypothesis testing, we used the Benjamini–Hochberg procedure³¹ to compute false discovery rate-adjusted *p*-values (*q*-values) and included all sites with a common major variant across the two groups for purposes of multiplicity adjustment. Sites with unadjusted *p*-value ≤ 0.05 and differing in frequency by ≥ 1 standard deviation of the mean frequency for the dominant amino acids were considered possible signatures of transmitted versus chronic infection. These tests were confirmed using R.³²

HLA typing of major histocompatibility complex class I

High-resolution HLA typing was performed for all study subjects as described previously.³³ Briefly, the AlleleSEQR HLA Sequencing-Based Typing kit (Celera, Alameda, CA) was used according to the manufacturer's instructions. Contig assembly and assignment of HLA alleles were implemented by Assign SBT ver. 3.5.1.42 (Conexio Genomics, Applecross, Australia). Ambiguous positions were resolved by resequencing; if this was not possible, ambiguous types were imputed to high-level resolution using a published statistical method (<http://atom.research.microsoft.com/HLACompletion/>).³⁴

Analysis of known CTL epitopes

Known HLA class I-restricted epitopes within subtype C Vif, Vpr, and Vpu were retrieved from the LANL Im-

munology Database (www.hiv.lanl.gov/content/immunology/ctl_search). Longitudinal analyses of the sequences of these epitopes were undertaken in early infected individuals carrying the known restricting HLA allele.

Prediction of epitope binding affinity

NetMHCpan2.8 was used to predict binding affinities of autologous Vif, Vpr, and Vpu peptides to the corresponding patient's HLA class I molecules (www.cbs.dtu.dk/services/NetMHCpan). We focused on HLA-A and HLA-B for which NetMHCpan2.8 has been shown to outperform alternative methods; HLA-C epitope binding predictions were not considered due to reduced reliability of these results and limited information available.³⁵ Anchor residues for HLA-A and HLA-B were defined according to the LANL Immunology Database (www.hiv.lanl.gov/content/immunology/motif_scan/motif.html).

Briefly, autologous Vif, Vpr, and Vpu sequences were cleaved into all possible peptide combinations of nine amino acids (9-mers) due to the preference of HLA class I molecules for epitopes of this length. For each patient, recurring epitope sequences were excluded to yield only unique peptides. The consensus sequences for Vif, Vpr, and Vpu derived from chronically infected HIV-1C patients were processed in the same manner. The epitope:HLA binding affinity is given as an IC_{50} , the predicted half-maximal concentration of peptide required to bind a given HLA. As such, a low IC_{50} is indicative of high binding affinity. NetMHCpan defines an IC_{50} of 500 nM threshold for weak peptide-HLA interactions, which we used as the cutoff to define a putative epitope for a given HLA allele.

GenBank accession numbers

Sequences from our cohort have been assigned GenBank accession numbers JQ895561–JQ896230.

Results

Clinical characteristics and interpatient viral quasispecies diversity over 500 days *p/s*

Sequences of HIV quasispecies were obtained from 17 individuals with acute/early HIV-1C infection (herein early infection) (Table 1) by single-genome amplification and sequencing. ML phylogenetic trees were constructed to visualize *vif*, *vpr*, and *vpu* intra- and interpatient diversity up to 500 days *p/s*. Patients harbored phylogenetically distinct HIV-1C sequences, with the exception of individuals, OK and OI, who represented a transmission pair (or chain) as described previously (Fig. 1).²⁵ The interpatient *vpu* diversity exceeded that of *vif* and *vpr*, which was consistent with previous observations in primary infection.^{25,36}

Vif and Vpr, but not Vpu, have low entropy in *t/f* viruses

To investigate HIV-1C Vif, Vpr, and Vpr amino acid diversity, we reconstructed each patient's *t/f* virus, using the consensus sequence from the earliest time point postinfection, and computed Shannon entropy scores across Vif, Vpr, and Vpu *t/f* alignments. The median number of sequences per patient used for reconstruction was 10 [interquartile range (IQR): 5–24]. The reconstructed *t/f* viruses exhibited relatively low interpatient amino acid entropy in Vif (median 0.00; IQR: 0.00–0.22) and Vpr (median 0.00; IQR: 0.00–

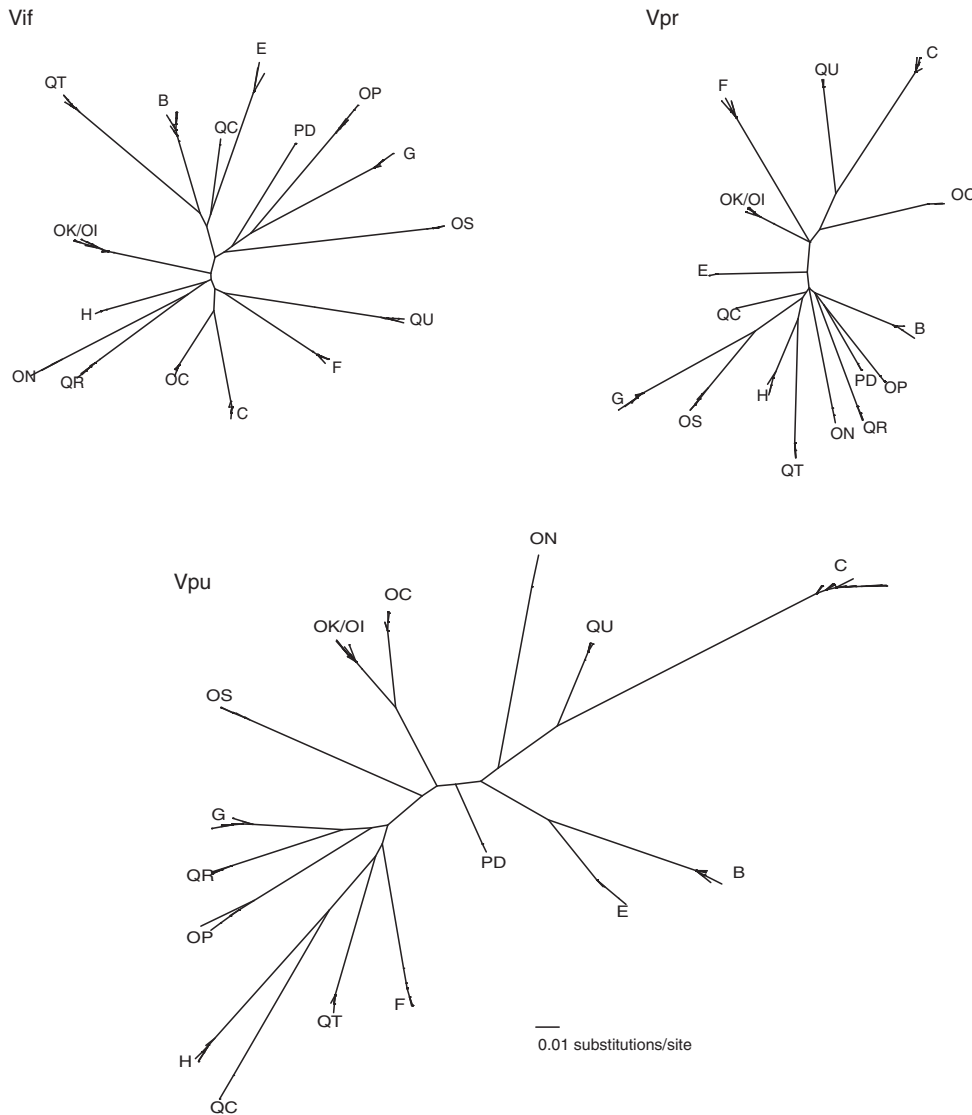


FIG. 1. Phylogenetic relationships of *vif*, *vpr*, and *vpu* sequences from 17 individuals during primary infection up to 500 days p/s. Clusters of patient quasispecies are labeled with patient IDs. Phylogenies drawn on the same distance scale highlight interpatient diversity during early HIV-1C infection and greater intrapatient distance for *vpu*.

0.22) (Fig. 2), suggesting that sequence conservation for these proteins might be important during the establishment of productive HIV-1 infection. In contrast, *t/f* viruses exhibited higher Vpu entropy (median 0.25; IQR: 0.00–0.65), confirming previous reports of substantial Vpu genetic variability.^{37–39}

As HIV-1C structural proteins generally exhibit lower entropy in early compared with chronic infection,¹⁸ we hypothesized that accessory proteins of *t/f* viruses would also demonstrate reduced interpatient variability compared with HIV-1C chronically infected patients. As our cohort was only followed up to 500 days p/s, we used HIV-1C sequences from the LANL HIV database (*Vif* and *Vpr* $n = 187$; *Vpu* $n = 197$), the majority from Southern African countries (Supplementary Table S1), to represent accessory protein diversity in chronic infection, defined as >1,000 days postinfection/seroconversion. When comparing *t/f* and chronic *Vif* amino acid sequences, significantly lower entropy was observed during early infection (0.00 [IQR: 0.00–0.22]) compared with chronic infection (0.07 [IQR: 0.00–0.37]; $p < .001$, Mann-Whitney test) (Fig. 2). Similarly, entropy of chronic *Vpr* was higher (median 0.07 [IQR: 0.03–0.22]) compared with *t/f*

viruses (median 0.00 [IQR: 0.00–0.22]; $p < .001$). Although both *t/f* and chronic *Vpu* sequences exhibited substantial variability, there was a trend toward higher entropy in chronic infection (median 0.33 [IQR: 0.11–0.69]) compared with early infection (median 0.25 [IQR: 0.00–0.65]; $p = .065$).

In *t/f* viruses, 117/192 (61%) of *Vif* codons and 62/96 (65%) *Vpr* codons were completely conserved compared with only 36/82 (44%) of *Vpu* codons. Regardless of the length of infection (early or chronic), conserved amino acids comprised 41% (78/192) of *Vif* and 43% (41/96) of *Vpr*, yet only 20% (17/86) of sites in *Vpu*. The locations of these conserved residues are highlighted in the context of accessory protein functional domains (Fig. 3).

Vif, *Vpr*, and *Vpu* display distinct amino acid signature patterns in *t/f* viruses

The significantly lower interpatient amino acid variability in *t/f* viruses led us to hypothesize that specific genetic transmission signatures might be present within *Vif* and *Vpr*, and possibly *Vpu*. Using the Viral Epidemiology Signature Pattern Analysis (VESPA) program from the LANL HIV

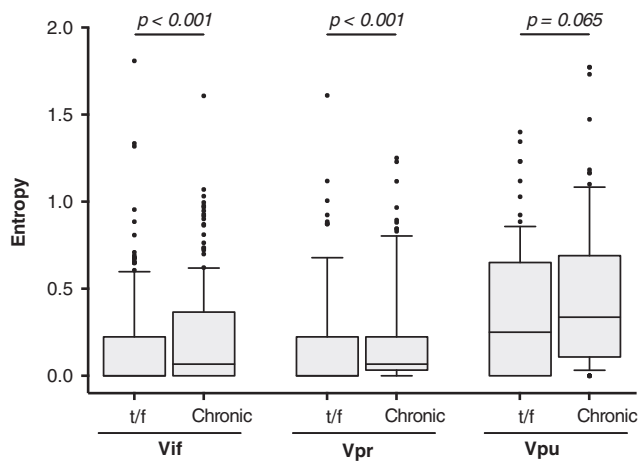


FIG. 2. A comparison of Shannon entropy between accessory proteins in *t/f* viruses compared with chronic infection. Shannon entropy was assessed for full-length protein Vif, Vpr, and Vpu sequences from the *t/f* viruses of our early infection cohort ($n = 17$), and HIV-1C chronically infected patients from LANL HIV database (Vif and Vpr $n = 187$; Vpu $n = 197$), defined as >1,000 days postinfection/seroconversion. The box-and-whisker plot shows the difference in median entropy with 25%–75%-ile (*box*) with median (*line*), 10%–90%-ile (*caps*), and outliers (*black points*). *t/f*, transmitted/founder; LANL, Los Alamos National Laboratory.

Sequence Database, amino acid frequencies in the *t/f* viruses of our early infection cohort were compared with those in chronically infected HIV-1C patients to identify codons where the consensus (most common) amino acid differed between the early and chronic datasets.⁴⁰ In cases where the most common amino acid was the same for both groups, we used an exact binomial test to compare those that differed in frequency by ≥ 1 standard deviation of the mean frequency for the dominant amino acid across the whole protein (Supplementary Table S2).

Signature patterns are highlighted in Figure 3. In Vif, six codons exhibited a different dominant amino acid in *t/f* compared with chronic viruses, of which two codons had significantly lower expression of the amino acid most commonly found in chronic infection: I31V ($p = .029$; $q = 0.26$) and R34K ($p = .004$; $q = 0.24$) (Table 2). We identified an additional 18 codons in Vif whereby the dominant amino acid was the same in the *t/f* viruses as that in chronic infection, yet differed in frequency by ≥ 1 standard deviation from the dominant amino acid frequency mean across Vif in chronic infection (Table 3).

In Vpr of the *t/f* viruses, the frequency of proline at codon 4 was reduced by almost 25% compared with those in chronic infection, although this was not significant ($p = .050$; $q = 0.52$). This was the only residue that expressed a different dominant amino acid in *t/f* compared with chronic viruses

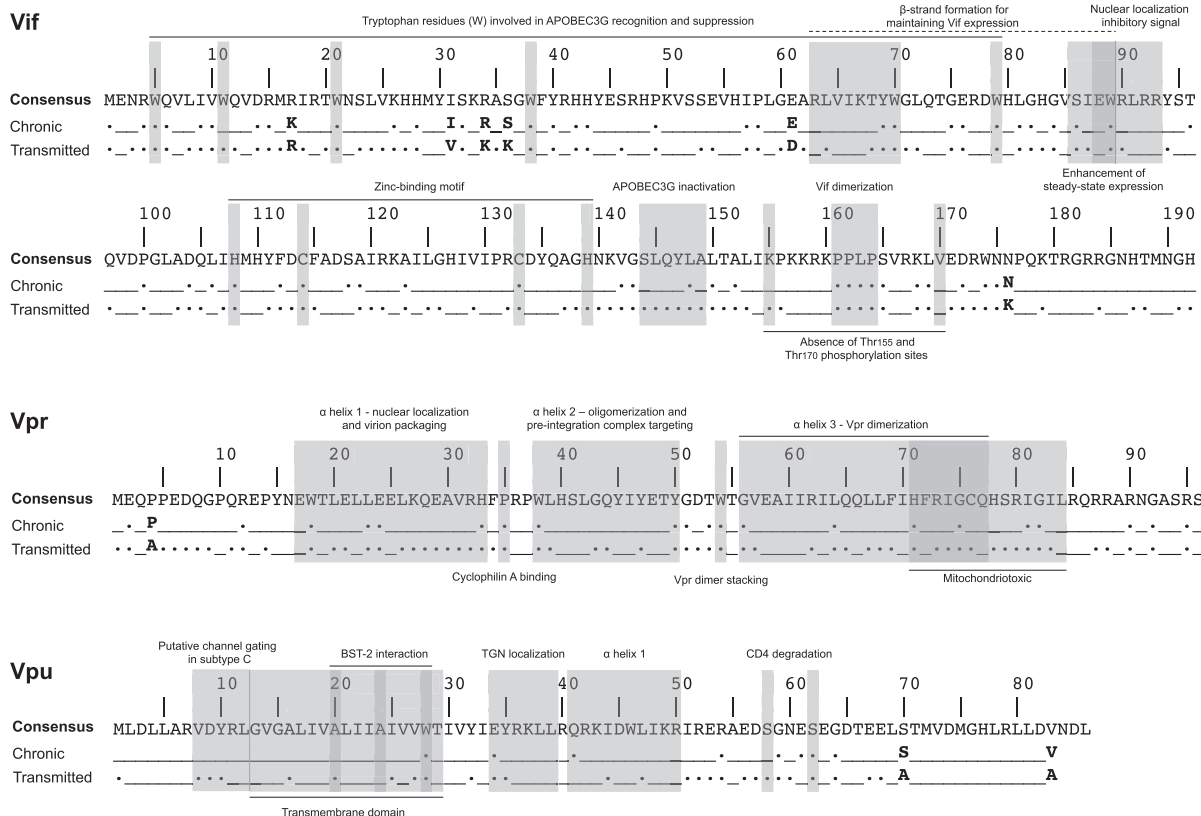


FIG. 3. Location of amino acid signature patterns and conservation of functional domain residues in accessory proteins. The HIV-1 subtype C consensus sequence was derived from www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html and was used as a reference. Conservation of a residue (.) indicates no variability at that position and the presence of the consensus amino acid; (□) indicates a site where one or more sequences in the dataset exhibited polymorphism. Signature patterns, whereby the dominant residue differed between the *t/f* viruses and those in chronic infection, are given as the single-letter amino acid code in *bold*.

TABLE 2. AMINO ACID FREQUENCY OF SIGNATURE PATTERN RESIDUES IN VIF

Codon	Infection stage	Residue frequency	Signature residues	p	q
17	Chronic t/f	K _{0.51} , R _{0.48} , E _{0.01} K _{0.47} , R _{0.53}	K R	.81	1.00
31	Chronic t/f	I _{0.51} , V _{0.43} , T _{0.032} , M _{0.016} A _{<0.01} R _{<0.01} S _{<0.01} V _{0.47} , I _{0.24} , T _{0.18} , G _{0.059} , M _{0.059}	I V	.029	0.26
34	Chronic t/f	K _{0.25} , R _{0.74} Q _{<0.01} S _{<0.01} K _{0.59} , R _{0.41}	R K	.004	0.24
36	Chronic t/f	S _{0.40} , N _{0.26} , K _{0.20} , A _{0.02} , D _{0.02} , G _{0.02} , Q _{0.02} , H _{0.01} , R _{0.01} E _{<0.01} M _{<0.01} T _{<0.01} K _{0.29} , N _{0.18} , S _{0.18} , G _{0.12} , H _{0.12} , Q _{0.06} , R _{0.06}	S K	.081	0.54
61	Chronic t/f	E _{0.55} , D _{0.41} , K _{0.02} , G _{0.02} A _{<0.01} N _{<0.01} D _{0.59} , E _{0.41}	E D	.33	0.92
176	Chronic t/f	N _{0.49} , K _{0.44} , D _{0.03} , E _{0.02} S _{<0.01} I _{<0.01} V _{<0.01} K _{0.65} , N _{0.24} , M _{0.06} , T _{0.06}	N K	.049	0.41

t/f, transmitted/founder.

The **bold** values highlight residues that can more likely be explored as possible transmission signatures that have a *p*-value <0.05 and a *q*-value threshold <0.3 to correct for multiple hypothesis testing, for which we used the Benjamini-Hochberg procedure to compute false discovery rate-adjusted *p*-values (*q*-values).

TABLE 3. RESIDUES IN VIF THAT DIFFERED IN FREQUENCY OF THE DOMINANT AMINO ACID

Codon	Infection stage	Residue frequency	Signature residue	p	q
7	Chronic t/f	V _{0.83} , A _{0.07} , G _{0.09} , R _{0.01} V _{0.94} , G _{0.06}	V	.34	0.92
19	Chronic t/f	R _{0.94} K _{0.06} R _{0.82} K _{0.12}	R	.075	0.86
50	Chronic t/f	K _{0.88} R _{0.12} K _{0.65} R _{0.35}	K	.013	0.24
73	Chronic t/f	Q _{0.82} H _{0.14} L _{0.02} N _{0.02} R _{0.01} Q _{0.71} H _{0.18} L _{0.06} N _{0.06}	Q	.22	0.86
80	Chronic t/f	H _{0.97} Q _{0.03} L _{0.01} H _{0.82} Q _{0.12} Y _{0.06}	H	.016	0.24
93	Chronic t/f	R _{0.69} K _{0.2} G _{0.04} S _{0.04} E _{0.02} N _{0.01} R _{0.41} K _{0.35} S _{0.12} E _{0.06} G _{0.06}	R	.018	0.24
117	Chronic t/f	D _{0.82} E _{0.16} G _{0.02} D _{0.71} E _{0.29}	D	.22	0.86
130	Chronic t/f	I _{0.65} S _{0.2} T _{0.06} F _{0.05} N _{0.02} V _{0.01} Y _{0.01} I _{0.82} A _{0.06} F _{0.06} Y _{0.06}	I	.13	0.86
134	Chronic t/f	D _{0.84} E _{0.13} N _{0.03} V _{0.01} D _{0.59} E _{0.41}	D	.012	0.24
136	Chronic t/f	Q _{0.82} P _{0.18} S _{0.01} Q _{0.94} P _{0.06}	Q	.34	0.92
154	Chronic t/f	I _{0.93} V _{0.04} L _{0.03} I _{0.76} V _{0.18} L _{0.06}	I	.027	0.26
159	Chronic t/f	R _{0.54} I _{0.42} A _{0.01} P _{0.01} T _{0.01} V _{0.01} R _{0.82} I _{0.12} K _{0.06}	R	.026	0.26
160	Chronic t/f	K _{0.99} R _{0.01} T _{0.01} K _{0.88} R _{0.12}	K	.014	0.24
167	Chronic t/f	R _{0.72} K _{0.11} Q _{0.07} S _{0.07} G _{0.01} N _{0.01} T _{0.01} W _{0.01} R _{0.94} K _{0.06}	R	.055	0.42
181	Chronic t/f	R _{0.93} , K _{0.07} R _{0.82} , K _{0.18}	R	.13	0.64
183	Chronic t/f	R _{0.88} H _{0.12} C _{0.01} R _{0.65} H _{0.29} C _{0.06}	R	.013	0.24
186	Chronic t/f	N _{0.91} S _{0.08} C _{0.01} D _{0.01} N _{0.76} S _{0.24}	N	.062	0.44
190	Chronic t/f	N _{0.94} S _{0.06} T _{0.01} N _{0.82} S _{0.12} P _{0.06}	N	.092	0.56

The **bold** values highlight residues that can more likely be explored as possible transmission signatures that have a *p*-value <0.05 and a *q*-value threshold <0.3 to correct for multiple hypothesis testing, for which we used the Benjamini-Hochberg procedure to compute false discovery rate-adjusted *p*-values (*q*-values).

TABLE 4. AMINO ACID FREQUENCY OF SIGNATURE PATTERN RESIDUES IN VPR

Codon	Infection stage	Residue frequency	Signature residues	p	q
4	Chronic t/f	P _{0.49} , A _{0.41} , S _{0.03} , T _{0.03} , V _{0.02} , F _{0.01} , I _{0.01} , L _{0.01} A _{0.65} , P _{0.24} , T _{0.12}	P A	.050	0.52

(Table 4). Of the t/f residues that differed by ≥ 1 standard deviation from the frequency of the dominant amino acid in chronic sequences, L22 was significantly enriched in the t/f viruses ($p=.007$; $q=0.14$) compared with chronic sequences, whereas I70 ($p=.023$; $q=0.28$) and A93 ($p=.005$; $q=0.14$) were significantly underrepresented compared with chronic sequences (Table 5). The only residue in Vpu that differed in the dominant amino acid was S70A, although this was not significant ($p=.21$; $q=1.00$) (Table 6), while no Vpu residue differed significantly in the dominant amino acid between t/f and chronic sequences (Table 7).

Association of Vif, Vpr, and Vpu mutations during early infection with known HLA epitopes

Given the reduced amino acid diversity of Vif, Vpr, and to a lesser degree Vpu in t/f viruses from our early infection cohort, we determined whether intrahost amino acid substitutions that arose in sequences collected up to 500 days p/s were associated with possible CTL responses in known HIV-1C-specific epitopes.⁴¹ Four of the 17 patients exhibited amino acid substitutions within a previously described Vif CTL epitope restricted by their respective class I HLA during the follow-up period (Table 8). By 199 days p/s, a Vif-V98I substitution arose in patient B_2865 in a B*18:01 epitope. Furthermore, patient B_2865 harbored an E88Q substitution in the t/f virus, which persisted during the follow-up period. This substitution was not present in any of the 187 sequences analyzed from chronically infected patients, suggesting that

B*18:01-mediated immune selection may have been responsible for the persistence of this variant. In patient E_3430, an A35T substitution within a known B*15:03 epitope arose in all Vif sequences by day 150 and persisted at least as far as day 372 p/s, at which K33E and F39S substitutions also predominated. Furthermore, the transmitted R34K and S36K variants, which we previously described as being more prevalent in HIV-1C t/f viruses, were maintained throughout the follow-up period in this individual, possibly due to HLA B*15:03 selection pressure.

Toward the end of the follow-up period, patient G_3603 developed E54A and K63I substitutions within a known A*68:01-restricted epitope, which occurred in *cis* as revealed by single-genome sequencing. Both of these amino acids were extremely rare in sequences from chronic infection, with 54A occurring in 1/187 sequences and 63I absent from the chronic dataset altogether. Patient QU_6029 also had a transmitted Vif variant, H56Y, which was extremely rare in chronic infection (1/187 sequences), yet persisted as the dominant variant within a B*15:10-restricted epitope at least as far as 174 days p/s. At that stage, a V51I variant emerged within the same B*15:10 epitope in 20% of the virus population.

Amino acid substitutions arising in early infection significantly reduce Vif and Vpr predicted epitope binding affinity

Due to the limited characterization of HLA-restricted CTL epitopes in HIV-1C accessory proteins, we used NetMHCpan

TABLE 5. RESIDUES IN VPR THAT DIFFERED IN FREQUENCY OF THE DOMINANT AMINO ACID

Codon	Infection stage	Residue frequency	Signature residue	p	q
19	Chronic t/f	A _{0.27} , I _{0.01} , T _{0.72} T _{0.88} , A _{0.12}	A	.180	0.70
22	Chronic t/f	L _{0.49} , I _{0.45} , V _{0.05} , M _{0.01} L _{0.82} , I _{0.18}	L	.007	0.14
41	Chronic t/f	S _{0.5} , G _{0.24} , N _{0.2} , D _{0.04} , A _{0.01} , H _{0.01} , I _{0.01} S _{0.65} , G _{0.35}	S	.332	0.80
55	Chronic t/f	T _{0.79} , A _{0.13} , M _{0.04} , L _{0.01} , P _{0.01} , R _{0.01} , S _{0.01} , I _{0.01} T _{0.65} , A _{0.12} , V _{0.12} , E _{0.06} , I _{0.06}	T	.143	0.70
60	Chronic t/f	I _{0.58} , L _{0.39} , M _{0.02} , V _{0.01} I _{0.76} , L _{0.18} , M _{0.06}	I	.147	0.70
61	Chronic t/f	I _{0.83} , L _{0.08} , M _{0.03} , T _{0.05} , V _{0.01} I _{0.71} , L _{0.18} , T _{0.06} , V _{0.06}	I	.193	0.70
70	Chronic t/f	I _{0.82} , V _{0.11} , T _{0.06} , M _{0.01} I _{0.59} , V _{0.29} , T _{0.12}	I	.023	0.28
84	Chronic t/f	L _{0.63} , I _{0.11} , M _{0.11} , V _{0.07} , W _{0.04} , T _{0.02} , G _{0.01} , S _{0.01} L _{0.47} , I _{0.12} , M _{0.12} , V _{0.12} , A _{0.06} , S _{0.06} , T _{0.06}	L	.209	0.70
93	Chronic t/f	A _{0.9} , S _{0.06} , T _{0.03} , D _{0.01} , P _{0.01} A _{0.65} , S _{0.18} , P _{0.12} , N _{0.06}	A	.005	0.14

The *bold* values highlight residues that can more likely be explored as possible transmission signatures that have a p -value < 0.05 and a q -value threshold < 0.3 to correct for multiple hypothesis testing, for which we used the Benjamini-Hochberg procedure to compute false discovery rate-adjusted p -values (q -values).

TABLE 6. AMINO ACID FREQUENCY OF SIGNATURE PATTERN RESIDUES IN VPU

Codon	Infection stage	Residue frequency	Signature residues	p	q
70	Chronic t/f	A _{0.45} , C _{0.01} , D _{0.01} , E _{0.04} , G _{0.01} , S _{0.50} A _{0.69} , S _{0.31}	S A	.21	1.00

TABLE 7. RESIDUES IN VPU THAT DIFFERED IN FREQUENCY OF THE DOMINANT AMINO ACID

Codon	Infection stage	Residue frequency	Signature residue	p	q
5	Chronic t/f	A _{0.06} , E _{0.01} , F _{0.01} , I _{0.18} , K _{0.01} , L _{0.55} , T _{0.11} , S _{0.02} A _{0.12} , I _{0.12} , L _{0.65} , Q _{0.06} , T _{0.06}	L	.220	1.00
8	Chronic t/f	A _{0.01} , E _{0.01} , I _{0.12} , V _{0.87} V _{1.00}	V	.056	1.00
38	Chronic t/f	I _{0.02} , L _{0.92} , V _{0.05} , W _{0.01} G _{0.06} , L _{0.76} , V _{0.12} , W _{0.06}	L	.038	1.00
46	Chronic t/f	C _{0.05} , E _{0.01} , K _{0.03} , N _{0.01} , Q _{0.04} , R _{0.23} , W _{0.66} C _{0.18} , R _{0.35} , W _{0.47}	W	.210	1.00
66	Chronic t/f	A _{0.01} , H _{0.03} , I _{0.11} , K _{0.01} , L _{0.01} , M _{0.01} , N _{0.02} , Q _{0.04} , T _{0.75} , V _{0.04} H _{0.06} , I _{0.35} , M _{0.06} , N _{0.06} , Q _{0.06} , T _{0.41}	T	.011	0.73
84	Chronic t/f	A _{0.03} , D _{0.02} , H _{0.05} , I _{0.07} , N _{0.81} , T _{0.01} , V _{0.01} , Y _{0.02} M _{0.08} , N _{0.92}	N	.055	1.00

TABLE 8. HIV-1C ACCESSORY PROTEIN MUTATIONS DURING PRIMARY INFECTION IN RELATION TO KNOWN HLA EPITOPES

Patient	Accessory protein	Days postseroconversion	Sequence	Epitope frequency	Patient HLA	Known HLA restriction																					
B_2865	Vif	Published CTL epitope t/f	⁸⁵ VSIEWRLRRYSTQVDPGL ¹⁰² 85...Q...G... ¹⁰² 85...Q...G... ¹⁰² 85...Q...G... ¹⁰² 85...Q...G... ¹⁰² 85...Q...G... ¹⁰² 85...Q...G...I... ¹⁰² 85...Q...G... ¹⁰²	1.00 1.00 1.00 0.88 0.12 1.00	A*23:01 A*80:01 B*18:01 B*53:01 Cw*02:02 Cw*06:02	B*18 Kiepiela <i>et al.</i> , ⁴¹																					
							E_3430	Vif	Published CTL epitope t/f	³² SKRASGWF [Y/F] ⁴⁰ 32..K.K...Y ⁴⁰ 32..K.K...Y ⁴⁰ 32..KTK...Y ⁴⁰ 32..EKTk...S Y ⁴⁰	1.00 1.00 1.00 1.00	A*29:02 A*74:01 B*13:02 B*15:03 Cw*02:10 Cw*06:02	B*15:03 Rousseau <i>et al.</i> ¹⁶														
														G_3603	Vif	Published CTL epitope t/f	⁵⁴ EVHIPLGEAR ⁶³ 54.....K ⁶³ 54.....K ⁶³ 54.....K ⁶³ 54.....K ⁶³ 54.....K ⁶³ 54.....K ⁶³ 54.....I ⁶³	1.00 1.00 1.00 1.00 0.71 0.29	A*30:01 A*68:01 B*14:02 B*15:03 Cw*02:10 Cw*08:02	A*68:01 Kiepiela <i>et al.</i> ⁴¹							
																					QU_6029	Vif	Published CTL epitope t/f	⁴¹ RHHYESRHPKVSSEVHI ⁵⁷ 41.....Y. ⁵⁷ 41.....Y. ⁵⁷ 41.....Y. ⁵⁷ 41.....I...Y. ⁵⁷	1.00 0.80 0.20	A*02:01 A*43:01 B*15:10 B*45:01 Cw*04:01 Cw*16:0	B*15:10 Kiepiela <i>et al.</i> ⁴¹

CTL, cytotoxic T lymphocyte; HLA, human leukocyte antigen.
The *bolded* patient HLA represents the one associated with the CTL epitope.

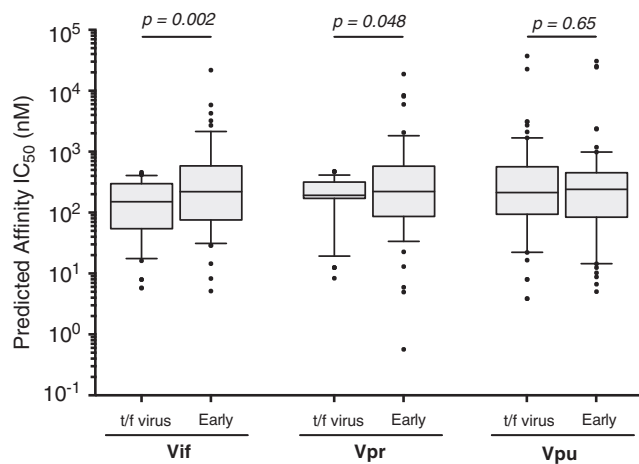


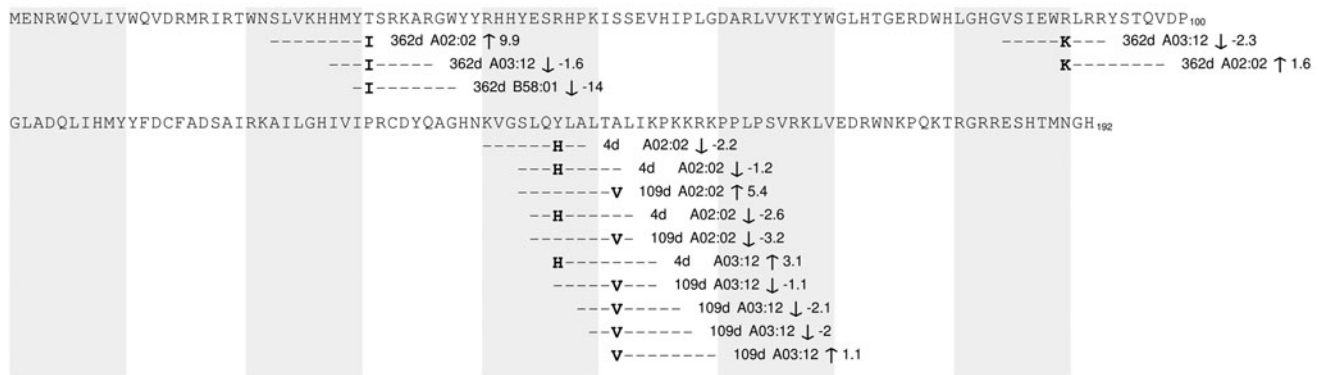
FIG. 4. Predicted epitope binding affinity in putative HLA class I epitopes in t/f viruses and variants arising in early infection. NetMHCpan 2.8 was used to predict the binding affinities of all possible Vif, Vpr, and Vpu 9-mer epitopes for each of the 17 patients against their corresponding MHC class I HLA alleles. Affinities are given as the IC₅₀ (nM), which is the peptide concentration at which half of the HLA molecules are occupied. As such, a lower IC₅₀ for a peptide is indicative of a stronger binding affinity for the HLA in question. We used an IC₅₀ cutoff of <500 nM to classify a putative peptide as being presented by HLA (Vif *n* = 58; Vpr *n* = 48; Vpu *n* = 69). Differences in binding affinity between putative epitopes in the t/f viruses were compared with the equivalent peptide found in early infection using the Wilcoxon matched-pairs signed rank test. HLA, human leukocyte antigen; MHC, major histocompatibility complex.

2.8 to predict autologous epitopes in Vif, Vpr, and Vpu and their binding affinities for each of the 17 patients. NetMHCpan, based on an artificial neural network, is reported to be the most accurate prediction currently available for MHC class I epitope binding. The output provides the predicted binding affinity of each putative epitope sequence to the restricting HLA allele. An IC₅₀ value threshold of <500 nM was used to identify putative autologous epitopes.

We used the t/f virus from each patient to represent Vif, Vpr, and Vpu peptides in their putative immunologically susceptible (i.e., non-HLA adapted) form and each unique variant observed during the follow-up period as a putative escape (i.e., HLA-adapted) form. Considering each patient’s HLA-A and HLA-B alleles and focusing on putative 9-mer epitopes, we evaluated any change in epitope binding affinity arising from a given amino acid substitution at each position within the epitope. We employed a sliding window approach whereby the query epitope was shifted by a single residue to avoid introducing bias by selecting only the epitopes with the strongest impact on HLA binding.

On average, amino acid substitutions seen in Vif during early infection increased the predicted median IC₅₀ of putative epitopes from 151 [IQR 54–299] nM in the t/f viruses to 221 (IQR: 76–585) nM in the escaped form (*p* < .001, Wilcoxon matched-pairs signed rank test) (Fig. 4), suggesting that the escape form would exhibit lower affinity for the restricting HLA allele. All Vif substitutions that were predicted to alter MHC class I epitope binding affinity were plotted onto patient-specific maps; examples for two patients are illustrated in Figure 5, while maps for all other patients are included as Supplementary Figure S2. Overall, similar

Patient C



Patient F

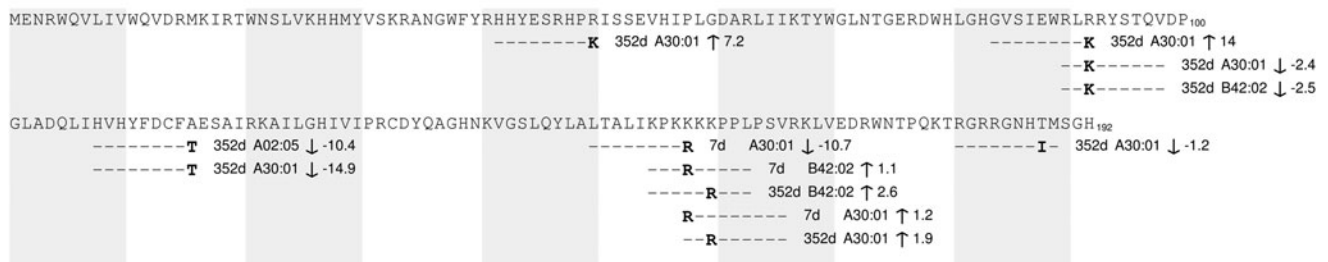


FIG. 5. Predicted binding affinities of putative Vif epitopes during early infection. Two patient transmitted/founder (MRCA) and subsequent putative predicted epitope binding fold change illustrated for patients C and F. MRCA, most recent common ancestor.

changes in predicted median IC_{50} of putative epitopes were observed for Vpr variants that arose during early infection (IC_{50} 243 [IQR: 99–601 nM]) compared with those encoded in *t/f* viruses (192 [IQR: 171–318] nM; $p = .048$) (Fig. 4). However, there was no significant difference in the binding affinity of putative Vpu epitopes found in the *t/f* viruses (214 [IQR: 94–564] nM) and those observed in early infection (240 [IQR: 84–452] nM; $p = .65$) (Fig. 4).

We hypothesized that mutations were more likely to occur in anchor residues of 9-mer epitopes, defined by known HLA subtype-specific anchor motifs (www.hiv.lanl.gov/content/immunology/motif_scan/motif.html). We presumed that mutations at anchor residues would reduce binding affinity to a greater degree than mutations in nonanchor residues. Indeed, amino acid substitutions at anchor residues resulted in a significant decrease in predicted epitope binding affinity for Vif (69 vs. 331 nM; $p < .001$, Wilcoxon matched-pairs signed rank test) and Vpr (192 vs. 283 nM; $p = .023$), although not Vpu (318 vs. 278 nM; $p = .413$) (Fig. 6). Amino acid substitutions at nonanchor residues did not significantly alter predicted epitope binding for any of the three accessory proteins (Fig. 6).

Discussion

We assessed amino acid signature patterns in HIV-1C accessory genes over the first 500 days *p/s* and addressed the potential role of MHC class I HLA restriction in virus evolution during early HIV-1C infection. Unsurprisingly, most of the critical functional domains in Vif, Vpr, and Vpu were completely conserved in early HIV infection. Vif tryptophan residues 5, 11, 21, and 38, involved in APOBEC3G antagonism, were completely conserved among transmitted viruses, nor was there any variation in the 144-SLQYLA-149 motif, supporting Vif anti-APOBEC3G function as critical during early infection.^{42–49} Despite previous reports of low sequence variability,³⁶ Vif amino acid residues 63–70, required for the formation of β -strand structures and critical in maintaining normal *vif* expression levels,⁵⁰ showed slight variations in the *t/f* viruses to a similar degree as those published in the extended subtype C consensus⁵¹ and the chronic infection subtype C analysis in this study. Residues Glu⁸⁸ and Trp⁸⁹ of the central hydrophilic region, which enhance steady-state expression of Vif, were conserved in the transmitted virus, although Gln⁸⁸, previously uncharacterized in subtype C, was present in one patient instead of Glu⁸⁸ (Val⁸⁸ was the only alternative amino acid during chronic infection at 0.5%). A previous study, however, indicates that the hydrophilicity and charge of the central region are dispensable for the function of Vif⁵⁰; therefore, the presence of the weak basic amino acid Gln in place of acidic Glu may not affect steady-state expression. The 90-RLRR-93 motif was not conserved in the transmitted viruses similar to previously reported subtype C analyses.^{36,52} The residues, Ser⁹⁵ and Thr⁹⁶, which are CKII and p44/42 mitogen-activated protein kinase phosphorylation sites, respectively, were relatively conserved in the transmitted viruses with variation present in one patient. As expected, the important His¹⁰⁸, Cys¹¹⁴, Cys¹³³, and His¹³⁹ (HCCH) zinc-binding motif was completely conserved in the transmitted viruses from all 17 patients, as identified in previous subtype B and C analyses,^{36,53,54} suggesting an important role for the HCCH-stabilized Vif–Cullin 5 interaction early in infection.

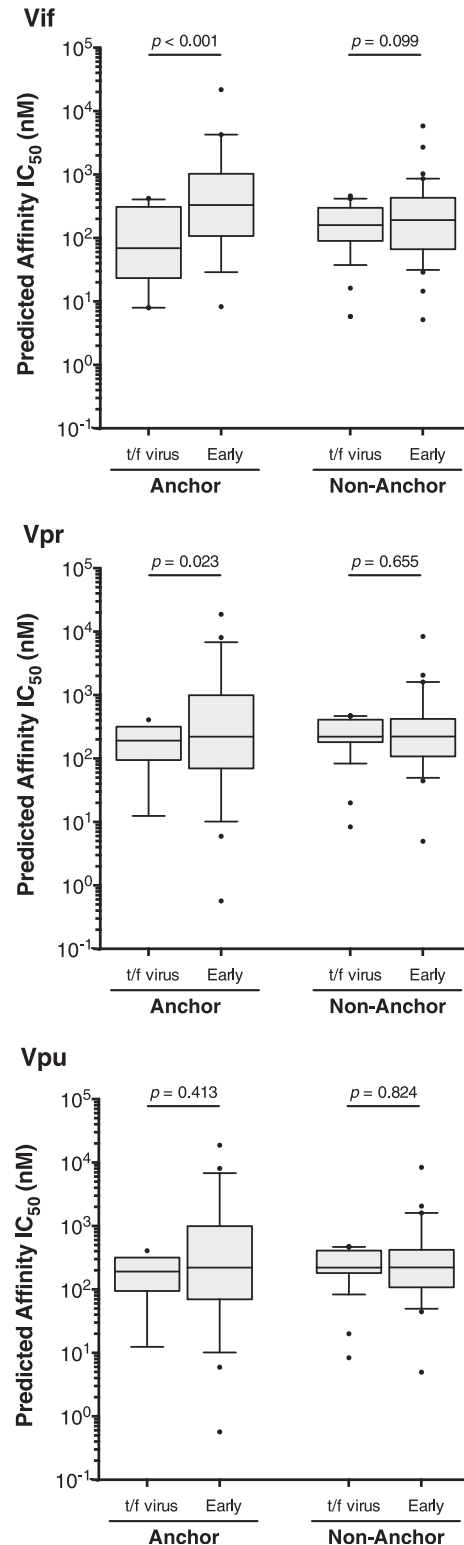


FIG. 6. Difference in binding affinity for HLA class I epitope substitutions that occur in anchored compared with nonanchored position. Amino acid substitutions were characterized as occurring in anchor or nonanchor residues using HLA subtype-specific anchor motifs: Vif anchor $n = 19$, nonanchor $n = 39$; Vpr anchor $n = 19$, nonanchor $n = 29$; Vpu anchor $n = 27$, nonanchor $n = 42$. Predicted peptide binding affinities were compared using the Wilcoxon matched-pairs signed rank test.

The HIV-1 subtype B Vif phosphorylation site Thr¹⁵⁵ is a highly conserved Lys¹⁵⁵ in subtype C, which is 100% conserved in the transmitted viruses in this study, with variation observed in 2.7% of chronic infection sequences. Substitutions at this site could have an effect on phosphorylation and activity of Vif during HIV infection. The Vif oligomerization site 161-PPLP-164 was conserved throughout, confirming its central role in Vif function and viral infectivity during all stages of infection. At the phosphorylation site position 170, our findings are similar to previously published HIV-1 subtype C studies^{36,52} where Val is the major amino acid in the transmitted viruses, although Ile is present in one patient.

Vpr t/f virus sequences were more conserved than those observed during chronic infection. The tertiary structure of Vpr comprises three α -helices folded around a hydrophobic core made up of leucine, isoleucine, valine, and aromatic residues.⁵⁵ The first α -helix (aa17-33) plays an essential role in nuclear localization and virion packaging, while the second (aa 38–50) is associated with oligomerization of Vpr and translocation of the preintegration complex to the perinuclear region. The third α -helix (aa 56–77) is associated with dimerization. Unlike previous findings in subtype C,^{36,56} we found reduced conservation of Pro¹⁰ and Pro³⁷ in the transmitted subtype C viruses. Try⁵⁴ and His⁷¹, previously described as aromatic amino acids involved in stabilizing the Vpr dimer and UNG2 binding,^{36,56} were completely conserved in the t/f viruses. Mutations such as R77A/G, observed in the early HIV-1C infection cohort in this study, are known to reduce the proapoptotic activity of Vpr.^{57,58} This variation is in line with the suggestion that proapoptotic activity of HIV-1 subtype C Vpr is different from HIV-1 subtype B.³⁶ R80A was not observed in any of 17 patients in this study and all transmitted viruses had arginine (R) at that position in Vpr. Therefore, it is plausible that in HIV-1 subtype C infection, two conserved arginine residues at positions 73 and 80 are sufficient for adenine nucleoside translocator interaction.³⁶ Other well-known substitutions, such as E21P, E24P, and A59P, which prevent the incorporation of Vpr into virus-like particles, as well as Q65R, which is correlated with impairment of Vpr docking at the nuclear envelope,⁵⁹ were not observed in this study. Neither were other known substitutions, underlying the integrity of Vpr during early HIV-1C infection.^{60–63}

All 17 t/f viruses had a five-residue N-terminal insertion in Vpu compared with HIV-1 subtype B sequences. The transmembrane (TM) domain aa 12–29 and the hinge region residues 34–42 in the transmitted viruses were conserved, as previously described.³⁶ Differences in Gly38 and Ile41 positions were observed in the transmitted viruses in this study. Thr29 is predicted to act as a channel anchor and the presence of isoleucine in the transmitted viruses from five patients, a similar frequency as seen in chronic infection, could support the suggestion that the 100% conserved Tyr32 is an alternative channel anchor in subtype C.³⁶ The predicted regions thought to be associated with subtype C channel linking showed slight variations in Ile22 (one patient) and Ile23 (two patients), whereas Ile25 was 100% conserved as described previously.³⁶ Gly13 is thought to be involved in interacting with the membrane carbonyl oxygen groups to stabilize closed states and was present in the majority of transmitted viruses in our study, with alternative residue alanine present in three patients. Similar to findings by Bell *et al.*,³⁶ no serine

residues were observed in the TM domain, and the 7-RVDYR-11 was highly conserved, as previously described, excluding variation observed in Arg7, although the second-most prevalent amino acid at that position, Lys7, is also a highly charged positive amino acid, supporting the prediction that in HIV-1 subtype C, this is the region that maintains Vpu channel selectivity.³⁶ Vpu protein acts as an adaptor for CD4 proteasomal degradation by recruiting CD4 and β -transducin repeat-containing protein (β TrCP), the receptor component of the multisubunit SCF- β TrCP E3 ubiquitin ligase complex. The recognition signal for these cellular ligands by β TrCP is the phosphorylation of one or two serine residues present in a conserved Vpu motif, 57-DSGNES-62.⁶⁴ Vpu proteins in all transmitted viruses are predicted to bind β TrCP through Ser58 and Ser62, which were highly conserved in the transmitted Vpu. However, it is worth noting that overall, Vpu function is highly preserved despite sequence variation, which argues against the idea of a genetic signature for Vpu, which is suggested by our study. Additionally relevant, a recent study testing alleles from transmitted and chronic viruses by Mlcochova *et al.*⁶⁵ suggests that immune evasion activities of accessory proteins, Vpu and Vif, are conserved in acute and chronic HIV-1 infection. Note that at the q-value threshold of 0.3 used in the present study, it is expected that no >30% of identified associations would represent false positives; as such, further work is needed to confirm these putative transmitted versus chronic subtype C infection signatures.

Immune escape pathways are broadly predictable based on host HLA class I profile.¹⁴ Nonconsensus variants (amino acid residues that differ from the HIV-1C consensus sequence) within known CTL-restricted epitopes in t/f virus sequences that were maintained over the first 500 days p/s cannot be conclusively associated with immune escape as these were presumably acquired at transmission. In contrast, variants that emerged within known or predicted CTL epitopes during follow-up are suggestive of immune escape⁶⁶ and are consistent with rapid initial accumulation of CTL escape mutations during the first weeks of infection.¹⁸ The observation that many of these emerging variants are predicted to reduce peptide-HLA binding affinity further supports these as CTL escape mutations;^{16,41,67} experimental validation of predicted epitopes and their selected variants will be addressed in future studies. It is also important to note that while we have used the terminology transmitted/founder viruses to describe the reconstructed MRCA HIV-1 sequence for each patient, we have not experimentally verified that these sequences are capable of being transmitted to multiple recipients, nor that the immunological pressures leading to changes in Vif, Vpr, and Vpu may have reduced the transmissibility of these viruses. Furthermore, to our knowledge, ours is the first study to investigate the impact of within-host evolution on the predicted binding affinities of putative novel epitopes in Vif, Vpr, and Vpu; as such, the generalizability of results to subtypes other than HIV-1C remains to be determined.

Our study was further constrained by the limited characterization of CTL epitopes in HIV-1 subtype C, particularly in accessory proteins. To partially address this limitation, we employed an artificial neural network-based prediction of epitope:HLA binding to identify putative epitopes. By comparing the affinity of predicted epitopes within the transmitted Vif, Vpr, and Vpu sequences to variants that arose within

the patient quasispecies over 500 days p/s, we determined that polymorphisms in Vif and Vpr were more likely to significantly decrease epitope binding when the mutation occurred in an anchor residue. These results highlight the need for further quantitative work on the impact of CTL escape mutations in HIV-1C accessory proteins.

In summary, our results highlight (1) highly conserved regions in accessory genes during both primary and chronic HIV-1 subtype C infection, (2) putative transmission signature residues in Vif and Vpr, but not Vpu, (3) the suggestion of possible escape mutations that are HLA-associated early p/s, and (4) that primary infection polymorphisms in Vif and Vpr accessory genes were more likely to decrease epitope binding when the mutation occurred at an anchor residue. These data contribute to the knowledge about primary HIV-1 subtype C infection and could be helpful toward HIV vaccine strategies that incorporate accessory genes as components.

Acknowledgments

The authors are grateful to the participants of the *Tshedimoso* study in Botswana—without their participation, this research would not have been possible (*Re a leboga le kamoso*). The authors thank the Botswana Ministry of Health, Gaborone City Council clinics, and the Gaborone VCT *Tebelopele* for collaboration and are grateful to Gaseboloke Mthowang, Florence Modise, S'khatele Molefhabangwe, Sarah Masole, and the late Melissa Ketunuti for their dedication and outstanding work in the clinic and outreach. The authors are grateful to Jan Walter De Neve, Lauren Margolin, Lauren Buck, Andrew J. Fiore-Gartland, Jeannie Baca, and Brian Foley for excellent technical assistance and thank Lendsey Melton for editorial assistance. They thank the Botswana Harvard AIDS Institute (Gaborone, Botswana) and the Essex Lab (Harvard T.H. Chan School of Public Health, Boston, MA) where the laboratory and data analyses were performed. The primary HIV-1 subtype C infection study in Botswana, the *Tshedimoso* study, was supported by the National Institute of Allergy and Infectious Diseases (R01 AI057027). R.R. was supported, in part, by the NIH Fogarty International Clinical Research Scholars and Fellows Program, the NIH Fogarty International Center AIDS International Training and Research Program (D43 TW000004), and the NIH Fogarty HIV Research Training Program for Low- and Middle-Income Countries (D43 TW009610). I.J.M. was supported by a Feasibility Award from the Harvard University Center for AIDS Research (CFAR), an NIH-funded program (P30 AI060354). Z.L.B. was supported by a New Investigator Award from the Canadian Institutes of Health Research (CIHR) and a Scholar Award from the Michael Smith Foundation for Health Research (MSFHR). This work was also supported, in part, by University of Botswana ORD, R812 (TKS).

Author Disclosure Statement

No competing financial interests exist.

References

1. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, Athreya GS, Treurnicht FK, Keele BF, Wood N, Salazar-Gonzalez JF, Bhattacharya T, Chu H, Hoffman I, Galvin S, Mpanje C, Kazembe P, Thebus R, Fiscus S, Hide W, Cohen MS, Karim SA, Haynes BF, Shaw GM, Hahn BH, Korber BT, Swanstrom R, Williamson C: Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-poisson distribution of transmitted variants. *J Virol* 2009; 83:3556–3567.
2. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, Heil ML, Kasolo F, Musonda R, Hahn BH, Shaw GM, Korber BT, Allen S, Hunter E: Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 2004;303:2019–2022.
3. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT: Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One* 2010;5:e12303.
4. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM: Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* 2008;105:7552–7557.
5. Novitsky V, Wang R, Margolin L, Baca J, Rossenkhan R, Moyo S, van Widenfelt E, Essex M: Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. *PLoS One* 2011;6:e16714.
6. Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, Gray RH, Serwadda D, Sewankambo NK, Shepherd JC, Toma J, Huang W, Quinn TC: Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *J Infect Dis* 2009;199:580–589.
7. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, Deymier MJ, Ende ZS, Klatt NR, DeZiel CE, Lin TH, Peng J, Seese AM, Shapiro R, Frater J, Ndung'u T, Tang J, Goepfert P, Gilmour J, Price MA, Kilembe W, Heckerman D, Goulder PJ, Allen TM, Allen S, Hunter E: HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. *Science* 2014;345:1254031.
8. Mota TM, Murray JM, Center RJ, Purcell DF, McCaw JM: Application of a case-control study design to investigate genotypic signatures of HIV-1 transmission. *Retrovirology* 2012;9:54.
9. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, Overbaugh J: Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J Virol* 2005;79:6528–6531.
10. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferott K, Aasa-Chapman MM, Pellegrino P, Williams I, Cohen MS, Gao F, Shaw GM, Hahn BH, Ochsenbauer C, Kappes JC, Borrow P: Relative resistance of HIV-1 founder viruses to control by interferon-alpha. *Retrovirology* 2013;10:146.
11. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNeven J, Zhao H, Wong K, Stoddard JN, Raugi D, Sorensen S, Genowati I, Birditt B, McKay A, Diem K, Maust BS, Deng W, Collier AC, Stekler JD, McElrath MJ, Mullins JI: De-

- mographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. *J Virol* 2011;85:7523–7534.
12. Walker BD, Chakrabarti S, Moss B, Paradis TJ, Flynn T, Durmo AG, Blumberg RS, Kaplan JC, Hirsch MS, Schooley RT: HIV-specific cytotoxic T lymphocytes in seropositive individuals. *Nature* 1987;328:345–348.
 13. Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, Block BL, Baker B, Kadie C, Markowitz M, Jessen H, Kelleher AD, Rosenberg E, Kaldor J, Yuki Y, Carrington M, Allen TM, Mallal S, Altfeld M, Heckerman D, Walker BD: Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J Virol* 2008;82:9216–9227.
 14. Brumme ZL, Brumme CJ, Heckerman D, Korber BT, Daniels M, Carlson J, Kadie C, Bhattacharya T, Chui C, Szinger J, Mo T, Hogg RS, Montaner JS, Frahm N, Brander C, Walker BD, Harrigan PR: Evidence of differential HLA class I-mediated viral evolution in functional and accessory/regulatory genes of HIV-1. *PLoS Pathog* 2007;3:e94.
 15. Kelleher AD, Booth BL, Jr., Sewell AK, Oxenius A, Cerundolo V, McMichael AJ, Phillips RE, Price DA: Effects of retroviral protease inhibitors on proteasome function and processing of HIV-derived MHC class I-restricted cytotoxic T lymphocyte epitopes. *AIDS Res Hum Retroviruses* 2001;17:1063–1066.
 16. 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: HLA class I-driven evolution of human immunodeficiency virus type 1 subtype C proteome: Immune escape and viral load. *J Virol* 2008;82:6434–6446.
 17. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, Decker JM, Wang S, Baalwa J, Kraus MH, Parrish NF, Shaw KS, Guffey MB, Bar KJ, Davis KL, Ochsenbauer-Jambor C, Kappes JC, Saag MS, Cohen MS, Mulenga J, Derdeyn CA, Allen S, Hunter E, Markowitz M, Hraber P, Perelson AS, Bhattacharya T, Haynes BF, Korber BT, Hahn BH, Shaw GM: Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. *J Exp Med* 2009;206:1273–1289.
 18. Treurnicht FK, Seoighe C, Martin DP, Wood N, Abrahams MR, Rosa Dde A, Bredell H, Woodman Z, Hide W, Mlisana K, Karim SA, Gray CM, Williamson C: Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure. *Virology* 2009;396:213–225.
 19. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Berlin AM, Malboeuf CM, Ryan EM, Gnerre S, Zody MC, Erlich RL, Green LM, Berical A, Wang Y, Casali M, Streeck H, Bloom AK, Dudek T, Tully D, Newman R, Axten KL, Gladden AD, Battis L, Kemper M, Zeng Q, Shea TP, Gujja S, Zedlack C, Gasser O, Brander C, Hess C, Gunthard HF, Brumme ZL, Brumme CJ, Bazner S, Rychert J, Tinsley JP, Mayer KH, Rosenberg E, Pereyra F, Levin JZ, Young SK, Jessen H, Altfeld M, Birren BW, Walker BD, Allen TM: Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 2012;8:e1002529.
 20. Joseph SB, Swanstrom R, Kashuba AD, Cohen MS: Bottlenecks in HIV-1 transmission: Insights from the study of founder viruses. *Nat Rev Microbiol* 2015;13:414–425.
 21. Novitsky V, Lagakos S, Herzig M, Bonney C, Kebaabetswe L, Rossenkhon R, Nkwe D, Margolin L, Musonda R, Moyo S, Woldegabriel E, van Widenfelt E, Makhema J, Essex M: Evolution of proviral gp120 over the first year of HIV-1 subtype C infection. *Virology* 2009;383:47–59.
 22. Novitsky V, Woldegabriel E, Wester C, McDonald E, Rossenkhon R, Ketunuti M, Makhema J, Seage GR, 3rd, Essex M: Identification of primary HIV-1C infection in Botswana. *AIDS care* 2008;20:806–811.
 23. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrandt C, Smith R, Conrad A, Kleinman SH, Busch MP: Dynamics of HIV viremia and antibody seroconversion in plasma donors: Implications for diagnosis and staging of primary HIV infection. *AIDS* 2003;17:1871–1879.
 24. Novitsky V, Wang R, Kebaabetswe L, Greenwald J, Rossenkhon R, Moyo S, Musonda R, Woldegabriel E, Lagakos S, Essex M: Better control of early viral replication is associated with slower rate of elicited antiviral antibodies in the detuned enzyme immunoassay during primary HIV-1C infection. *J Acquir Immune Defic Syndr* 2009;52:265–272.
 25. Rossenkhon R, Novitsky V, Sebunya TK, Musonda R, Gashe BA, Essex M: 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* 2011;7:e35491.
 26. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, Davidov O, Zhao LP, Mullins JI: HIV quasispecies and resampling. *Science* 1996;273:415–416.
 27. Deng W, Maust BS, Nickle DC, Learn GH, Liu Y, Heath L, Kosakovsky Pond SL, Mullins JI: DIVEIN: A web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques* 2010;48:405–408.
 28. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;52:696–704.
 29. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;28:2731–2739.
 30. Korber BT, Kunstman KJ, Patterson BK, Furtado M, McEvilly MM, Levy R, Wolinsky SM: Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: Evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *J Virol* 1994;68:7467–7481.
 31. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Statist Soc B* 1995;57:289–300.
 32. R Core Team. R: A language and environment for statistical computing [computer program]. Vienna: R Foundation for Statistical Computing; 2015. Available at: www.R-project.org, accessed June 6, 2016.
 33. Novitsky V, Wang R, Margolin L, Baca J, Kebaabetswe L, Rossenkhon R, Bonney C, Herzig M, Nkwe D, Moyo S, Musonda R, Woldegabriel E, van Widenfelt E, Makhema J, Lagakos S, Essex M: Timing constraints of in vivo gag mutations during primary HIV-1 subtype C infection. *PLoS One* 2009;4:e7727.
 34. Listgarten J, Brumme Z, Kadie C, Xiaojiang G, Walker B, Carrington M, Goulder P, Heckerman D: Statistical reso-

- lution of ambiguous HLA typing data. *PLoS Comput Biol* 2008;4:e1000016.
35. Rasmussen M, Harndahl M, Stryhn A, Boucherma R, Nielsen LL, Lemonnier FA, Nielsen M, Buus S: Uncovering the peptide-binding specificities of HLA-C: A general strategy to determine the specificity of any MHC class I molecule. *J Immunol* 2014;193:4790–4802.
 36. Bell CM, Connell BJ, Capovilla A, Venter WD, Stevens WS, Papathanasopoulos MA: Molecular characterization of the HIV type 1 subtype C accessory genes *vif*, *vpr*, and *vpu*. *AIDS Res Hum Retroviruses* 2007;23:322–330.
 37. Hasan Z, Carlson JM, Gatanaga H, Le AQ, Brumme CJ, Oka S, Brumme ZL, Ueno T: Minor contribution of HLA class I-associated selective pressure to the variability of HIV-1 accessory protein Vpu. *Biochem Biophys Res Commun* 2012;421:291–295.
 38. Pickering S, Hue S, Kim EY, Reddy S, Wolinsky SM, Neil SJ: Preservation of tetherin and CD4 counter-activities in circulating Vpu alleles despite extensive sequence variation within HIV-1 infected individuals. *PLoS Pathog* 2014;10:e1003895.
 39. Siwka W, Schwinn A, Baczko K, Pardowitz I, Mhalu F, Shao J, Rethwilm A, ter Meulen V: *vpu* and *env* sequence variability of HIV-1 isolates from Tanzania. *AIDS Res Hum Retroviruses* 1994;10:1753–1754.
 40. Korber B, Myers G: Signature pattern analysis: A method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses* 1992;8:1549–1560.
 41. 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: CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* 2007;13:46–53.
 42. Conticello SG, Harris RS, Neuberger MS: The Vif protein of HIV triggers degradation of the human antiretroviral DNA deaminase APOBEC3G. *Curr Biol* 2003;13:2009–2013.
 43. Goila-Gaur R, Strebel K: HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology* 2008;5:51.
 44. Malim MH: Natural resistance to HIV infection: The Vif-APOBEC interaction. *C R Biol* 2006;329:871–875.
 45. Mariani R, Chen D, Schrofelbauer B, Navarro F, Konig R, Bollman B, Munk C, Nymark-McMahon H, Landau NR: Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* 2003;114:21–31.
 46. Romani B, Engelbrecht S, Glashoff RH: Antiviral roles of APOBEC proteins against HIV-1 and suppression by Vif. *Arch Virol* 2009;154:1579–1588.
 47. Sheehy AM, Gaddis NC, Choi JD, Malim MH: Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 2002;418:646–650.
 48. Sheehy AM, Gaddis NC, Malim MH: The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* 2003;9:1404–1407.
 49. Yang B, Gao L, Li L, Lu Z, Fan X, Patel CA, Pomerantz RJ, DuBois GC, Zhang H: Potent suppression of viral infectivity by the peptides that inhibit multimerization of human immunodeficiency virus type 1 (HIV-1) Vif proteins. *J Biol Chem* 2003;278:6596–6602.
 50. Fujita M, Akari H, Sakurai A, Yoshida A, Chiba T, Tanaka K, Strebel K, Adachi A: Expression of HIV-1 accessory protein Vif is controlled uniquely to be low and optimal by proteasome degradation. *Microbes Infect* 2004;6:791–798.
 51. Novitsky V, Smith UR, Gilbert P, McLane MF, Chigwedere P, Williamson C, Ndung'u T, Klein I, Chang SY, Peter T, Thior I, Foley BT, Gaolekwe S, Rybak N, Gaseitsiwe S, Vannberg F, Marlink R, Lee TH, Essex M: Human immunodeficiency virus type 1 subtype C molecular phylogeny: Consensus sequence for an AIDS vaccine design? *J Virol* 2002;76:5435–5451.
 52. Jacobs GB, Nistal M, Laten A, van Rensburg EJ, Rethwilm A, Preiser W, Bodem J, Engelbrecht S: Molecular analysis of HIV type 1 *vif* sequences from Cape Town, South Africa. *AIDS Res Hum Retroviruses* 2008;24:991–994.
 53. Wieland U, Hartmann J, Suhr H, Salzberger B, Eggers HJ, Kuhn JE: In vivo genetic variability of the HIV-1 *vif* gene. *Virology* 1994;203:43–51.
 54. Scriba TJ, Treurnicht FK, Zeier M, Engelbrecht S, van Rensburg EJ: Characterization and phylogenetic analysis of South African HIV-1 subtype C accessory genes. *AIDS Res Hum Retroviruses* 2001;17:775–781.
 55. Le Rouzic E, Benichou S: The Vpr protein from HIV-1: Distinct roles along the viral life cycle. *Retrovirology* 2005;2:11.
 56. Romani B, Glashoff R, Engelbrecht S: Molecular and phylogenetic analysis of HIV type 1 *vpr* sequences of South African strains. *AIDS Res Hum Retroviruses* 2009;25:357–362.
 57. Jacotot E, Ravagnan L, Loeffler M, Ferri KF, Vieira HL, Zamzami N, Costantini P, Druillennec S, Hoebeke J, Briand JP, Irinopoulou T, Daugas E, Susin SA, Cointe D, Xie ZH, Reed JC, Roques BP, Kroemer G: The HIV-1 viral protein R induces apoptosis via a direct effect on the mitochondrial permeability transition pore. *J Exp Med* 2000;191:33–46.
 58. Lum JJ, Cohen OJ, Nie Z, Weaver JG, Gomez TS, Yao XJ, Lynch D, Pilon AA, Hawley N, Kim JE, Chen Z, Montpetit M, Sanchez-Dardon J, Cohen EA, Badley AD: Vpr R77Q is associated with long-term nonprogressive HIV infection and impaired induction of apoptosis. *J Clin Invest* 2003;111:1547–1554.
 59. Tungaturthi PK, Sawaya BE, Ayyavoo V, Murali R, Srinivasan A: HIV-1 Vpr: Genetic diversity and functional features from the perspective of structure. *DNA Cell Biol* 2004;23:207–222.
 60. Jacquot G, Le Rouzic E, David A, Mazzolini J, Bouchet J, Bouaziz S, Niedergang F, Pancino G, Benichou S: Localization of HIV-1 Vpr to the nuclear envelope: Impact on Vpr functions and virus replication in macrophages. *Retrovirology* 2007;4:84.
 61. Kamata M, Nitahara-Kasahara Y, Miyamoto Y, Yoneda Y, Aida Y: Importin-alpha promotes passage through the nuclear pore complex of human immunodeficiency virus type 1 Vpr. *J Virol* 2005;79:3557–3564.
 62. Kogan M, Rappaport J: HIV-1 accessory protein Vpr: Relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* 2011;8:25.
 63. Sawaya BE, Khalili K, Gordon J, Srinivasan A, Richardson M, Rappaport J, Amini S: Transdominant activity of human immunodeficiency virus type 1 Vpr with a mutation at residue R73. *J Virol* 2000;74:4877–4881.
 64. Besnard-Guerin C, Belaidouni N, Lassot I, Segéral E, Jobart A, Marchal C, Benarous R: HIV-1 Vpu sequesters beta-transducin repeat-containing protein (betaTrCP) in the

- cytoplasm and provokes the accumulation of beta-catenin and other SCFbetaTrCP substrates. *J Biol Chem* 2004; 279:788–795.
65. Mlcochova P, Apolonia L, Kluge SF, Sridharan A, Kirchhoff F, Malim MH, Sauter D, Gupta RK: Immune evasion activities of accessory proteins Vpu, Nef and Vif are conserved in acute and chronic HIV-1 infection. *Virology* 2015;482:72–78.
66. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Gansov VV, Keele BF, Learn GH, Turnbull EL, Salazar MG, Weinhold KJ, Moore S, Letvin N, Haynes BF, Cohen MS, Hraber P, Bhattacharya T, Borrow P, Perelson AS, Hahn BH, Shaw GM, Korber BT, McMichael AJ: The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J Exp Med* 2009;206:1253–1272.
67. Bazhan SI, Karpenko LI, Ilyicheva TN, Belavin PA, Seregin SV, Danilyuk NK, Antonets DV, Ilyichev AA: Rational design based synthetic polyepitope DNA vaccine for eliciting HIV-specific CD8+ T cell responses. *Mol Immunol* 2007;47:1507–1515.

Address correspondence to:

M. Essex, DVM, PhD

Lasker Professor of Health Sciences

Chair, Harvard T.H. Chan School

of Public Health AIDS Initiative

FXB 402

651 Huntington Avenue

Boston, MA 02115

E-mail: messex@hsph.harvard.edu