Supplementary Information Guide:

Defective proviruses rapidly accumulate during acute HIV-1 infection

Authors: Katherine M. Bruner, Alexandra J. Murray, Ross A. Pollack, Mary G. Soliman, Sarah B. Laskey, Adam A. Capoferri, Jun Lai, Matthew C. Strain, Steven M. Lada, Rebecca Hoh, Ya-Chi Ho, Douglas D. Richman, Steven G. Deeks, Janet D. Siliciano and Robert F. Siliciano*

Supplementary Figure 1. - Characterization of proviral sequences in subjects initiating ART during chronic infection.

Supplementary Figure 2. – Deletions are nonrandom and occur at hotspots in the HIV-1 genome.

Supplementary Figure 3. - High proportion of hypermutated sequences contain mutated start codons and premature stop codons.

Supplementary Figure 4. - 2LTR-circles are present at low levels in resting CD4+ T cells of subjects treated during acute or chronic infection.

Supplementary Figure 5. - Characterization of proviral sequences in subjects initiating ART during acute/early infection.

Supplementary Figure 6. - Acutely-treated subjects have a higher fraction of hypermutated proviruses and a lower fraction of deleted proviruses than chronically-treated subjects.

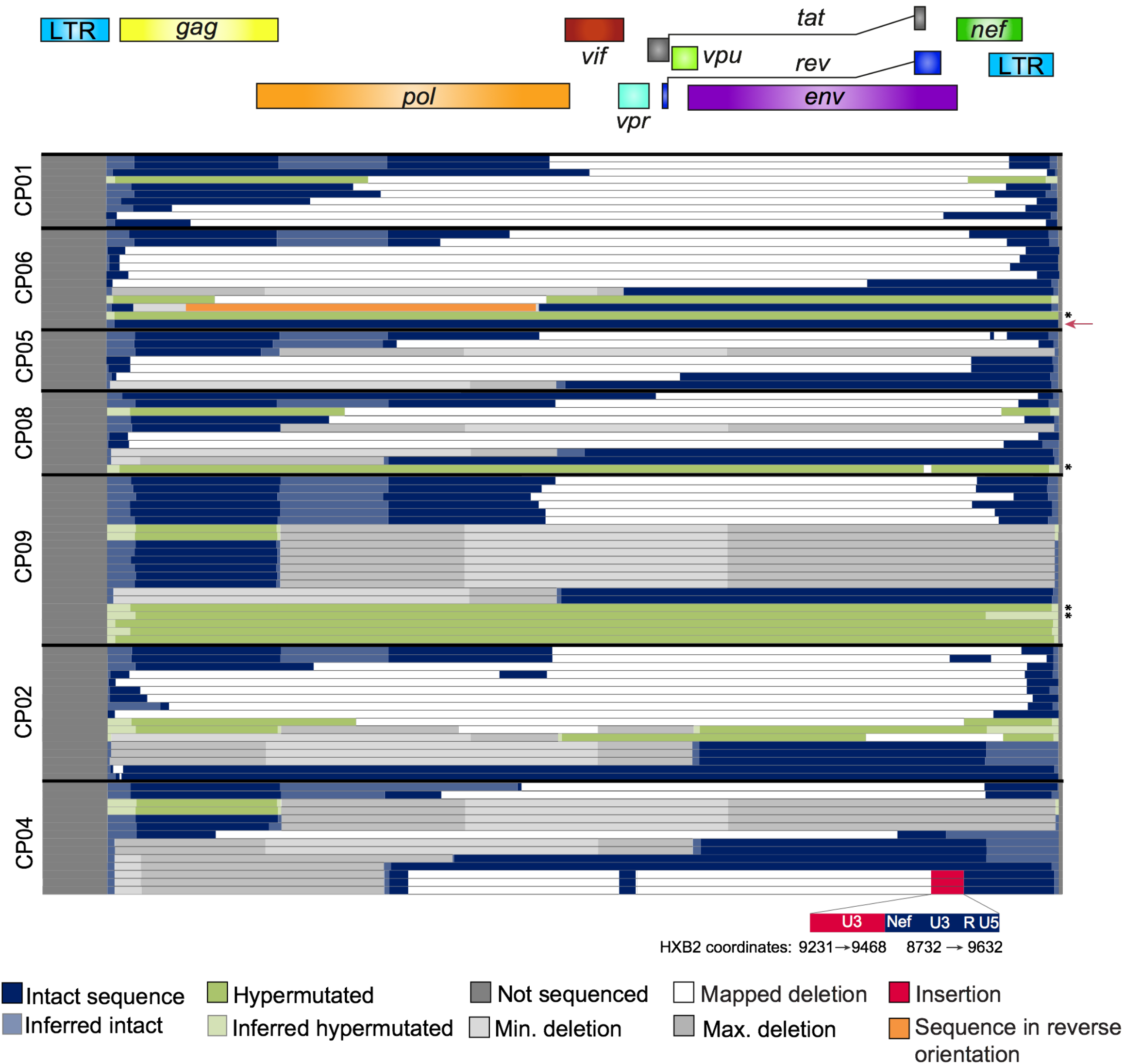
Supplementary Figure 7. - The number of intact proviruses does not correlate well with current assays.

Supplementary Table 1. - Study participant characteristics.

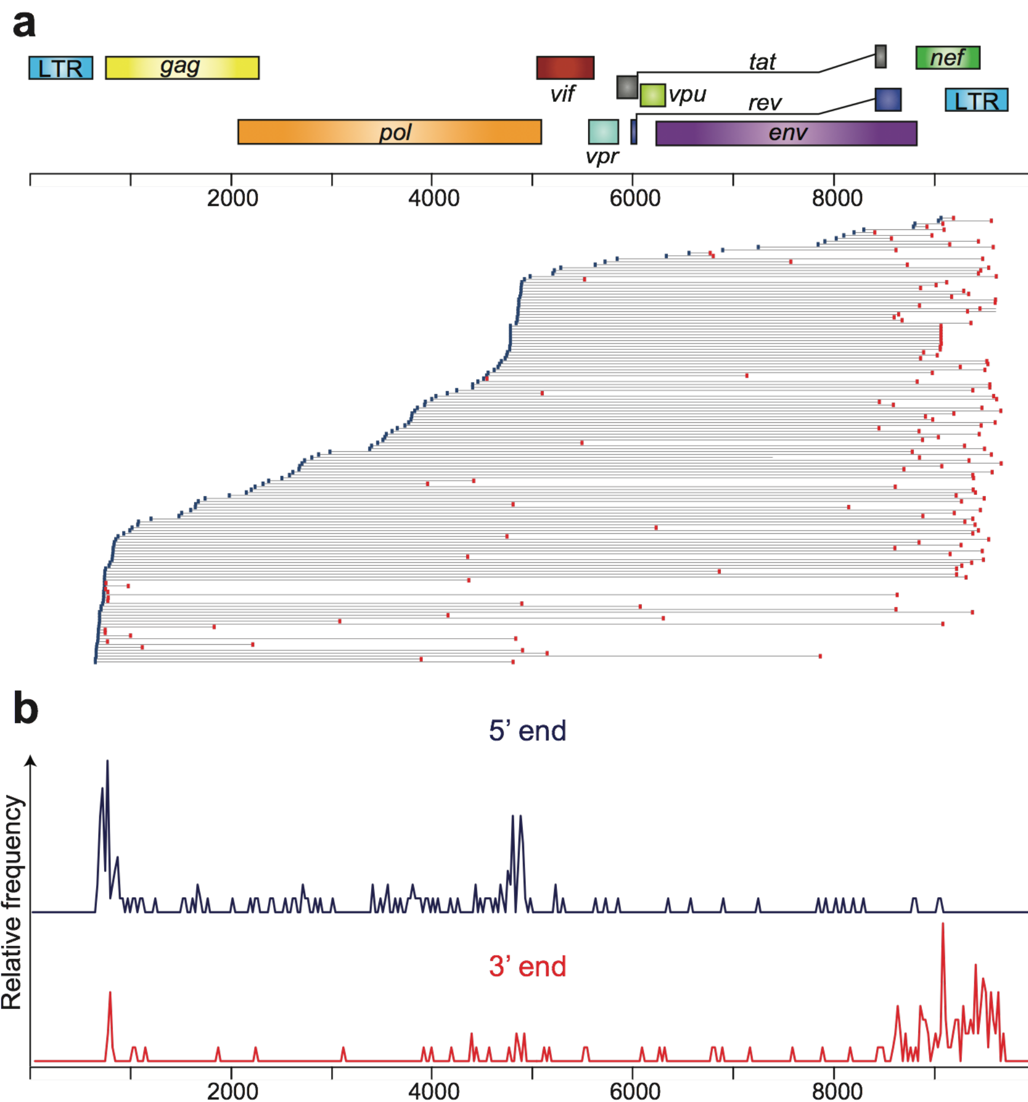
Supplementary Table 2. - PCR primers and conditions.

Supplementary Information:
Defective proviruses rapidly accumulate during acute HIV-1 infection

Authors: Katherine M. Bruner, Alexandra J. Murray, Ross A. Pollack, Mary G. Soliman, Sarah B. Laskey, Adam A. Capoferri, Jun Lai, Matthew C. Strain, Steven M. Lada, Rebecca Hoh, Ya-Chi Ho, Douglas D. Richman, Steven G. Deeks, Janet D. Siliciano and Robert F. Siliciano*

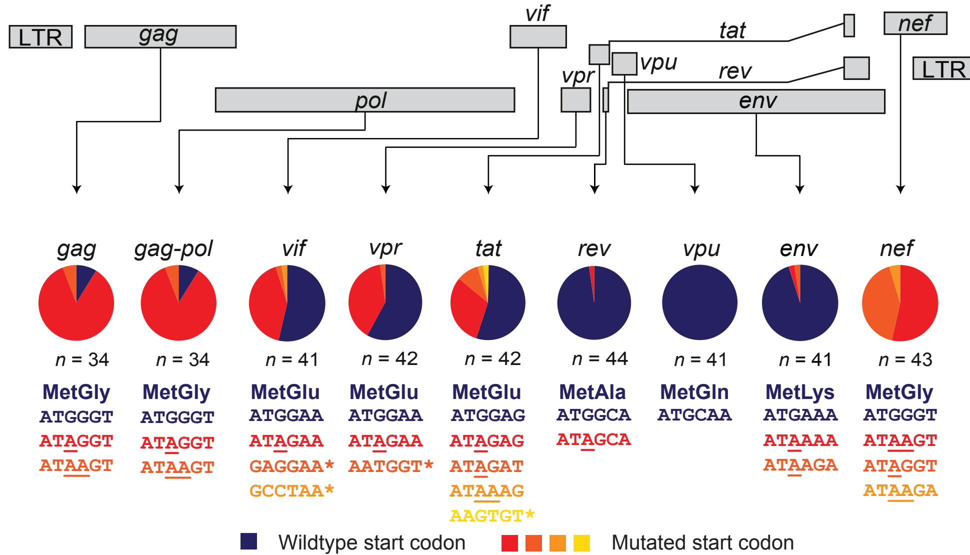


Supplemental Figure 1. Characterization of proviral sequences in subjects initiating ART during chronic infection. Sequences maps of proviral clones from seven additional subjects treated during the chronic phase of infection. Each horizontal bar represents an individual clone identified through full genome sequencing. In cases where a deletion could not be precisely mapped, likely due to a deletion encompassing multiple forward or reverse primer binding sites, the maximum and minimum deletions sizes are plotted (in grey). If sequencing data shows a mapped deletion that removes primer-binding sites for other amplicons, the resulting missing sequence was inferred to be present (light blue or green). A pink arrow denotes full-length, genetically intact sequences. Black asterisks indicate likely full-length hypermutated proviruses that were not fully sequenced due to extreme hypermutation.

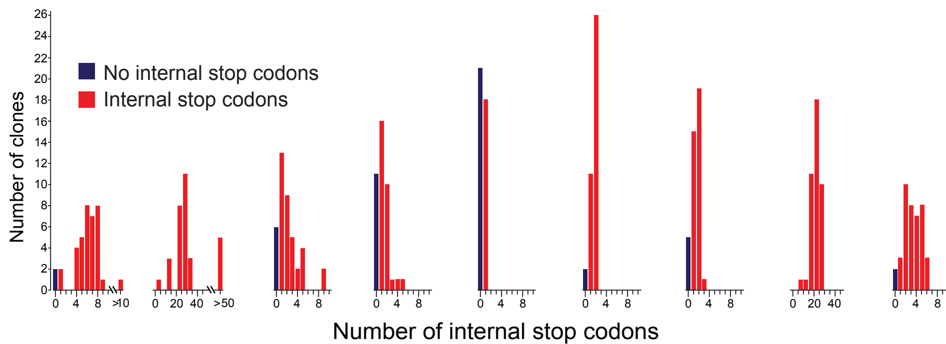


Supplemental Figure 2. Deletions are nonrandom and occur at hotspots in the HIV-1 genome. (**a**) Distribution of deletions in subject sequences. Each horizontal line indicates a unique proviral deletion from a clone identified by near full-length genome sequencing. Sequences from all treated subjects (acute and chronic phase) are included. In cases of identical (expanded) clones, only one clone was included in this analysis. For proviral clones with multiple deletions, one horizontal bar is shown for each deletion. Blue and red circles indicate the 5' and 3' ends of the deletion, respectively. (**b**) Distribution of the 5' (blue) and 3' (red) ends of the deletions. Hotspots for deletions, indicated by spikes in the frequency distributions, may be related to structural features in the RNA genome.

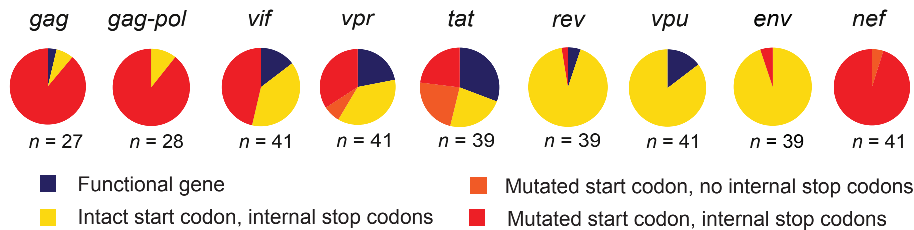
a



b

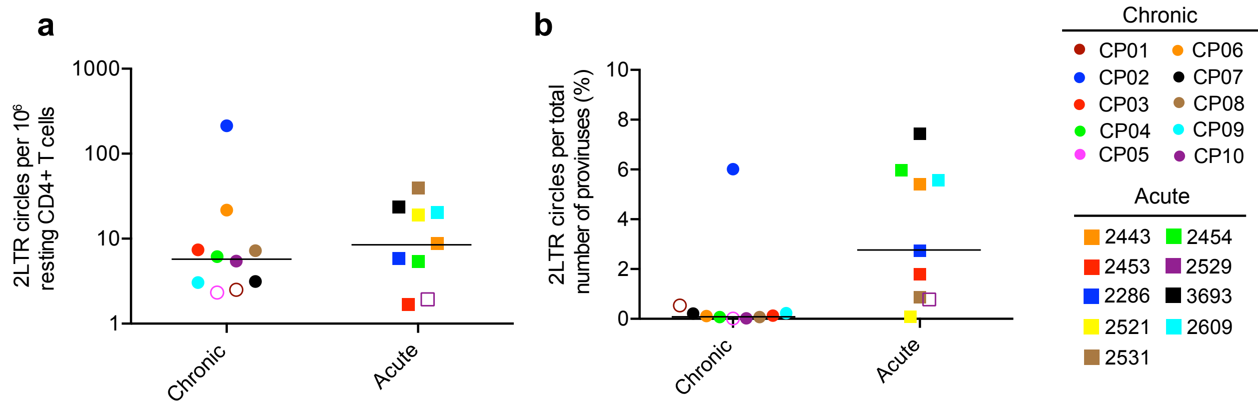


c

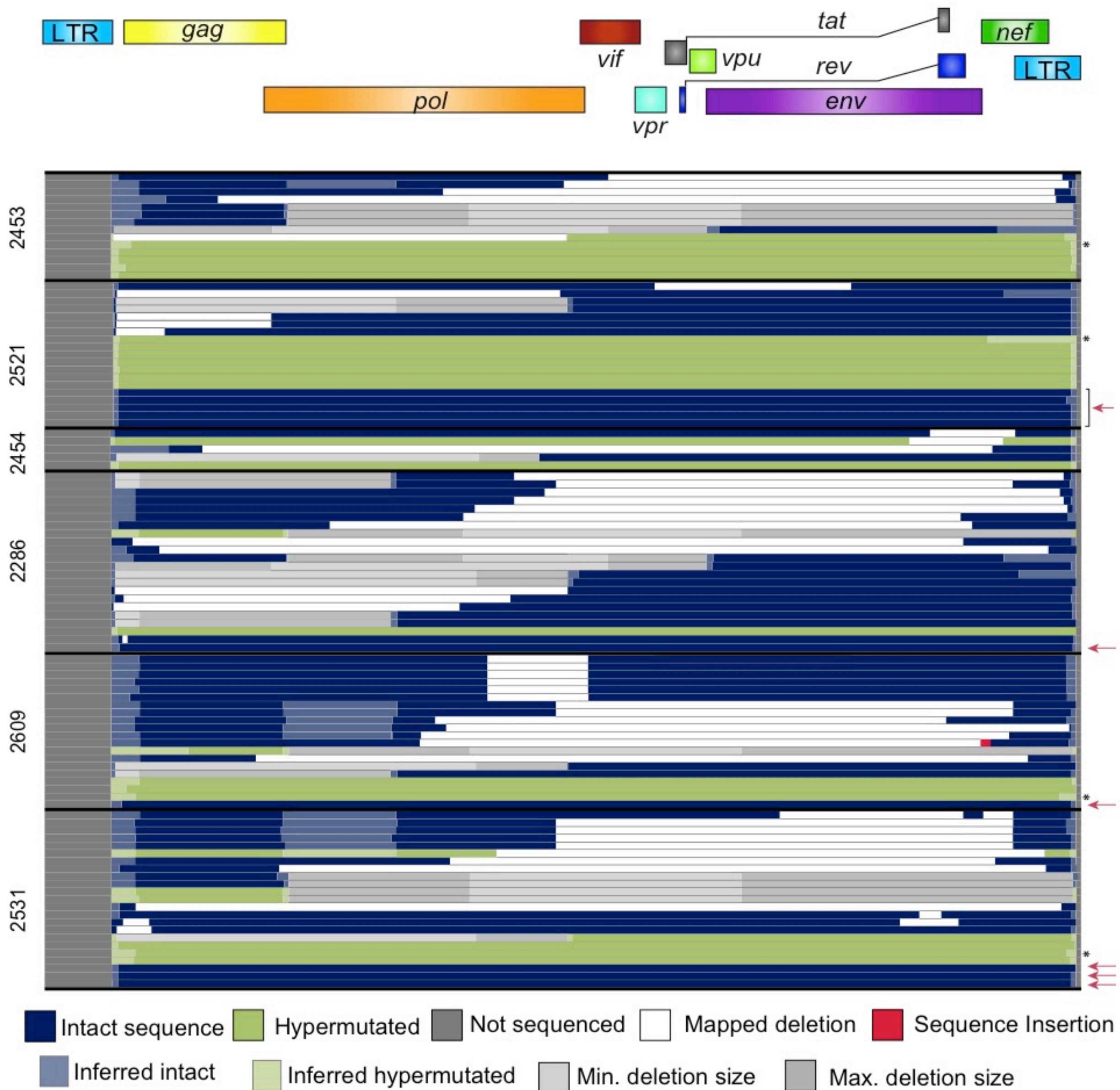


Supplemental Figure 3. High proportion of hypermutated sequences contain mutated start codons and premature stop codons. (**a**) Fraction of hypermutated sequences that contain a functional start codon for each HIV-1 gene. The ATG and second codon as well as the translated amino acid sequence are shown in blue. Sequences with G to A hypermutation by APOBEC3F/G affecting the ATG start codon are shown in red and orange with hypermutated nucleotides underlined. An asterisk denotes other types of mutations affecting the start codon (frameshift mutation, etc). Hypermutated sequences were identified using the Los Alamos hypermut algorithm[42]. This analysis was performed on all unique hypermutated sequences from acute and chronic phase subjects for which the relevant start codon was present. *n* indicates the number of

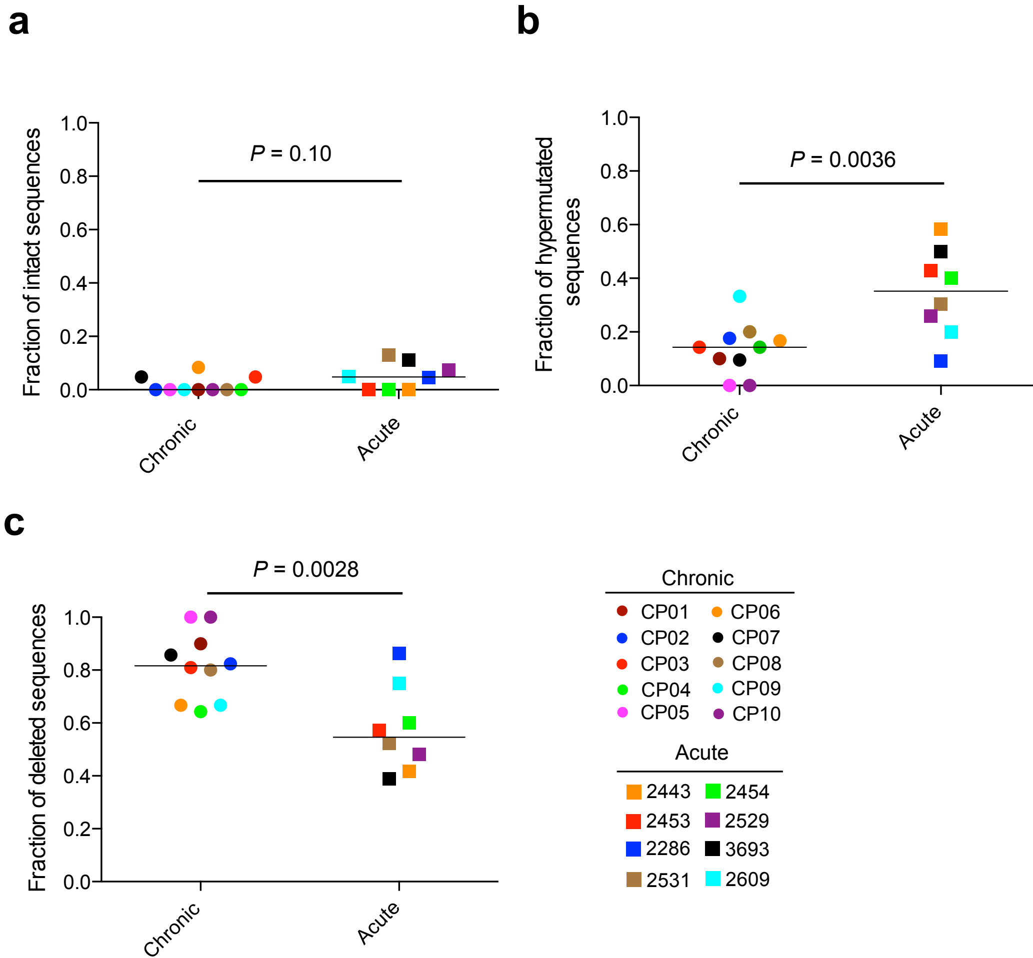
sequences analyzed. (**b**) Frequency of hypermutated sequences that contain internal, premature stop codons in each HIV-1 gene. Wild type (blue) sequences contained no premature stop codons, and mutated (red) sequences contained 1-50 premature stop codons per gene. (**c**) Frequency of hypermutated sequences that contain functional or defective copies of each HIV-1 gene. Proviruses with both an intact start codon and no internal stop codons were relatively rare and were considered to make a functional copy of a given gene (blue). The remaining proviruses contained defective copies of a given gene and had a functional start codon but contained internal stop codons (yellow), a mutated start codon but no internal stop codons (orange) or contained both a mutated start codon and internal stop codons (red).



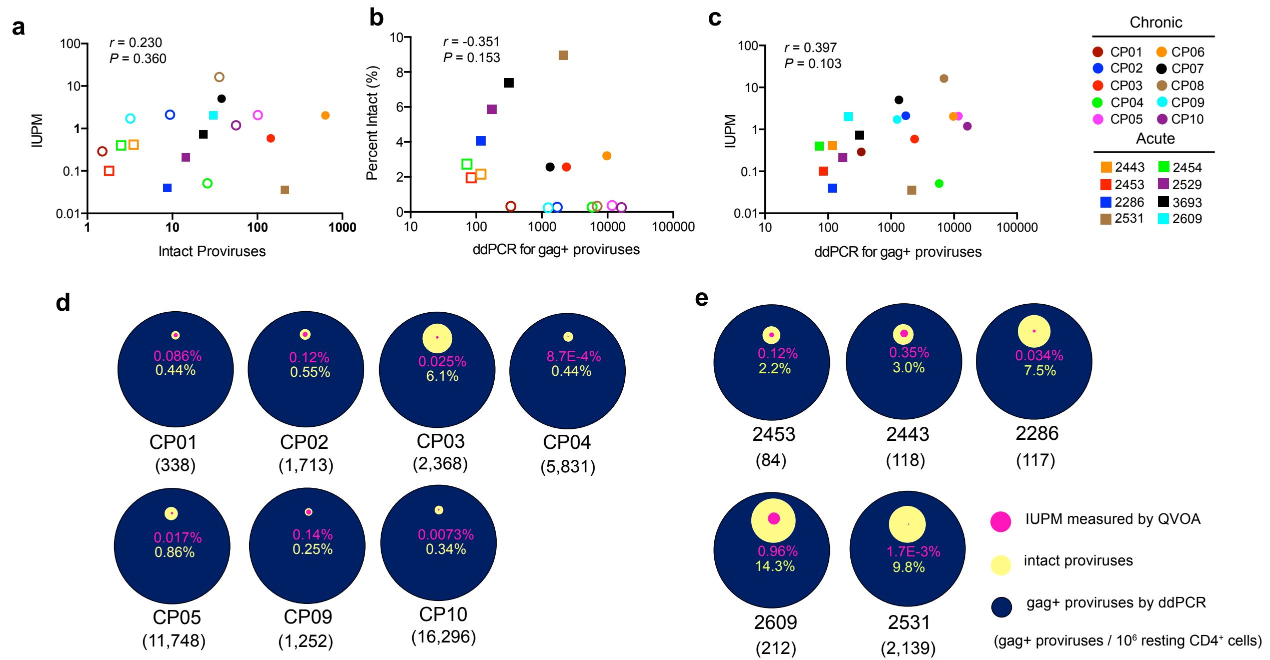
Supplemental Figure 4. 2LTR-circles are present at low levels in resting $CD4^+$ T cells of subjects treated during acute or chronic infection. (**a**) 2LTR circles were measured by ddPCR[14,40]. The cellular gene RPP30 was measured in a replicate well by ddPCR and used to calculate the total number of cells. Levels of 2LTR circles were expressed as a frequency per 10^6 resting $CD4^+$ T cells. In cases where no 2LTR circles were detected (represented by an open symbol), the limit of detection is plotted as determined by the cellular input for that sample. (**b**) Frequency of 2LTR circles relative to the total number of proviruses. The number of 2LTR circles was measured as in (a) and was normalized to the total number of proviruses per 10^6 resting $CD4^+$ T cells. The predicted total number of proviruses was calculated for each subject by correcting the *gag* ddPCR result (see Fig. 4) for the fraction of proviruses with deletions in *gag.*



Supplemental Figure 5. Characterization of proviral sequences in subjects initiating ART during acute/early infection. Maps of proviral clones from six additional subjects treated during the acute/early phase of infection. Each horizontal bar represents an individual clone identified through full genome sequencing. In cases where a deletion could not be precisely mapped, likely due to a deletion encompassing multiple forward or reverse primer binding sites, the maximum and minimum deletions sizes are plotted (in grey). If sequencing data shows a mapped deletion that removes primer-binding sites for other amplicons, the resulting missing sequence was inferred to be present (light blue or green). A pink arrow denotes full-length, genetically intact sequences. Black asterisks indicate likely full-length hypermutated proviruses that were not fully sequenced due to extreme hypermutation.



Supplemental Figure 6. Acutely-treated subjects have a higher fraction of hypermutated proviruses and a lower fraction of deleted proviruses than chronically-treated subjects. (**a**,**b**,**c**) Sequences were obtained by near full-length genome sequencing performed at limiting dilution. Horizontal bars indicate median values. Statistical significance was determined using a two-tailed unpaired t-test. The variance between groups was not statistically different. (**a**) The fraction of intact sequences was calculated by dividing the number of intact clones by the total number of clones observed for each subject. (**b**) The fraction of hypermutated sequences was calculated as described in (a) but using the number of hypermutated clones instead of the number of intact clones. Hypermutation was determined for each sequence by the Los Alamos hypermut algorithm[42]. (**c**) The fraction of deleted sequences was calculated as described in (a) but using the number of deleted clones instead of the number of intact clones.



Supplemental Figure 7. The number of intact proviruses does not correlate well with current assays. (**a**) Pearson correlation between the number of intact proviruses and the frequency of infected cells as measured by the QVOA (IUPM). The number of intact proviruses was calculated as the frequency of infected cells times the fraction of intact proviruses estimated for each subject using an empirical Bayesian model. Open symbols indicate subjects in which no intact proviruses were detected and the Bayesian value was plotted. (**b**) Pearson correlation between the fraction of intact proviruses and the frequency of cells with *gag*$^+$ proviruses as determined by ddPCR[14,40]. (**c**) Pearson correlation between the IUPM and frequency of cells with *gag*$^+$ proviruses as measured in (b). (**d,e**) Comparison of IUPM (pink), number of intact proviruses (yellow), and *gag*$^+$ proviral DNA (blue). All values are plotted as a percentage of the frequency of cells with *gag*$^+$ proviral DNA. The number of *gag*$^+$ proviruses per million resting CD4$^+$ T cells as measured by ddPCR is indicated in parenthesis for each subject. (**d**) Comparison for subjects treated during chronic infection (**e**) Comparison for subjects treated during acute/early infection.

Supplementary Table 1. Study participant characteristics

Pt. ID	Age	Sex	Race[a]	ART regimen	Infection Duration (months)	Time on ART (months)	Time on Suppressive ART (months)[b]	Treatment start time after infection (days)[c]	viral load (copies/ mL)[d]
2453	59	M	W	TDF/FTC, NVP	134	133	130	24	<50
2454	34	M	W	TDF/FTC/RPV	85	84	77	17	<50
2443	52	M	W	TDF/FTC/EVG/COBI	80	79	70	29	<50
2529	32	M	AS	TDF/FTC/RPV	29	26	25	97	<50
2521[e]	47	M	W	TDF/FTC, DGV	17	15	8	53	<50
3693	46	M	W	TDF/FTC, DGV	11	10	9	25	<50
2286	49	M	W	TDF/FTC/RPV	147	145	142	39	<50
2609	33	M	W	ABC/3TC/DTG	26	24	20	50	<50
2531	40	M	W/LA	TDF/FTC, RAL	34	32	28	57	<50
CP01	64	M	LA	TDF, 3TC, ETR	>179	130	127	>180	<50
CP02	61	M	AA	ABC/3TC, DRV/r, RAL	131	89	58	>180	<50
CP03[f]	76	M	W	ABC/3TC, ATV/r	>194	192	191	>180	<50
CP04	49	F	AA	TDF/FTC/EFV, DRV/r	322	210	129	>180	<50
CP05	56	F	AA	ABC, FTC, ETR, RAL	>214	>187	31	>180	<50
CP06	64	M	W	ABC, FTC, ETR	>263	>203	103	>180	<50
CP07	36	M	N/A	TDF/FTC/EFV	>124	65	55	>180	<50
CP08	45	M	AA	ETR, MVC, RAL	>131	>53	16	>180	<50
CP09	45	M	AA	3TC, DRV/r, RAL	311	132	103	>180	<50
CP10	58	M	AA	ETR, DRV/r, RAL	261	>189	91	>180	<50
VM01	33	M	AA	TDF/FTC/EFV	137	106	-	>180	4,719
VM02	20	M	AA	-	unknown	-	-	-	101,484

[a]AA, African-American; W, Caucasian; AS, Asian; LA, Latino; N/A, Not reported. Race was self-reported by study participants.
[b]Time of continuous suppression of plasma HIV-1 RNA to <50 copies/mL prior to enrollment.
[c]Treatment initiated within the first 3 months of infection is considered to be treated in the acute phase of infection.
[d]Plasma HIV-1 RNA levels for all patients were <50 copies/mL for at least 8 months before this study.
[e]Subject had a blip (~1000 copies/mL) 4 months before sampling. Proviral sequences from this individual are shown but were excluded from any analyses of reservoir size.
[f]Subject may have initiated treatment early in the course of infection but records are unclear and there is no prior negative HIV-1 test. Thus this individual is included as a chronically-treated subject.

Drug abbreviations ABC, abacavir; ATV, atazanavir; COBI, cobicistat; DRV, darunavir; DTG, dolutegravir; EFV, efavirenz; EVG, elvitegravir; ETR, etravirine; FTC, emtricitabine; MVC, maraviroc; NVP, nevirapine; RAL, raltegravir; RPV, rilpivirine; TDF, tenofovir disoproxil fumarate; 3TC, lamivudine; /r, boosted with ritonavir.

Supplementary Table 2. PCR primers and conditions

PCR	Length	Primer Name	HXB2 position	Primer sequence	Extension time
Outer PCR	9,064	BLOuterF	623 - 649	AAATCTCTAGCAGTGGCGCCCGAACAG	10 m
		BLOuterR	9,662 - 9,686	TGAGGGATCTCTAGTTACCAGAGTC	
Inner PCRs					
gag	1,448	5GagIn	836 - 857	GGGAAAAAATTCGGTTAAGGCC	1 m 30 s
		3GagIn	2,264 - 2,283	CGAGGGGTCGTTGCCAAAGA	
env	2,841	5EnvIn	6,201 - 6,231	GAGAAAGAGCAGAAGACAGTGGCAATG AGAG	3 m 30 s
		3EnvIn	9,007 - 9,042	CTTGTAAGTCATTGGTCTTAAAGGTACC TGAGGTCTG	
A	4,449	275F	646 - 666	ACAGGGACCTGAAAGCGAAAG	5 m
		3INOut	5,072 - 5,094	AATCCTCATCCTGTCTACTTGCC	
B	5,793	263F	651 - 672	GACCTGAAAGCGAAAGGGAAAC	6 m
		3AccOut	6,421 - 6,443	GGCATGTGTGGCCCARAYATTAT	
C	6,385	5INOut	3,248 - 3,270	ACTCCATCCTGATAAATGGACAG	6 m 30 s
		BLInnerR	9,604 - 9,632	GCACTCAAGGCAAGCTTTATTGAGGCTA	
D	4,778	5AccOut	4,899 - 4,922	CGGGTTTATTACAGGGACARCARA	5 m
		280R	9,650 - 9,676	CTAGTTACCAGAGTCACACAACAGACG	

Outer PCR cycling conditions:
94°C for 2 m; then 94°C for 30 s, 64°C for 30 s, 68°C for 10 m for 3 cycles; 94°C for 30 s, 61°C for 30 s, 68°C for 10 m for 3 cycles; 94°C for 30 s, 58°C for 30 s, 68°C for 10 m for 3 cycles; 94°C for 30 s, 55°C for 30 s, 68°C for 10 m for 21 cycles; then 68°C for 10 m. (30 total cycles)

Inner PCR cycling conditions: (**X** = extension time from the table above)
94°C for 2 m; then 94°C for 30 s, 64°C for 30 s, 68°C for **X m** for 3 cycles; 94°C for 30 s, 61°C for 30 s, 68°C for **X m** for 3 cycles; 94°C for 30 s, 58°C for 30 s, 68°C for **X m** for 3 cycles; 94°C for 30 s, 55°C for 30 s, 68°C for **X m** for 36 cycles; then 68°C for **X m**. (45 total cycles)