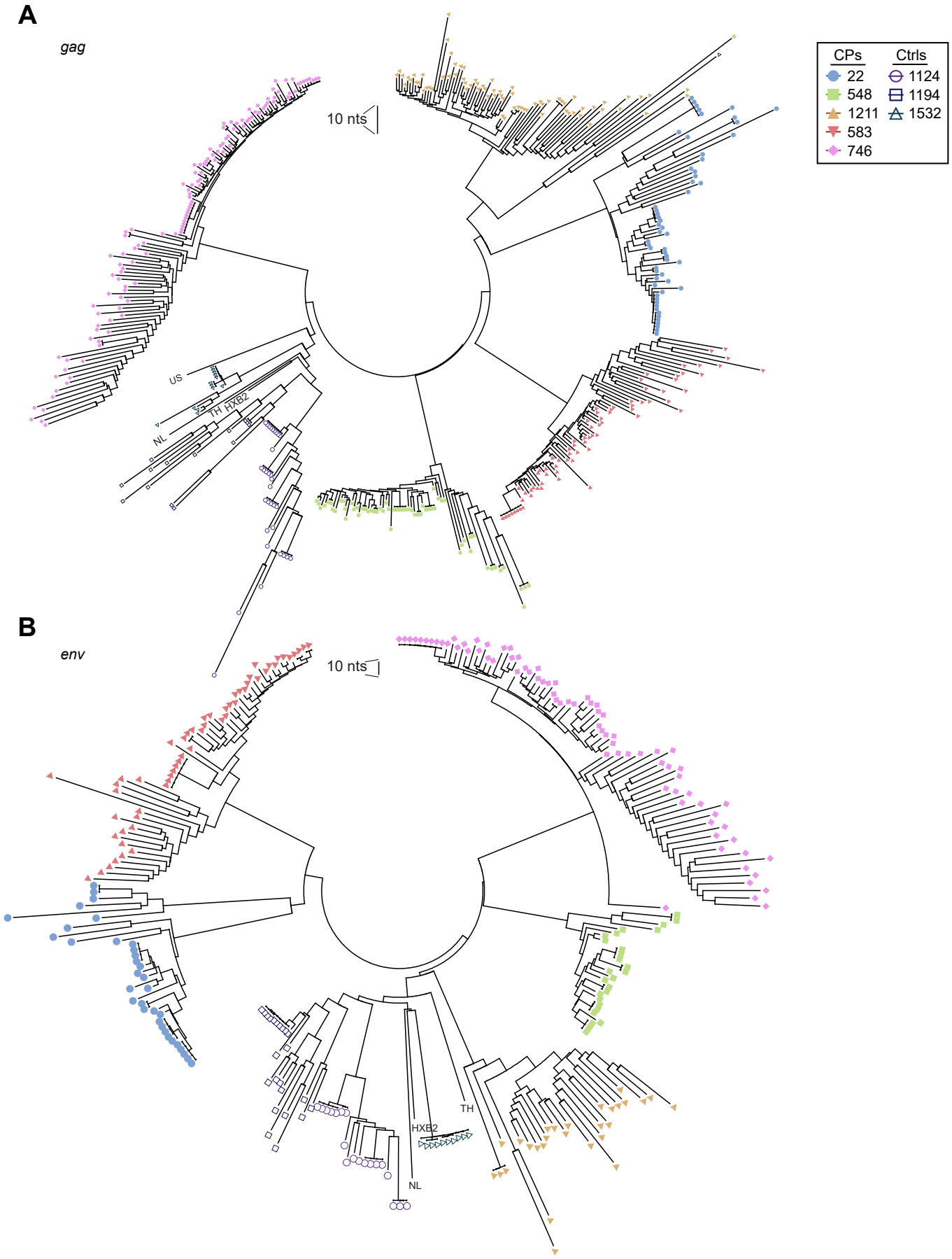
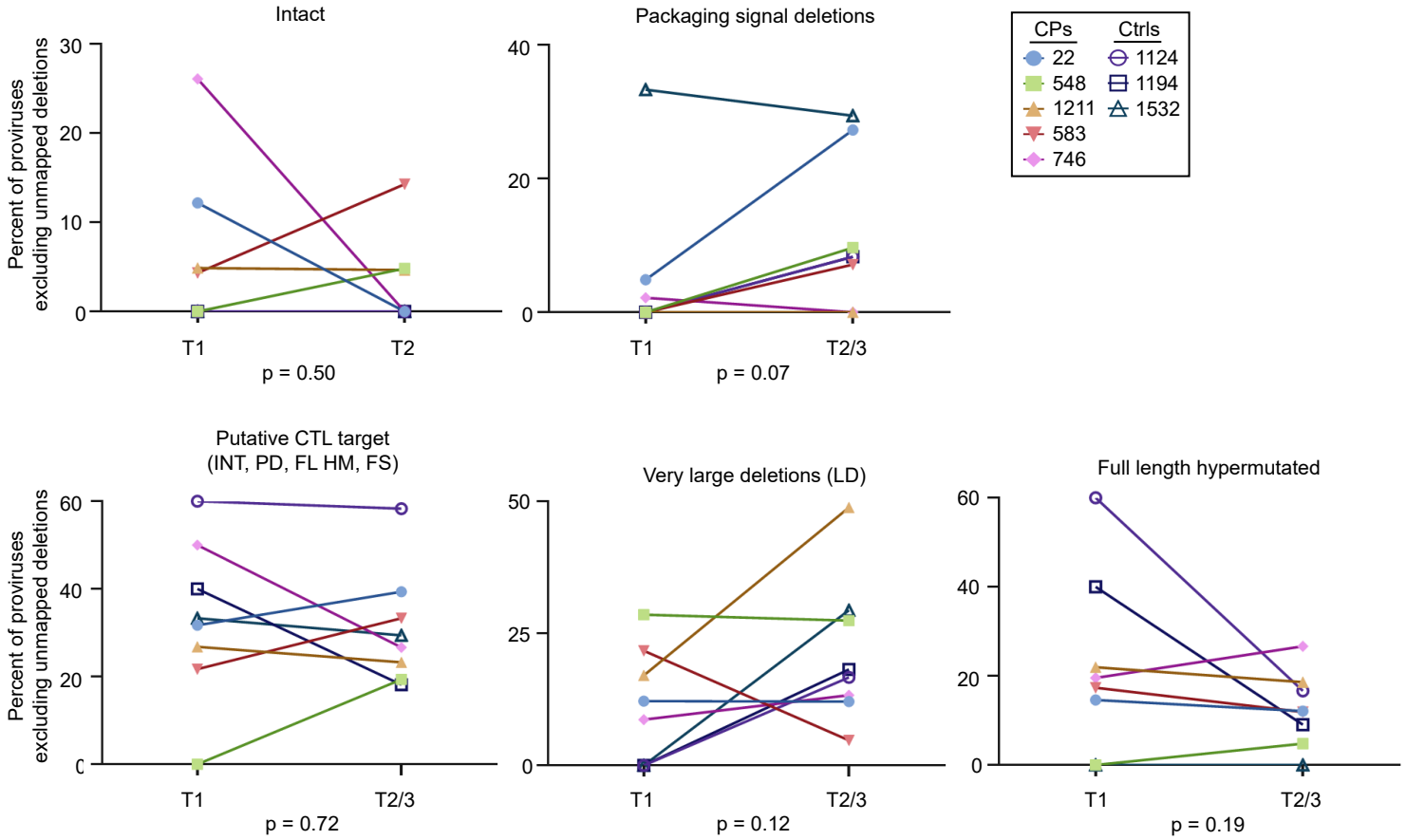
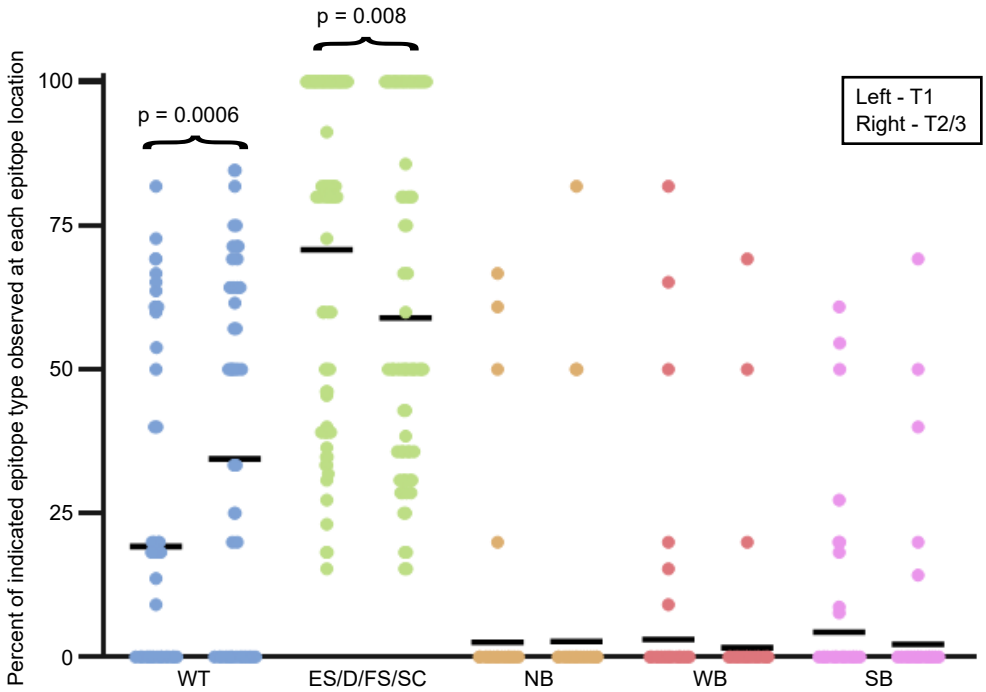


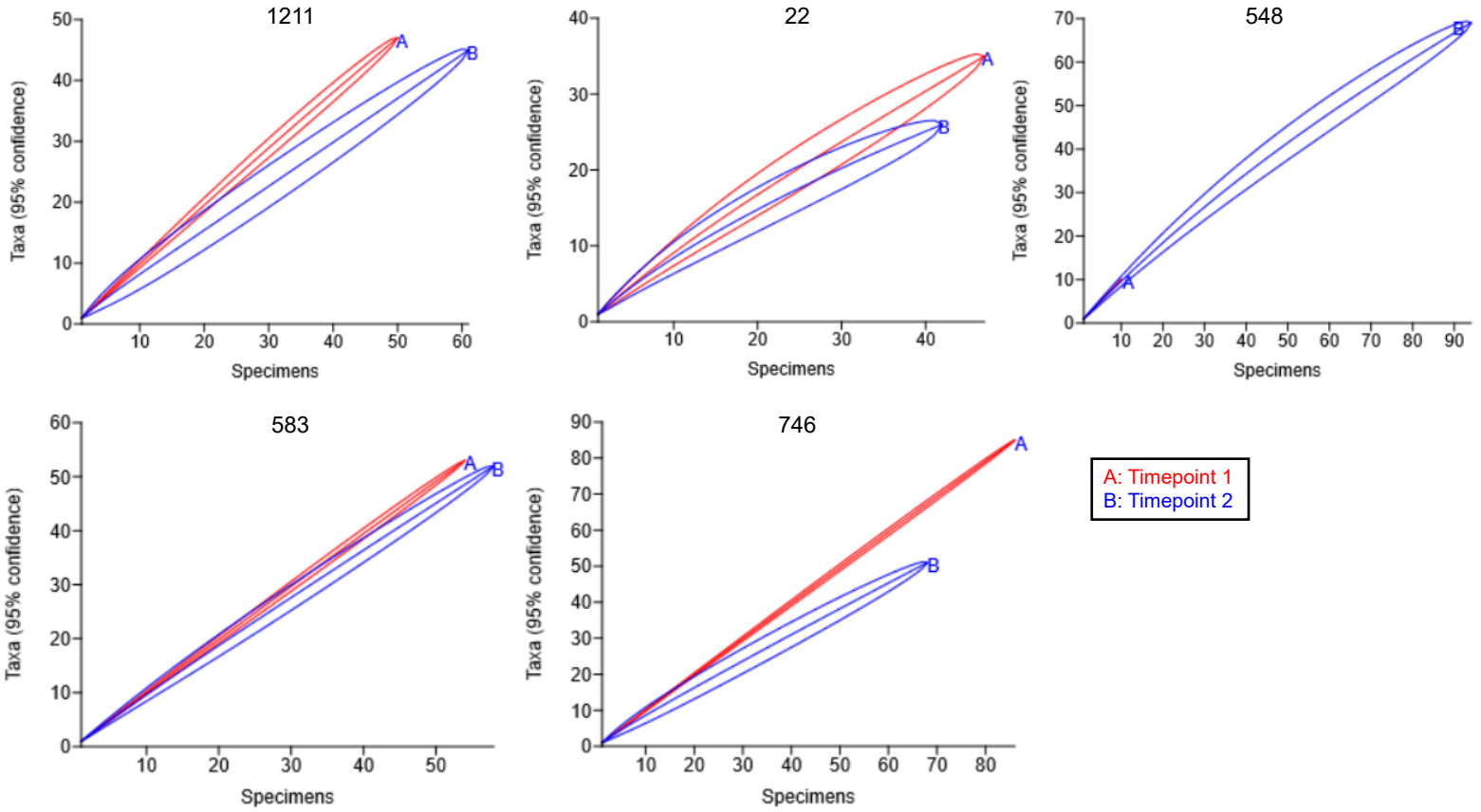
Supplementary Figure 1: Phylogenetic trees of HIV-1 provirus sequences from eight individuals on suppressive ART.
Phylogenetic trees of HIV-1 gag (A) and env (B) sequences obtained by near-full genome sequencing from eight participants who maintained suppressed viral loads on ART for years. Each individual's sequences clustered together. All sequences are HIV-1 subtype B, and reference subtype B sequences were used as outgroups to root the tree. Reference sequences HXB2, NL, TH, and US correspond to GenBank accession numbers K03455, AY423387, AY173951, and DQ853437, respectively. (A) Sequence analysis of gag for the 451 proviruses that had known sequence with no deletions between HXB2 positions 888 and 2143. One hypermutated sequence each from two individuals clustered with a third individual's hypermutated sequences due to commonly mutated positions. (B) Sequence analysis of env for the 277 proviruses that had known sequence with no deletions between HXB2 positions 6324 and 8190. CPs, chronic progressors. Ctrls, controllers.



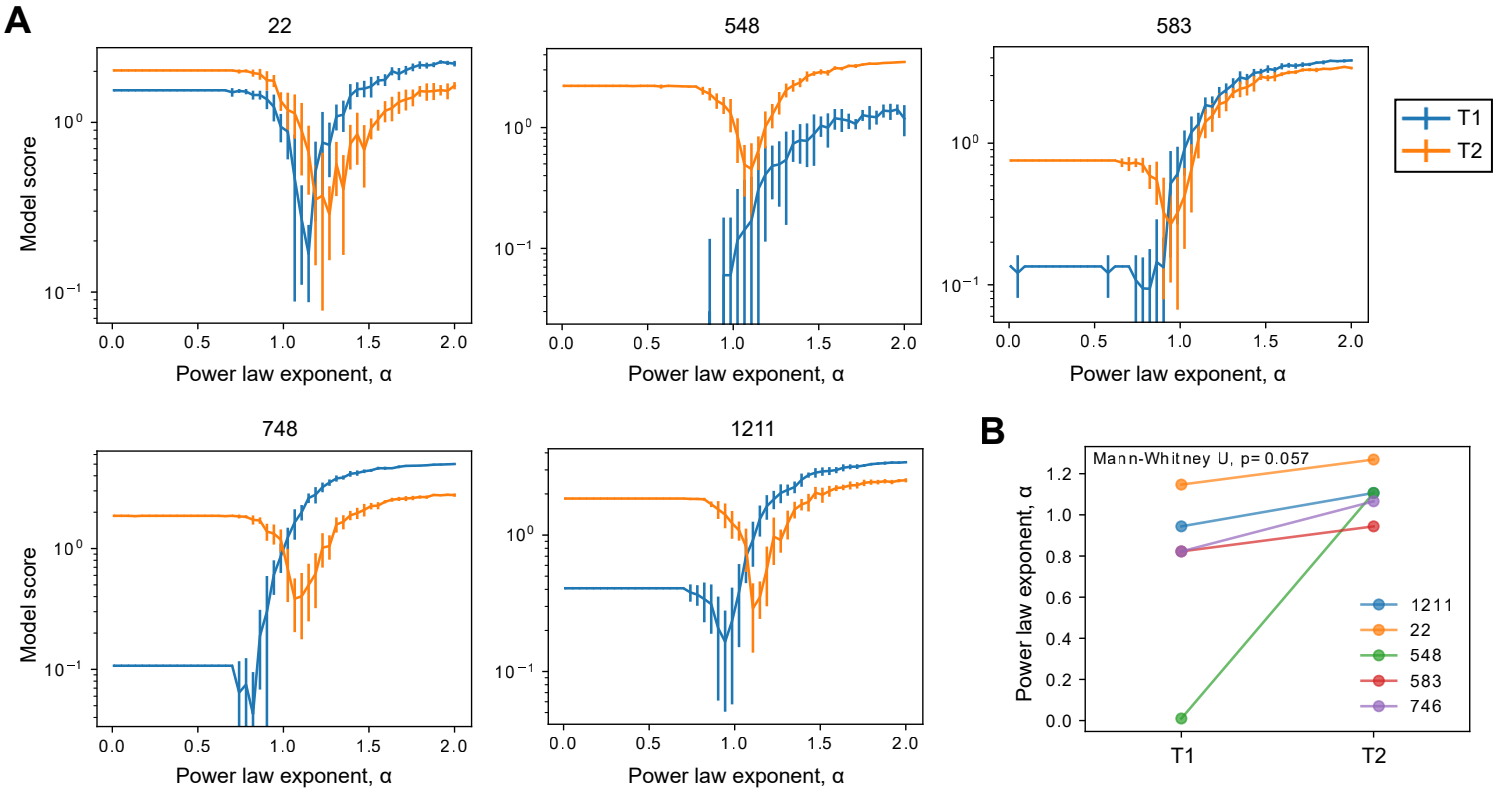
Supplementary Figure 2: Longitudinal analysis of provirus types over time on ART by near-full length sequencing omitting sequences with unmapped deletions. Percent of each indicated provirus type or grouping among total proviruses. Total proviruses here do not include those with unmapped deletions. No intact proviruses were found in the controllers on ART. The packaging signal deletion grouping includes PD proviruses that are hypermutated and those that are not hypermutated. A paired, two-tailed student's t test was applied; p values are shown. CPs, chronic progressors; Ctrls, controllers; INT, intact; PD, deletion within packaging signal or major splice donor site; FL HM, full-length hypermutated; FS, full-length and not hypermutated with one or more frameshifts affecting genes required for replication; LD, deletion of >75% of genome length.



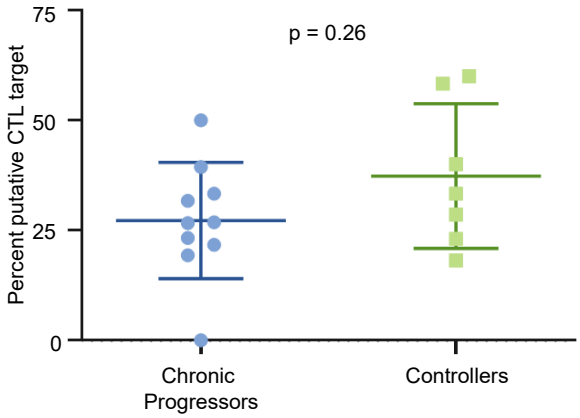
Supplementary Figure 3: Percentages of unrecognized epitope sequences in Gag, Pol, and Nef in intact, frameshifted, full-length hypermutated, and packaging signal deletion proviruses decrease over time in people on suppressive ART.
For each of 86 predicted dominant epitopes from 6 individuals on suppressive ART, the percent of the indicated epitope type among the proviruses with epitope sequence available at that site is plotted as one dot with mean shown in black. In each pair, the left represents T1 and the right,T2 and 3. Unrecognized epitopes include escaped epitopes (ES), epitopes that lie within a deleted region of the provirus (D), and epitopes with a preceding frameshift (FS) mutation or stop codon (SC). Epitope sequences that were uncharacterized in the literature are categorized as non-binder (NB), weak binder (WB), or strong binder (SB) based on predictive software (NetMHC 4.0) and relevant HLA allele. P values were generated to test the hypothesis that the percentage of each type of epitope did not change over time on ART using the nonparametric Wilcoxon signed-rank test. Only proviruses that were intact, full-length with a frameshift, full-length hypermutated, and packaging signal deleted were included in this analysis because they can theoretically generate epitopes recognized by cytotoxic T lymphocytes. Because participants 548 and 1532 each had one timepoint in which 0-1 proviruses in this grouping were observed, they were excluded from this analysis. WT, wild-type.



Supplementary Figure 4: Rarefaction curves demonstrating decreases in observed proviral richness over long periods of time on suppressive ART. Rarefaction curves of the 5 chronic progressor participants demonstrate that provirus richness is lower at T2 compared to T1. Participant 548 had a low sampling depth at T1. Participants 1211 and 746 are also shown in Fig. 7e. Rarefaction analysis by Past3 (https://folk.uio.no/ohammer/past/).



Supplementary Figure 5. Estimated power law exponents to extrapolate the clonal abundance of the entire HIV-infected CD4+ T cell pool. For each individual and each time point, the observed rank abundance of infected resting CD4+ T cell clones was used to extrapolate the clonal abundance of the entire HIV-infected resting CD4+ T cell pool. True richness was assumed to be on the order of 1e6 infected cells, proportional to total HIV DNA by IPDA in each individual. An inference tool was implemented in Python to estimate the best-fit power law exponent. 50 power law exponents were simulated ranging from 0-2, and 10 replicate simulations were performed for each exponent. (A) Results of fitting are illustrated. Lower model score is better, and lines represent average score with error bars representing standard deviation of score across replicate simulations. (B) For each individual, the best-fit power law exponent increased between timepoint 1 and 2. A non-parametric Mann-Whitney U test suggests a significant trend towards increasing power law exponent over time during suppressive antiretroviral therapy, meaning the HIV-infected resting CD4+ T cells are increasingly concentrated among predominant large clones over time. By later in suppressive ART, more than 50% of HIV-1 proviruses may be contained in the top 100-1000 clones.



Supplementary Figure 6: Percentages of proviruses that are hypothesized to be immunogenic upon expression are not different in chronic progressors and controllers on ART.
A plot of the percent putative CTL target proviruses at each timepoint for chronic progressors (left) and controllers (right) on ART by near-full length sequencing. Here, putative CTL target means the grouping of intact proviruses, full-length hypermutated proviruses, full-length and containing one frameshift proviruses, and packaging signal deleted proviruses. The nonparametric Kruskal-Wallis test was applied and p value is shown.

Supplementary Table 1. Intact provirus decay rates by IPDA

Participant ID	Intact provirus $t_{1/2}$ (month)	Decay rate (/month)
22	34.10	0.020
548	23.13	0.030
1211	-71.95	-0.010
583	27.36	0.025
746	36.37	0.019
1124	21.33	0.032
1194	27.42	0.025
1532	22.68	0.031

Supplementary Table 2. Best-defined epitopes in Gag, Pol, and Nef for each participant's HLA type.

Participant	Epitope (in HXB2)	HXB2 coordinates	Relevant MHC
	Pol		
22	ITLWQRPLV (HXB2 vTLWQRPLV)	2259-2285	A*68:02
22	GKKAIGTVL (HXB2 GhKAIGTVL)	2454-2480	B*15:03
22	IAMESIVIW (HXB2 IttESIVIW)	3672-3698	B*58:01
22	GAETFYVDGA	3855-3884	A*68:02
22	IQQEFGIPY (HXB2 IkQEFGIPY)	4632-4658	B*15:03
22	RKAKIIRDY	5016-5042	B*15:03
	Nef		
22	KAAFDLSFF (HXB2 KAAvDLShF)	9040-9066	B*58:01
22	HTQGYFPDW (HXB2 HTQGYFPD*)	9142-9168	B*58:01
22	WRFDSRLAF	9343-9369	B*15:03
	Gag		
22	RSLYNTVATLY	1015-1047	B58
22	VKVVEEKAF	1255-1281	B*15:03
22	TSTLQEQIGW	1507-1536	B*58:01
22	YVDRFFKTL (HXB2 YVDRFyKTL)	1676-1701	B*15:03
22	QASQEVKNW	1711-1737	B*58:01
	Pol		
548	TVLDVGDAY	2868-2894	B*3501
548	VPLDEDFRKY	2901-2930	B*3501
548	NETPGIRYQY	2958-2987	B18
548	AIFQSSMTK	3021-3047	A*1101
548	HPDIVIYQY (HXB2 nPEIVIYQY)	3072-3098	B18 and B*3501
548	IPLTEEAEL	3426-3452	B*3501
548	QIIEQLIKK	4107-4133	A*1101
548	AVFIHNFKRK	4764-4793	A*1101
548	IIATDIQTK	4836-4862	A11
	Nef		
548	VPLRPMTY	9016-9039	B*3501
548	AVDLSHFLK	9046-9072	A*1101
548	RRQDILDLWVY or KRQEILDLWVY	9109-9141	B18 and Cw7
548	YFPDWQNYT (HXB2 YFPD*QNYT)	9154-9180	B*3501
548	YPLTFGWCY	9199-9225	B*1801
	Gag		
548	WASRELERF	895-921	B*3501
548	TLYCVHQK (HXB2 TLYCVHQr)	1039-1062	A*1101
548	NSSKVSQNY (HXB2 hSnqVSQNY)	1159-1185	B*3501
548	HPVHAGPIA	1435-1461	B*3501
548	PPIPVGDIY (HXB2 PPIPVGeIY)	1549-1575	B*3501
548	FRDYVDRFYK	1666-1695	B*1801
548	ACQGVGGPGHK	1834-1866	A*1101
	Pol		

746	GKKAIGTVL (HXB2 GhKAIGTVL)	2454-2480	B*1503
746	LVGPTPVNI	2478-2504	A*0201
746	ALVEICTEM	2646-2672	A*0201
746	YTAFTIPSV (HXB2 YTAFTIPSi)	2928-2954	A2
746	VIYQYMDDL	3084-3110	A*0201
746	ILKEPVHGV	3474-3500	A*0201
746	IQQEFGIPY (HXB2 IkQEFGIPY)	4632-4658	B*1503
746	RKAKIIRDY	5016-5042	B*1503
	Nef		
746	YPLTFGWCY	9199-9225	B*5301
746	WRFDSRLAF	9343-9369	B*1503
	Gag		
746	SLYNTVATL	1018-1044	A*0201, A*0202
746	VKVIEEKAF (HXB2 VKVvEEKAF)	1255-1281	B*1503
746	YVDRFFKTL (HXB2 YVDRFyKTL)	1675-1701	B*1503
746	QASQEVKNW	1711-1737	B*5301
	Pol		
1211	ITLWQRPLV (HXB2 vTLWQRPLV)	2259-2285	A*6802
1211	DTVLEEWNL (HXB2 DTVLEEmsL)	2340-2366	A*6802
1211	KQNPDIVIY	3066-3092	A*3002
1211	KLNWASQIY	3336-3362	A*3002
1211	RMRGAHTNDV	3615-3644	A*3002
1211	GAETFYVDGA	3855-3884	A*6802
1211	KIQNFRVYY	4884-4910	A*3002
	Gag		
1211	RSLYNTVATLY	1015-1047	A*3002
	Pol		
583	LVGPTPVNI	2478-2504	A*0201
583	IETVPVKL	2562-2585	B*4001
583	ALVEICTEM	2646-2672	A*0201
583	YTAFTIPSV (HXB2 YTAFTIPSi)	2928-2954	A2
583	VIYQYMDDL	3084-3110	A*0201
583	IEELRQHLL	3153-3179	B*4001
583	ILKEPVHGV	3474-3500	A*0201
583	IVTDSQYAL	4032-4058	Cw*0802
	Nef		
583	LEKHGAITS	8905-8931	B*4001
583	AAVDLSHFL	9043-9069	Cw*0802 AND Cw3
583	KEKGGLEGL	9070-9096	B*4001
583	PLTFGWCYKL	9202-9231	A*0201
583	VLEWRFDSRL	9334-9363	A*0201
	Gag		
583	SLYNTVATL	1018-1044	A*0201
583	IEIKDTKEAL	1063-1092	B*4001
583	TPQDLNTML	1327-1353	Cw*0802

583	DRFYKTLRA	1681-1707	B*1402
583	CRAPRKKGC	2002-2028	B14
	Pol		
1124	ITLWQRPLV (HXB2 vTLWQRPLV)	2259-2285	A*6802
1124	DTVLEEWNL (HXB2 DTVLEEmsL)	2340-2366	A*6802
1124	GAETFYVDGA	3855-3884	A*6802
1124	NNETPGIRY	2955-2981	B18
1124	NPEIVIYQY (HXB2 NPdIVIYQY)	3072-3098	B18
1124	HTDNGSNF	4569-4592	Cw5
1124	STTVKAACWW (HXB2 gaTVrAACWW)	4596-4625	B57
	Nef		
1124	QDILDLWVY (HXB2 QDILDLWiY)	9115-9141	B18
1124	HTQGYFPDW (HXB2 HTQGYFPD*)	9142-9168	B*5703
1124	YPLTFGWCY	9199-9225	B*1801
	Gag		
1124	KAFSPEVI	1273-1296	B*5703
1124	FRDYVDRFY	1666-1692	B*1801
1124	AEQASQEVKNWM	1705-1740	Cw5
	Pol		
1194	ALVEICTEMEK	2646-2678	A*0301
1194	LVDFRELNK	2769-2795	A*0301
1194	GIPHPAGLK	2826-2852	A*0301
1194	SPAIFQSSM	3015-3041	B7 AND A0301 (2 aa longer)
1194	QIYPGIKVR	3354-3380	A*0301
1194	IVTDSQYAL	4032-4058	Cw*0802
1194	AVFIHNFKRK	4764-4793	A*0301
	Nef		
1194	FPVTPQVPLR	8998-9027	B*0702
1194	QVPLRPMTYK	9013-9042	A*0301
1194	RPMTYKAAL (HXB2 RPMTYKAAv)	9025-9051	B*0702
1194	AAVDLSHFL	9043-9069	Cw*0802 AND A0301
	Gag		
1194	KIRLRPGGK	841-867	A*0301
1194	RLRPGGKKK	847-873	A*0301
1194	SPRTLNAWV	1231-1257	B*0702
1194	TPQDLNTML	1327-1353	B*0702 AND Cw*0802
1194	HPVHAGPIA	1435-1461	B7
1194	DRFYKTLRA	1681-1707	B*1402
1194	GPGHKARVL	1852-1878	B*0702
	Pol		
1532	LVGPTPVNI	2478-2504	A*0201
1532	ALVEICTEM	2646-2672	A*0201
1532	YTAFTIPSV (HXB2 YTAFTIPSi)	2928-2954	A2
1532	AIFQSSMTK	3021-3047	A*1101

1532	VIYQYMDDL	3084-3110	A*0201
1532	ILKEPVHGV	3474-3500	A*0201
1532	IYQEPFKNLK	3570-3599	A*1101
1532	QIIEQLIKK	4107-4133	A*1101
1532	AVFIHNFKRK	4764-4793	A*1101
1532	IIATDIQTK	4836-4862	A11
	Nef		
1532	QVPLRPMTYK	9013-9042	A*1101
1532	AVDLSHFLK	9046-9072	A*1101
1532	KRQEILDLWVY (HXB2 rRQdILDLWiY)	9109-9141	Cw7
1532	PLTFGWCYKL	9202-9231	A*0201
1532	VLEWRFDSRL	9334-9363	A*0201
	Gag		
1532	SLYNTVATL	1018-1044	A*0201
1532	TLYCVHQK (HXB2 TLYCVHQr)	1039-1062	A*1101
1532	ACQGVGGPGHK	1834-1866	A*1101

Supplementary Table 3. Proportion of indicated epitope types in Gag, Pol, and Nef over number of proviruses with known sequence at the epitope location for all proviruses.

Epitope type	Mean proportion not found in clones	Mean proportion found in clones	Wilcoxon two tailed p-value
Escaped, Deleted, Preceded by Stop Codon or Frameshift	0.6632	0.5361	<0.00001
Wild-type	0.2424	0.3549	0.00194
Deleted	0.3218	0.2924	0.05118
Frameshift	0.0092	0.0120	0.06876
Stop codon	0.2880	0.1707	<0.00001
Non-binder	0.0419	0.0529	0.72786
Escaped	0.0445	0.0609	0.76418
Strong binder	0.0282	0.0239	0.63836
Weak binder	0.0244	0.0322	0.95216