Intact HIV-1 proviruses accumulate at distinct chromosomal positions during prolonged antiretroviral therapy

Kevin B. Einkauf, Guinevere Q. Lee, Ce Gao, Radwa Sharaf, Xiaoming Sun, Stephane Hua,
Samantha M. Y. Chen, Chenyang Jiang, Xiaodong Lian, Fatema Z. Chowdhury,
Eric S. Rosenberg, Tae-Wook Chun, Jonathan Z. Li,
Xu G. Yu, Mathias Lichterfeld

Supplemental Material

Supplemental Material

Supplemental Figure 1: Comparison of intact near full-length HIV-1 sequences with and without prior whole-genome amplification in the three study participants. Horizontal phylogenetic trees demonstrating near full-length, intact HIV-1 sequences generated by single-genome amplification with (black) or without (red/blue) prior multiple displacement amplification. Proviral sequences amplified using primers spanning near full-length HIV-1 are denoted by "FLv3"; sequences amplified using five overlapping primer sets spanning near full-length HIV-1 are classified by "TM5". Sequences retrieved from viral outgrowth assays are indicated as "qVOA" and highlighted in blue. Grey boxes highlight sequences with 100% sequence-identity; primer binding sites were not considered for clonality analysis. *Highlights proviral species that shares a similar integration site with other intact sequences but differs from these sequences by 1bp. For patient 1, all sequences were generated from the same cell sample; for patients 2 and 3, sequences were generated from an identical sample collected in 2016, unless a different year is indicated in the sequence identification number.

Supplemental Figure 2: Molecular diagrams of the HIV-1 reservoir landscape in the three study participants. Diagrams demonstrate individual proviral species identified by near full-length whole genome amplification in the three study participants. Striped bars indicate internal inversion.

Supplemental Figure 3: Site-specific analysis of chromosomal integration sites at 3'-LTR and 5'-LTR viral-host borders. Diagrams show proviral sequences for which 5'-LTR and 3'-LTR integration site coordinates were separately assessed. 5'- and 3'-LTR-associated integration sites of the same proviral sequence consistently showed a 5bp duplication of host DNA at both the 5' and 3' human-HIV-1 junctions, consistent with previous literature. This results in a 4bp difference in integration site coordinates, which are defined as the last human base before the 5'-LTR in the 5'-LTR assay, and as the first human base after the 3'-LTR in the 3'-LTR assay.

Supplemental Figure 4: Analysis of HIV-1 3'-LTR and 5'-LTR promoter sequences in proviral species isolated from the three study participants. Figure shows HIV-1 Highlighter plots (Los Alamos National Laboratory) of the 5'-LTR regions identified during chromosomal integration site analysis in patient 2. In patient 1 and 3, 5'-LTR sequences were not always available and were inferred from 3'-LTR sequence data. In all three patients, the promoter region was highly conserved, and diversity was mainly attributed to single base substitutions. Primer-binding sites were removed from all analyses to remove primer bias.

Supplemental Figure 5: Individual patient analyses of chromosomal integration site features of intact and defective proviruses. (A-B): For each individual patient, pie charts reflecting proportion of intact and defective HIV-1 sequences located in genic vs. non-genic/pseudogenic regions (A), and with the same or opposite orientation relative to host genes (among sequences integrated in genes, B). Integration sites associated with multiple genes and mixed orientations to host genes were not considered for the analysis in (B). Significance was tested using two-tailed chi-square tests; nominal p-values are reported. Clonal sequences were only counted once.

Supplemental Figure 6: Chromosomal integration site features of intact and defective proviruses, counting all clonal sequences individually. (A-B): Pie charts reflecting proportion of intact and defective HIV-1 sequences located in genic vs. non-genic/pseudogenic regions (A), and with the same or opposite orientation relative to host genes (among sequences integrated in genes, B). Integration sites associated with multiple genes and mixed orientations to host genes were not considered for the analysis in (B). (C-D): Pie charts indicating the proportion of intact and defective HIV-1 sequences located in regions with defined repetitive genetic elements (C) (SINE: short interspersed nuclear element; LINE: long interspersed nuclear element; LTR: long-terminal repeat retrotransposon; DNA: DNA transposon) and in exons or introns (D). (A-D): Clonal proviruses were counted as individual sequences.

Supplemental Figure 7: Positions of proviral chromosomal integration sites relative to host transcriptional start sites and accessible chromatin, counting all clonal sequences individually. (A/D): Combined individual value/box and whisker plots indicating the chromosomal distance between HIV-1 integration sites and the most proximal TSS listed in Ensembl v86 (databank), or identified through analysis of expressed RNA-species located within the boundaries of the host gene, using autologous RNA-Seq data from indicated cell populations. (B/E): Combined individual value/box and whisker plots reflecting the chromosomal distance between integration sites and the center of the most proximal ATAC-Seq peaks in indicated CD4 T cell populations. (C/F): Gene expression intensity of host genes harboring intact or defective proviral integration sites, normalized to the chromosomal distance between integration sites and the most proximal TSS determined using autologous RNA-Seq data as described in (A/D). Box and whiskers reflect median, 25% and 75% percentiles, and minimum/maximum levels. Significance was calculated using two-tailed Mann Whitney U tests; nominal p-values are reported [* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$, **** indicates $p<0.0001$]. (A, B, C) reflect data from study patient 1 [41 intact, 23 defective proviruses]; (D, E, F) show data from study patients 2 and 3 [59 intact, 61 defective proviruses]. For all data in this figure, clonal proviral sequences were included as individual datapoints.

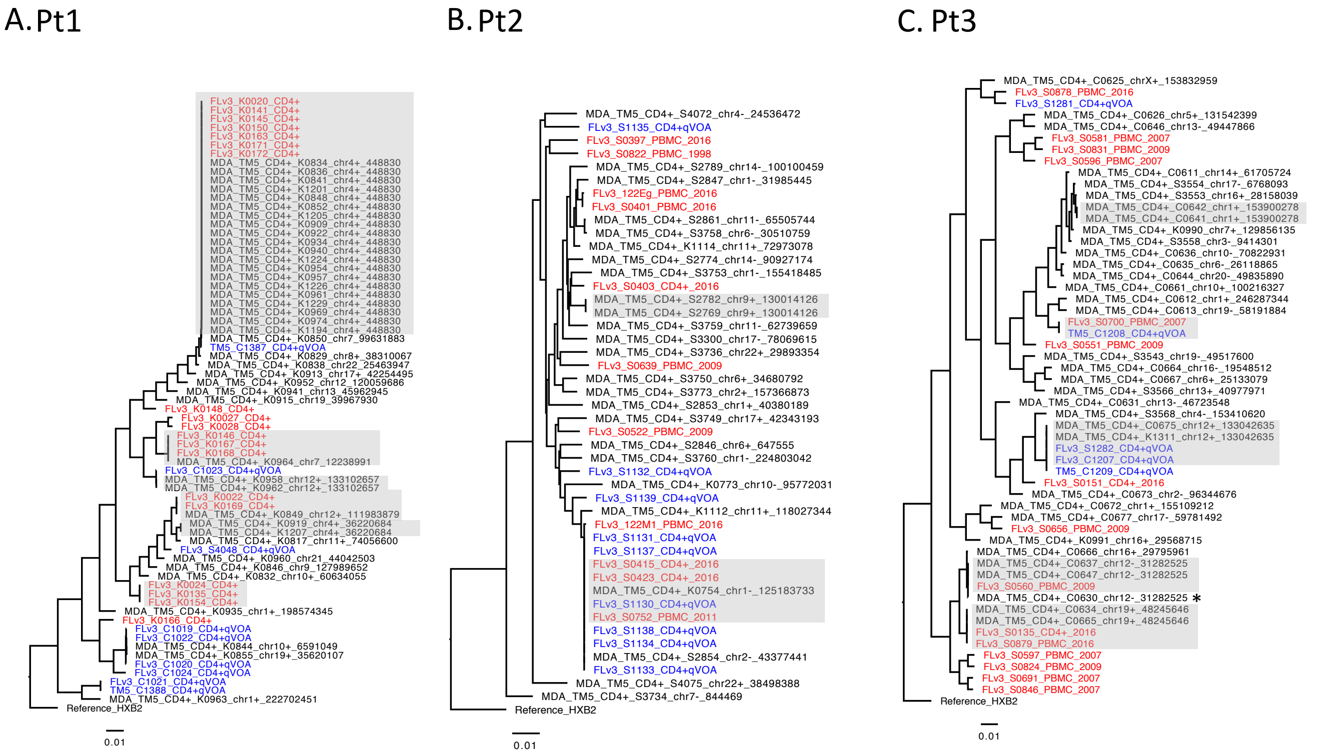
Supplemental Figure 8: Transcriptional pattern of host genes harboring HIV-1 integration sites. Violin plots show transcriptional activity of host genes containing intact or defective HIV-1 proviruses; transcriptional data of all expressed genes (background) are shown for comparison in each patient. Significance was tested using non-FDR-adjusted, two-tailed Wilcoxon tests (* indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$, **** indicates $p<0.0001$).

Supplemental Table 1: Clinical and demographical characteristics of study participants.

Supplemental Table 2: Phylogenetic distance between intact proviral sequences analyzed by near full-length sequencing after multiple displacement amplification.

Supplemental Table 3: List of integration sites for intact and defective proviruses in the three study patients. ISLA: integration site loop amplification; nrLAM: non-restrictive linear amplification-mediated PCR; 5'/3'-SSPCR: site-specific PCR at 5'-/3'-LTR viral-host border; SINE: short interspersed nuclear element; LINE: long interspersed nuclear element; LTR: long-terminal repeat retrotransposon; DNA: DNA transposon. For LentiX assays, restriction enzyme used is listed in parentheses.

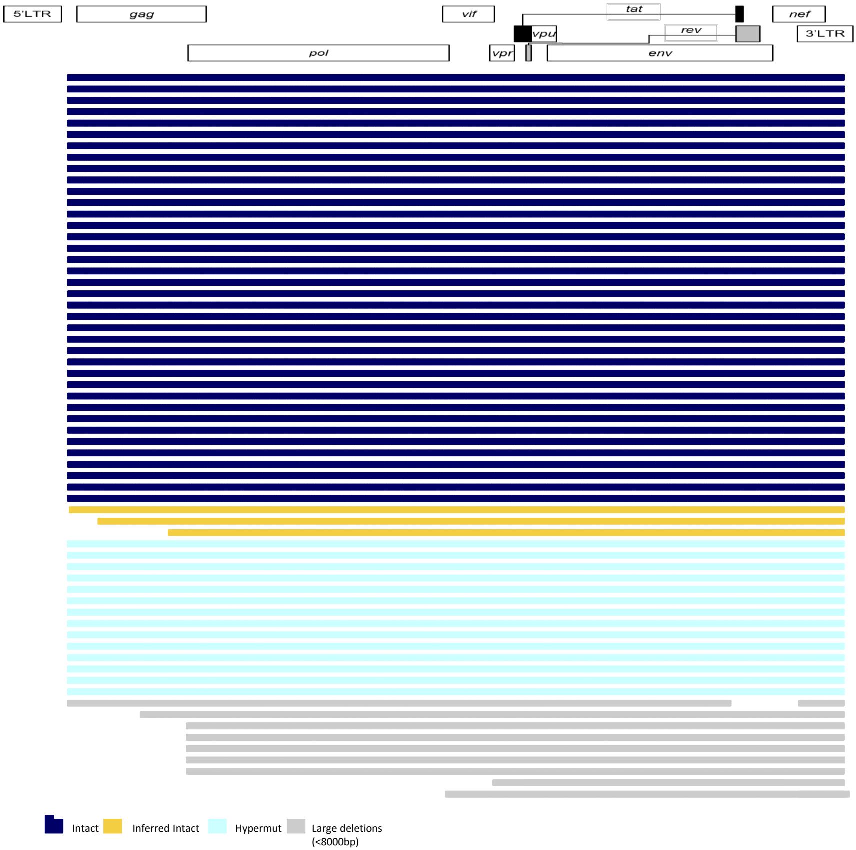
Supplemental Figure 1



* C0630 was 1bp different from other members in the same clone family

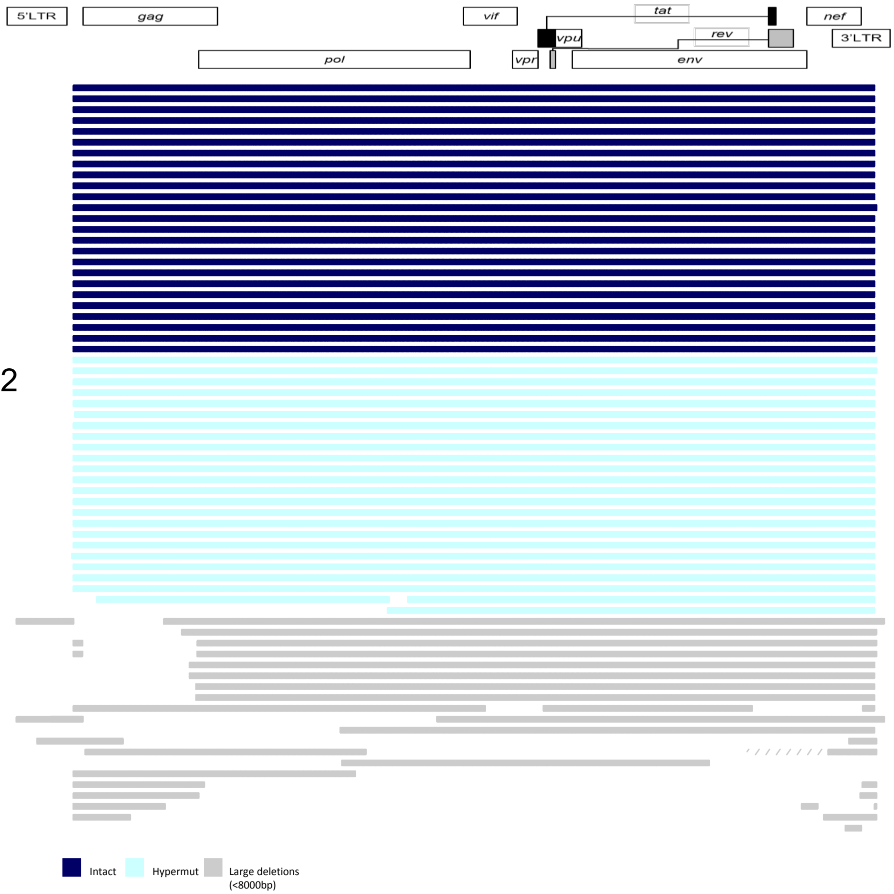
Supplemental Figure 2A

Virogram: Pt1



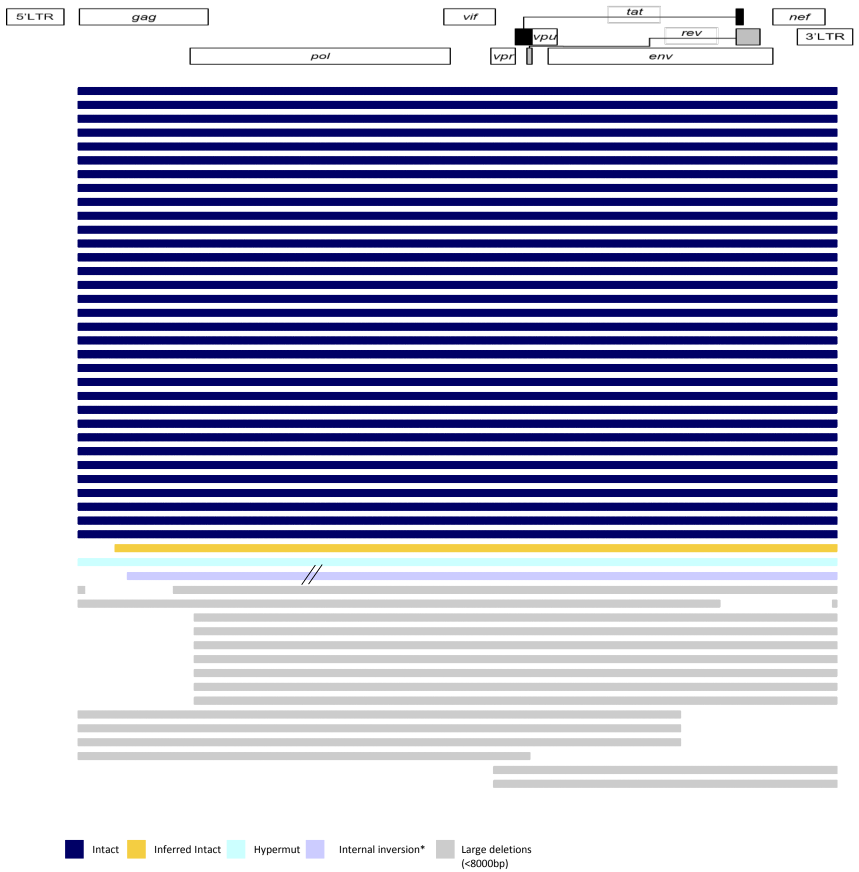
Supplemental Figure 2B

Virogram: Pt2



Supplemental Figure 2C

Virogram: Pt3



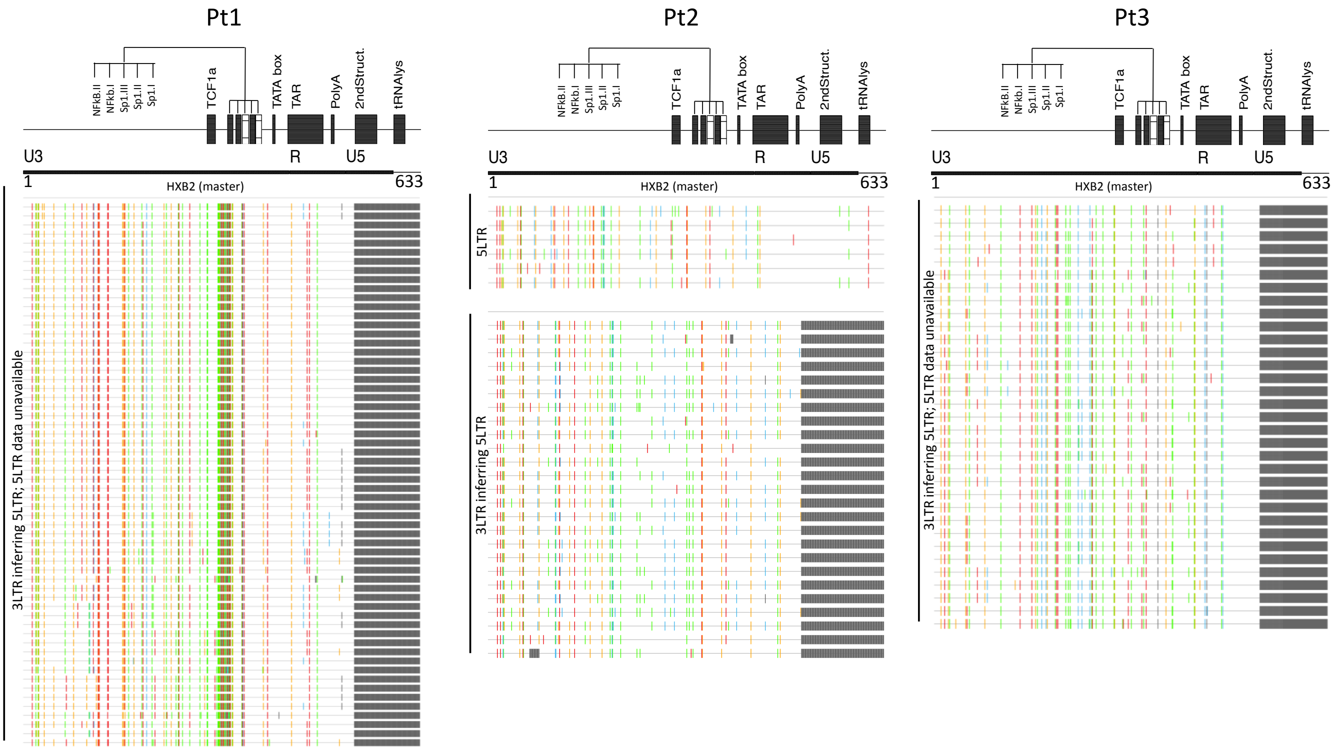
*Slashes indicate amplicon coordinate 1-28 which mapped to HXB2 2724<2697 (inversion)

Supplemental Figure 3

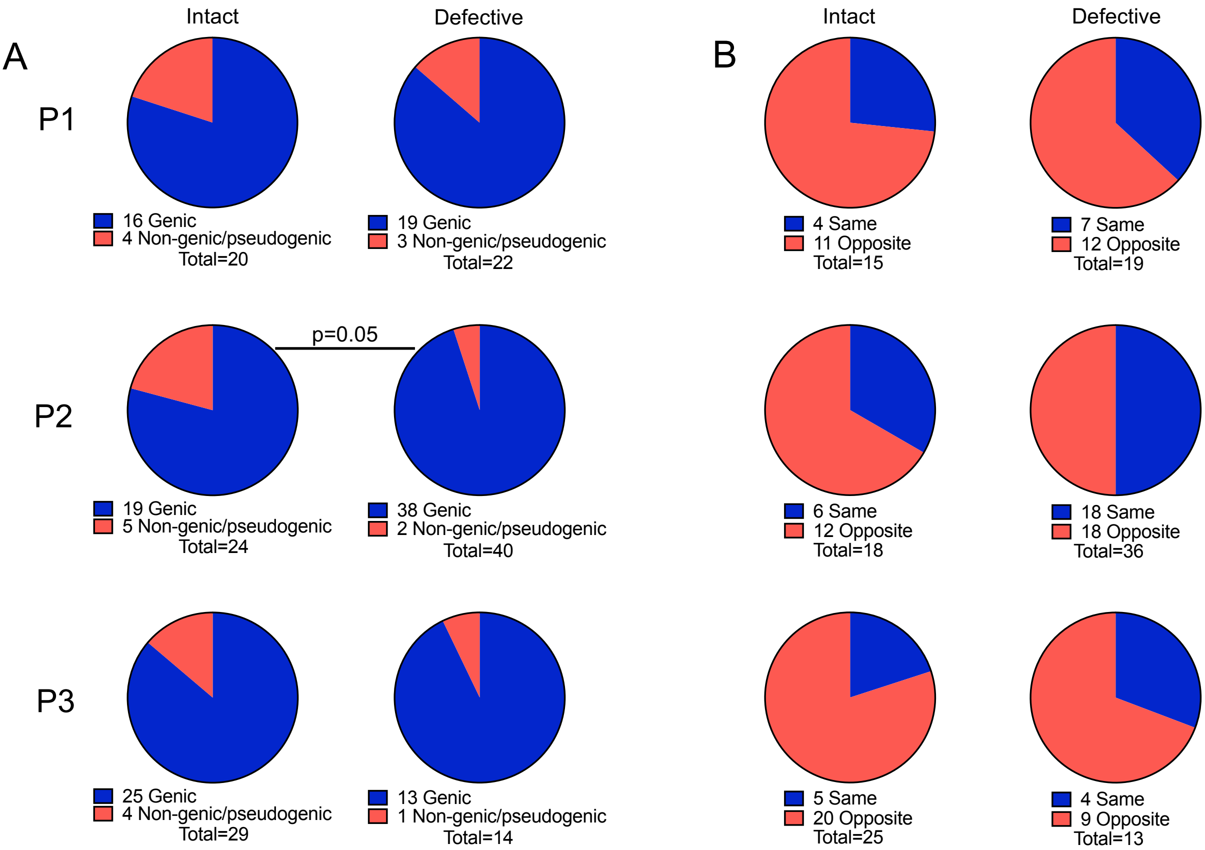
	Provirus	Host-5'LTR Junction	HIV full-genome	3'LTR-Host Junction
*8E5LAV	S3783	chr13-_67485903	……………………………………………………	chr13-_67485907
Pt1	K0834	chr4+_448834	……………………………………………………	chr4+_448830
	K0836	chr4+_448834	……………………………………………………	chr4+_448830
	K0841	chr4+_448834	……………………………………………………	chr4+_448830
	K0848	chr4+_448834	……………………………………………………	chr4+_448830
	K0919	chr4+_36220688	……………………………………………………	chr4+_36220684
	K1207	chr4+_36220688	……………………………………………………	chr4+_36220684
Pt2	S3759	chr11-_62739655	……………………………………………………	chr11-_62739659
	S3786	chr17+_31236424	……………………………………………………	chr17+_31236420
	S2778	chr1+_13704699	……………………………………………………	chr1+_13704695
	S2776	chr1+_13704699	……………………………………………………	chr1+_13704695
	S3769	chr19+_11932626	……………………………………………………	chr19+_11932622
	S4077	chr16+_31715313	……………………………………………………	chr16+_31715309
Pt3	C0647	chr12-_31282521	……………………………………………………	chr12-_31282525
	C0630	chr12-_31282521	……………………………………………………	chr12-_31282525
	C0675	chr12+_133042639	……………………………………………………	chr12+_133042635

*NIH AIDS Reagent Catalog #95, positive control

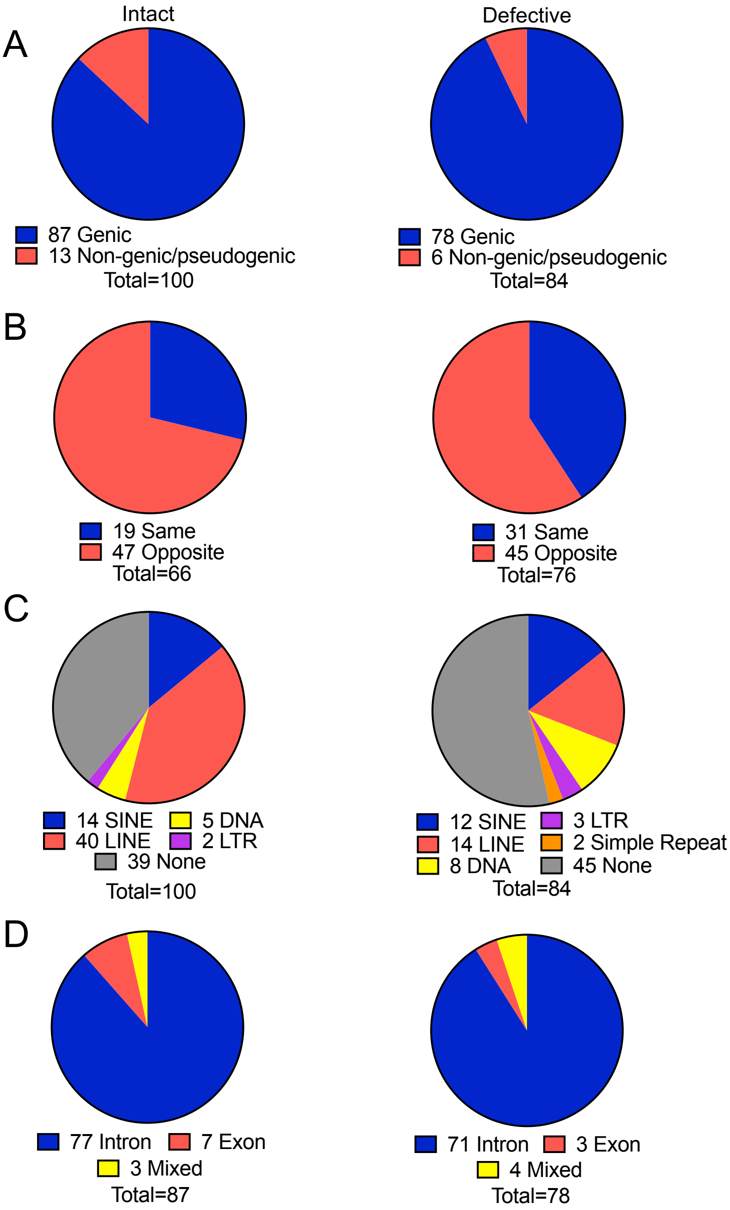
Supplemental Figure 4



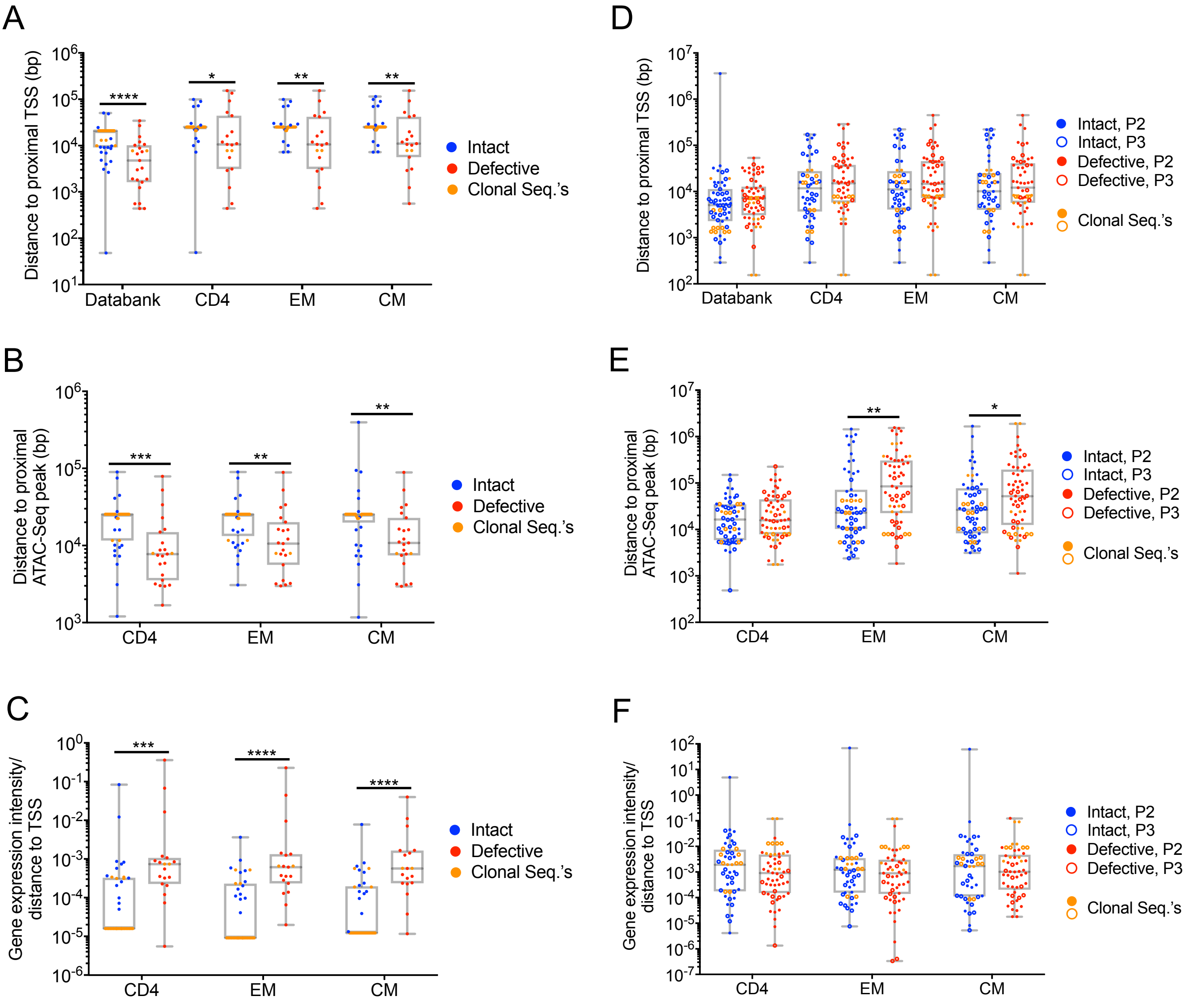
Supplemental Figure 5



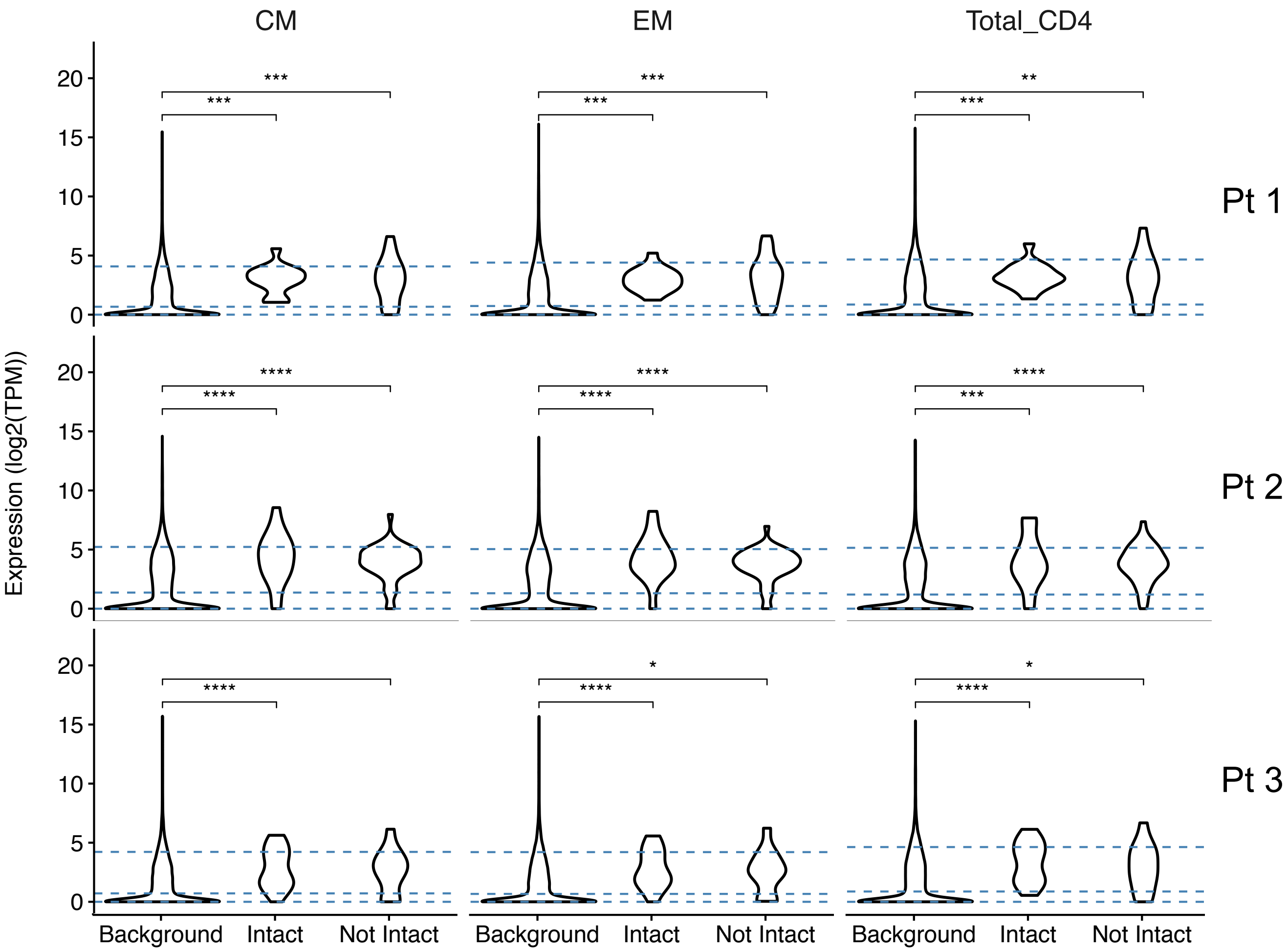
Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 8



Supplemental Table 1: Clinical and demographical characteristics of the study patients.

Patient	Age	Gender	CD4 T cell count (cells/ul)	Viral load (HIV-1 RNA copies/ml)	Time with continuous suppressive HAART at the time of PBMC sampling
P1	59	M	726	<40	11.5 years
P2	55	M	612	<20	13 years
P3	59	M	599	<20	8 years

Supplemental Table 2: Phylogenetic distance between intact proviral sequences analyzed by near full-length sequencing after multiple displacement amplification.

Hamming distances (bp) within hosts			
Within host	**Median**	**Q1**	**Q3**
P1	179	137.5	481.5
P2	329	234	379
P3	177	144.5	211

Hamming distances (bp) across hosts			
Across hosts	**Median**	**Q1**	**Q3**
P1 vs P2	967	941	1018
P1 vs P3	941	924	966
P2 vs P3	873	852	904

Supplemental Table 3A

Patient 1

Sequence No.	Viral Seq. Intactness	Chrom.	Position	Strand	Assays Used	Gene	Orientation	Intron/ Exon	Repeats
1	Intact	chr4	448830	+	ISLA; 5'-SSPCR; 3'-SSPCR	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
2	Intact	chr4	448830	+	ISLA; 5'-SSPCR; 3'-SSPCR	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
3	Intact	chr4	448830	+	ISLA; 5'-SSPCR; 3'-SSPCR	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
4	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
5	Intact	chr4	448830	+	ISLA; 5'-SSPCR; 3'-SSPCR	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
6	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
7	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
8	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
9	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
10	Intact	chr4	448830	+	ISLA; 5'-SSPCR	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
11	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
12	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
13	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
14	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
15	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
16	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
17	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
18	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
19	Intact	chr4	448830	+	ISLA	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
20	Intact	chr4	448830	+	LentiX (DraI); LentiX (SspI)	ZNF721; ABCA11P; RP11-2H3	Multiple	Intron	LINE : L1MB2
21	Intact	chr12	133102657	+	ISLA; 3'-SSPCR	ZNF140	Same	Intron	SINE: AluSx
22	Intact	chr12	133102657	+	ISLA; 5'-SSPCR; 3'-SSPCR	ZNF140	Same	Intron	SINE: AluSx
23	Intact	chr4	36220684	+	ISLA; 5'-SSPCR; 3'-SSPCR	ARAP2	Opposite	Intron	DNA: Tigger3b
24	Intact	chr4	36220684	+	ISLA; 5'-SSPCR; 3'-SSPCR	ARAP2	Opposite	Intron	DNA: Tigger3b
25	Intact	chr11	74056600	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	C2CD3	Opposite	Intron	SINE: MIRb
26	Intact	chr8	38310067	+	ISLA; LentiX (DraI); LentiX (HpaI)	WHSC1L1	Opposite	Intron	LINE: L1MB3
27	Inferred Intact	chr10	60634055	+	ISLA	ANK3	Opposite	Intron	
28	Intact	chr22	25463947	-	ISLA	CRYBB2P1 (Pseudogene)	n/a	n/a	LINE: L1ME3E
29	Intact	chr10	6591049	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	PRKCQ-AS1	Same	Exon	
30	Intact	chr9	127989652	-	ISLA	Non-genic	n/a	n/a	LINE: L2c
31	Intact	chr12	111983879	+	LentiX (DraI); LentiX (SspI)	TMEM116	Opposite	Intron	LINE: L1MB3
32	Inferred Intact	chr7	99631883	-	ISLA; LentiX (DraI); LentiX (SspI)	ZSCAN25	Opposite	Exon	
33	Intact	chr19	35620107	+	ISLA; LentiX (DraI); LentiX (SspI)	HAUS5	Same	Exon	
34	Intact	chr17	42254495	+	ISLA; LentiX (DraI); LentiX (HpaI)	STAT5B	Opposite	Intron	
35	Intact	chr19	39967930	-	ISLA	Non-genic	n/a	n/a	
36	Intact	chr1	198574345	+	ISLA	Non-genic	n/a	n/a	
37	Intact	chr13	45962945	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	ZC3H13	Same	Mixed	
38	Intact	chr12	120059686	-	ISLA	CCDC64/BICDL1 (synonyms)	Opposite	Intron	SINE: AluSx
39	Inferred Intact	chr21	44042503	-	LentiX (DraI)	TRAPPC10	Opposite	Intron	LINE: L1MEc
40	Intact	chr1	222702451	+	ISLA	AIDA	Opposite	Intron	
41	Intact	chr7	12238991	-	ISLA	TMEM106B	Opposite	Exon	DNA: Tigger1
42	Hypermut	chr4	44718733	+	ISLA	GNPDA2	Opposite	Intron	
43	Hypermut	chr4	44718733	+	ISLA	GNPDA2	Opposite	Intron	
44	Hypermut	chr11	73545872	+	ISLA	FAM168A	Opposite	Intron	SINE: MIRb
45	Hypermut	chr8	100934615	+	ISLA	YWHAZ	Opposite	Intron	SINE: AluSx1

46	Hypermut	chr1	235192458	-	ISLA	ARID4B	Same	Intron	LINE: L1ME3D
47	Hypermut	chr12	763265	-	ISLA	WNK1	Opposite	Intron	DNA: MER44B
48	Hypermut	chr17	65053724	+	ISLA	GNA13	Opposite	Exon	
49	Hypermut	chr20	5584799	+	ISLA	GPCPD1	Opposite	Mixed	
50	Hypermut	chr10	119219504	+	ISLA	GRK5	Same	Intron	DNA: Tigger4a
51	Hypermut	chr12	93501308	+	ISLA	MRPL42	Same	Mixed	
52	Hypermut	chr17	27482594	-	ISLA	KSR1	Opposite	Intron	
53	Hypermut	chr12	69250618	+	ISLA	CPSF6	Same	Intron	Simple_repeat: (AAAGA)n
54	Hypermut	chr17	29717125	+	ISLA	SSH2	Opposite	Intron	DNA: Charlie8
55	Hypermut	chr12	132107029	+	ISLA	EP400NL (Pseudogene)	n/a	n/a	
56	Large Deletion	chr1	178539487	+	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	TEX35	Same	Intron	
57	Large Deletion	chr11	65670471	-	ISLA	Non-genic	n/a	n/a	LINE: L1M4
58	Large Deletion	chr21	33203033	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	Non-genic	n/a	n/a	
59	Large Deletion	chr22	40330910	+	ISLA	TNRC6B	Same	Exon	
60	Large Deletion	chr15	65699881	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	DENND4A	Opposite	Intron	SINE: AluJb
61	Large Deletion	chr17	42256522	-	ISLA	STAT5B	Same	Intron	DNA: Tigger7
62	Large Deletion	chr1	32833702	-	ISLA	S100PBP	Opposite	Intron	LINE: L3
63	Large Deletion	chr12	48954552	+	ISLA	ARF3; RP11-302B13.5	Opposite	Intron	SINE: MIR
64	Large Deletion	chr14	34709052	+	ISLA; LentiX (DraI)	CFL2	Opposite	Exon	Simple_repeat: (TA)n

Supplemental Table 3B

Patient 2

Sequence No.	Viral Seq. Intactness	Chrom.	Position	Strand	Assays Used	Gene	Orientation	Intron/ Exon	Repeats
1	Intact	chr14	100100459	-	ISLA; nrLAM	EVL	Opposite	Intron	
2	Intact	chr14	90927174	-	nrLAM	RPS6KA5	Same	Intron	LINE: L1MB2
3	Intact	chr1	31985445	-	ISLA; nrLAM	Non-genic	n/a	n/a	LINE: L2b
4	Intact	chr1	40380189	+	ISLA	SMAP2	Same	Intron	
5	Intact	chr2	43377441	-	ISLA; nrLAM	THADA	Same	Intron	
6	Intact	chr10	95772031	-	nrLAM	ENTPD1; ENTPD1-AS1	Multiple	Intron	DNA: Tigger1
7	Intact	chr6	647555	+	ISLA	EXOC2	Opposite	Intron	
8	Intact	chr17	78069615	-	ISLA	TNRC6C	Opposite	Intron	LINE: L1ME4b
9	Intact	chr11	65505744	-	ISLA	MALAT1	Opposite	Mixed	
10	Intact	chr7	844469	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	SUN1	Opposite	Intron	
11	Intact	chr22	29893354	+	LentiX (DraI)	MTMR3	Same	Intron	
12	Intact	chr6	34680792	+	LentiX (HpaI)	C6orf106	Opposite	Intron	
13	Intact	chr1	155418485	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	ASH1L	Same	Intron	LINE: L1MEc
14	Intact	chr11	62739659	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	TTC9C	Opposite	Intron	SINE: AluSz6
15	Intact	chr1	224803042	-	LentiX (DraI)	Non-genic	n/a	n/a	DNA: MER112
16	Intact	chr2	157366873	+	LentiX (SspI)	Non-genic	n/a	n/a	
17	Intact	chr9	130014126	+	ISLA; nrLAM	FNBP1	Opposite	Intron	LINE: L1ME3D
18	Intact	chr9	130014126	+	ISLA; nrLAM	FNBP1	Opposite	Intron	LINE: L1ME3D
19	Intact	chr11	118027344	+	ISLA; LentiX (DraI); LentiX (HpaI)	TMPRSS4-AS1	Opposite	Intron	
20	Intact	chr11	72973078	+	LentiX (DraI); LentiX (SspI)	FCHSD2	Opposite	Intron	LINE: L1ME3D
21	Intact	chr1	125183733	-	ISLA	Non-genic	n/a	n/a	
22	Intact	chr17	42343193	+	LentiX (SspI); LentiX (HpaI)	STAT3	Opposite	Intron	
23	Intact	chr6	30510759	-	LentiX (SspI); LentiX (HpaI)	Non-genic	n/a	n/a	LINE: L1MC4a
24	Intact	chr4	24536472	-	LentiX (DraI); LentiX (SspI)	DHX15	Same	Intron	
25	Intact	chr22	38498388	+	nrLAM; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	DDX17	Opposite	Intron	
26	Large Deletion with Inversion	chr20	32399018	+	LentiX (DraI); LentiX (HpaI)	ASXL1	Same	Intron	SINE: AluSq2
27	Large Deletion	chr6	90063564	-	LentiX (DraI); LentiX (SspI)	BACH2	Same	Intron	
28	Large Deletion	chr16	1596037	-	LentiX (DraI); LentiX (SspI)	LA16c-395F10.2; IFT140	Multiple	Intron	
29	Large Deletion	chr12	1477652	-	LentiX (DraI); LentiX (HpaI)	ERC1	Opposite	Intron	LINE: L1ME2z
30	Large Deletion	chr17	80794211	-	LentiX (DraI)	RPTOR	Opposite	Intron	LINE: L1ME1
31	Large Deletion	chr16	3772413	+	nrLAM	CREBBP	Opposite	Intron	DNA: MER5B
32	Large Deletion	chr16	3772413	+	nrLAM	CREBBP	Opposite	Intron	DNA: MER5B
33	Large Deletion	chr17	5319738	-	LentiX (DraI)	RABEP1	Opposite	Intron	LINE: L1MA4
34	Large Deletion	chr3	142533494	-	LentiX (DraI)	ATR	Same	Intron	
35	Large Deletion	chr17	14052865	+	LentiX (HpaI)	COX10-AS1	Opposite	Intron	
36	Large Deletion	chr2	27397467	+	LentiX (DraI)	PPM1G	Opposite	Intron	LINE: L1MB7
37	Large Deletion	chr11	65284473	-	LentiX (DraI)	POLA2	Opposite	Intron	SINE: AluJo
38	Large Deletion	chr9	121196173	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	RAB14	Same	Intron	
39	Large Deletion	chr9	121196173	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	RAB14	Same	Intron	
40	Large Deletion	chr21	46466153	-	nrLAM	DIP2A; DIP2A-IT1	Opposite	Intron	DNA: Charlie1a
41	Large Deletion	chr19	4659367	+	nrLAM	MYDGF	Opposite	Intron	SINE: AluSz
42	Large Deletion with Hypermut	chr22	31288715	-	nrLAM	PIK3IP1	Same	Intron	
43	Large Deletion with Hypermut	chr17	59858622	-	ISLA; LentiX (DraI)	Non-genic	n/a	n/a	
44	Large Deletion	chr3	51579313	+	LentiX (DraI); LentiX (SspI)	RAD54L2	Same	Intron	LTR: LTR1D1

45	Large Deletion with Hypermut	chr17	31236420	+	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	NF1	Same	Intron	
46	Hypermut	chr10	12113655	-	ISLA; nrLAM; LentiX (DraI); LentiX (SspI)	DHTKD1	Opposite	Intron	SINE: AluJo
47	Hypermut	chr11	72862543	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	FCHSD2	Same	Intron	LINE: L1MEc
48	Hypermut	chr19	21036090	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	ZNF430	Opposite	Mixed	
49	Hypermut	chr19	41556748	+	LentiX (DraI); LentiX (SspI)	CEACAM21	Same	Intron	LTR: LTR49-int
50	Hypermut	chr6	36531291	+	LentiX (DraI); LentiX (HpaI)	STK38	Opposite	Intron	
51	Hypermut	chr1	35384809	+	LentiX (DraI); LentiX (SspI)	ZMYM4	Same	Intron	DNA: MER112
52	Hypermut	chr5	132400412	-	LentiX (DraI); LentiX (SspI)	Non-genic	n/a	n/a	LINE: L1PA6
53	Hypermut	chr1	183483346	-	LentiX (SspI)	SMG7	Opposite	Intron	
54	Hypermut	chr1	13704695	+	nrLAM; LentiX (DraI); LentiX (SspI)	PRDM2	Same	Intron	
55	Hypermut	chr1	13704695	+	ISLA; nrLAM; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	PRDM2	Same	Intron	
56	Hypermut	chr11	66182302	-	LentiX (DraI)	PACS1	Opposite	Intron	
57	Hypermut	chr12	48671753	+	nrLAM	KANSL2	Opposite	Intron	
58	Hypermut	chr12	48671753	+	nrLAM	KANSL2	Opposite	Intron	
59	Hypermut	chr10	129499970	-	LentiX (HpaI)	MGMT	Opposite	Intron	
60	Hypermut	chr19	7400123	-	LentiX (DraI)	CTD-2207O23.3; ARHGEF18	Opposite	Intron	
61	Hypermut	chr2	7955911	-	LentiX (DraI); LentiX (SspI)	LINC00298	Same	Intron	
62	Hypermut	chr19	11932622	+	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	ZNF700; CTD-2006C1.13	Same	Intron	
63	Hypermut	chr17	75360595	+	LentiX (DraI)	RP11-16C1.3; GRB2	Multiple	Intron	
64	Hypermut	chr3	132690458	+	nrLAM	NPHP3-ACAD11; NPHP3	Opposite	Intron	
65	Hypermut	chr10	14908637	-	nrLAM	DCLRE1C	Same	Mixed	
66	Hypermut	chr9	109085865	-	LentiX (DraI)	TMEM245	Same	Intron	
67	Hypermut	chr16	31715309	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	ZNF720	Same	Intron	LTR: MER92B
68	Hypermut	chr6	42590685	+	LentiX (DraI); LentiX (SspI)	UBR2	Same	Intron	LINE: L1M5
69	Hypermut	chr12	101333996	+	nrLAM	UTP20	Same	Intron	

Supplemental Table 3C

Patient 3

Sequence No.	Viral Seq. Intactness	Chrom.	Position	Strand	Assays Used	Gene	Orientation	Intron/ Exon	Repeats
1	Inferred Intact	chrX	153832959	+	ISLA; LentiX (DraI)	Non-genic	n/a	n/a	LTR: ERVL-E-int
2	Intact	chr5	131542399	+	ISLA	RAPGEF6; CTC-432M15.3	Opposite	Intron	SINE: MIR
3	Intact	chr14	61705724	+	ISLA	HIF1A	Same	Intron	
4	Intact	chr6	26118865	-	ISLA	HIST1H2BC	Same	Intron	
5	Intact	chr17	6768093	-	ISLA	XAF1	Opposite	Intron	LINE: L1MEd
6	Intact	chr7	129856135	+	ISLA	UBE2H	Opposite	Intron	
7	Intact	chr1	246287344	+	ISLA	SMYD3	Opposite	Intron	
8	Intact	chr19	58191884	-	ISLA; LentiX (HpaI)	ZNF274	Opposite	Intron	LINE: L2a
9	Intact	chr13	49447866	-	ISLA; LentiX (HpaI)	SETDB2	Opposite	Intron	LINE: L1ME3C
10	Intact	chr13	46723548	-	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	LRCH1	Opposite	Intron	SINE: AluJo
11	Intact	chr4	153410620	-	ISLA	MND1	Opposite	Intron	
12	Intact	chr19	49517600	-	ISLA	FCGRT	Opposite	Intron	LINE: L1MC5
13	Intact	chr10	100216327	+	ISLA	CHUK	Opposite	Intron	LINE: MamRTE1
14	Intact	chr16	19548512	-	ISLA	CCP110	Opposite	Mixed	
15	Intact	chr16	29795961	+	ISLA	KIF22	Same	Intron	
16	Intact	chr6	25133079	+	ISLA	CMAHP (Pseudogene)	n/a	n/a	LTR: MER4-int
17	Intact	chr2	96344676	-	ISLA	NCAPH	Opposite	Intron	SINE: MIR
18	Intact	chr16	28158039	+	LentiX (DraI)	XPO6	Opposite	Intron	LINE: L1ME5
19	Intact	chr20	49835890	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	SLC9A8	Opposite	Intron	SINE: AluSx1
20	Intact	chr1	155109212	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	Non-genic	n/a	n/a	
21	Intact	chr12	31282525	-	LentiX (DraI); LentiX (SspI)	FAM60A	Same	Exon	SINE: AluSq
22	Intact	chr12	31282525	-	ISLA; LentiX (DraI); LentiX (SspI); 5'-SSPCR; 3'-SSPCR	FAM60A	Same	Exon	SINE: AluSq
23	Intact	chr12	31282525	-	ISLA; LentiX (SspI); 5'-SSPCR; 3'-SSPCR	FAM60A	Same	Exon	SINE: AluSq
24	Intact	chr13	40977971	+	ISLA; LentiX (SspI)	ELF1	Opposite	Intron	LINE: L1M5
25	Intact	chr17	59781492	-	ISLA	VMP1	Opposite	Intron	
26	Intact	chr3	9414301	-	ISLA	SETD5	Opposite	Intron	
27	Intact	chr1	153900278	+	ISLA; LentiX (DraI); 3'-SSPCR	GATAD2B	Opposite	Intron	SINE: AluSg
28	Intact	chr1	153900278	+	ISLA; 3'-SSPCR	GATAD2B	Opposite	Intron	SINE: AluSg
29	Intact	chr16	29568715	+	ISLA	SMG1P2 (Pseudogene)	n/a	n/a	
30	Intact	chr10	70822931	-	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	SGPL1	Opposite	Intron	
31	Intact	chr19	48245646	+	LentiX (SspI)	CARD8	Opposite	Intron	
32	Intact	chr19	48245646	+	LentiX (SspI)	CARD8	Opposite	Intron	
33	Intact	chr12	133042635	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI); 5'-SSPCR; 3'-SSPCR	ZNF84	Same	Intron	
34	Intact	chr12	133042635	+	ISLA; LentiX (DraI); LentiX (SspI); LentiX (HpaI)	ZNF84	Same	Intron	
35	Hypermut	chr12	122588565	-	ISLA; LentiX (DraI); LentiX (HpaI)	KNTC1	Opposite	Intron	LINE: L1ME2
36	Large Deletion	chr9	121195377	+	ISLA	RAB14	Opposite	Intron	
37	Large Deletion	chr19	8452805	-	LentiX (DraI); LentiX (SspI)	HNRNPM	Opposite	Intron	
38	Large Deletion	chr19	8452805	-	LentiX (DraI); LentiX (SspI)	HNRNPM	Opposite	Intron	
39	Large Deletion	chr19	8452805	-	LentiX (DraI); LentiX (SspI)	HNRNPM	Opposite	Intron	
40	Large Deletion	chr19	8452805	-	LentiX (DraI); LentiX (SspI)	HNRNPM	Opposite	Intron	
41	Large Deletion	chr12	24825366	-	LentiX (DraI); LentiX (SspI)	BCAT1	Same	Intron	LINE: L1MA4
42	Large Deletion	chr1	31013049	+	LentiX (DraI); LentiX (SspI)	PUM1	Opposite	Intron	
43	Large Deletion with Hypermut	chr2	43141299	+	LentiX (DraI); LentiX (SspI)	Non-genic	n/a	n/a	SINE: AluJb
44	Large Deletion with Hypermut	chr17	44492696	+	LentiX (DraI)	GPATCH8	Opposite	Intron	SINE: MIRb
45	Large Deletion	chr17	65019576	+	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	GNA13	Opposite	Intron	LINE: L1MA2
46	Large Deletion	chr17	81659786	-	LentiX (DraI)	PDE6G	Same	Intron	
47	Large Deletion	chr9	120960997	+	LentiX (DraI)	C5	Opposite	Intron	
48	Large Deletion	chr8	131948310	-	LentiX (DraI); LentiX (SspI); LentiX (HpaI)	EFR3A	Opposite	Intron	SINE: MIR
49	Large Deletion	chr1	111155126	+	ISLA; LentiX (DraI)	CEPT1	Same	Intron	LINE: L1MA5
50	Large Deletion	chr5	177239690	-	ISLA	NSD1	Opposite	Intron	

51	Internal Inversion	chr8	128428551	-	LentiX (DraI); LentiX (SspI)		LINC00824	Same	Intron	SINE: MIRb