Supplementary Materials for

Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection

Thor A. Wagner, Sherry McLaughlin, Kavita Garg, Charles Y. K. Cheung, Brendan B. Larsen, Sheila Styrchak, Hannah C. Huang, Paul T. Edlefsen, James I. Mullins, Lisa M. Frenkel*

*Corresponding author. E-mail: lfrenkel@uw.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S3
Tables S1 and S2
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/science.1256304/DC1)

Table S3 (as Excel file)

MATERIALS and METHODS

Study population. HIV-1 infected individuals were enrolled prior to initiation of ART, and had study blood collected at regular clinic visits, in accordance with a protocol approved by the Seattle Children's Institutional Review Board for Human Subjects. Clinical parameters, including plasma HIV RNA, CD4 cell counts and prescribed antiretroviral drugs were extracted from participants' medical records. Children with median plasma HIV-1 RNA levels of <50 copies/mL for >10 years of ART with available banked specimens were selected for this study (*2, 8, 44*).

HIV integration site loop amplification (ISLA) assay. The ISLA assay is outlined in Figure S1. To develop and optimize the method, DNA extracted from ACH-2 cells, a T cell clone with one common integrated proviral copy (*46, 47*), was mixed with DNA from Molt-4, a HIV-uninfected lymphoblast clone (*48, 49*) at diminishing concentrations of ACH-2 DNA.

Extraction and quantification of HIV DNA. Nucleic acids were extracted from cell lines and HIV-infected participants' peripheral blood mononuclear cells (PBMC) using the Gentra DNA purification system (Qiagen, Valencia, CA) or the ArchivePure DNA purification system (5 Prime, Gaithersburg, MD). HIV DNA was quantified by nested PCR of *env* (C2V5 region) using serially diluted nucleic acids (three replicates of each 5-fold dilution) and the QUALITY software program (*50*); http://indra.mullins.microbiol.washington.edu/quality/.

Linear extension reactions. For unidirectional amplification of the proviral DNA from HIV *env* through the site of chromosomal integration two forward primers were used, one in HIV *env* and one in HIV *nef* (Figure S1, Step 1). These primers were combined with ~3 copies of proviral DNA, as quantitated above (3 copies were typically required to achieve 30% success in integration site amplification reactions). Multiple linear reactions were conducted with each specimen (~100 reactions to achieve 30 positive reactions). The DNA was combined with 15 pmol each of primers env.6880F (5'-CCCCGGCTGGTTTTGCGATTCTAAAGTGTA-3') and

nef.8948F (5'-CCAATGCTGATTGTGCCTGGCTAGAAGCA-3'), and 1x Advantage2 buffer (Clontech, Palo Alto, CA), 200 uM dNTPs, 5 U NotI (New England Biolabs, Ipswich, MA), and 0.5uL of Advantage2 50x polymerase in 50 uL total reaction volumes. The reactions were incubated as follows: 30 minutes (min) at 37ºC, 2 min at 95ºC, followed by 30 cycles of 20 seconds (sec) at 94ºC, 30 sec at 60ºC and 3 min at 72ºC.

<u>Bind random decamers tailed with a U5-specific primer sequence</u>. To create a reverse primer site for loop formation of the proviral integration site and adjacent human sequence, random decamers tailed with a U5-specific primer (deca1.U5; 5'-TCAAGTAGTGTGTGCCCGTCTGTNNNNNNNNNN-3') were hybridized to an aliquot of the single-stranded linear extension reaction generated in Step 1 (Figure S1, Step 2). This primer was extended by combining 13 uL of the linear extension reaction with 40 pmol of deca1.U5 and 2.5 U of MyTaq in 0.5x MyTaq buffer (Bioline, London, UK) in a total volume of 20 uL. Concurrently, to stabilize the unused single-stranded template during storage, the remaining ~37uL of each linear-extension reaction was converted to double-stranded DNA by adding 15 pmol of deca1.U5 primer and 2.5 U of MyTaq in a total volume of 50 uL.

Both sets of reactions were incubated at 68ºC for 2 min, cooled to 65ºC for 1 min, and then progressively cooled by 1ºC per minute until reaching 25ºC. The reactions were reheated to 60ºC and ramped down by 1ºC per minute, as before, until reaching 20ºC. Following the two ramped cooling steps, the 50uL reactions were stored at -20ºC.

<u>Trim ends of duplex fragments</u>. Using the 20uL reactions from Step 2, the single-stranded DNA 3' to the random decamer annealing site was removed by addition of 10 U of Exonuclease 1 (Exo1) (New England Biolabs, Ipswich, MA), followed by incubation at 37ºC for 45 min (Figure S1, Step 3a). To ensure complete digestion by Exo1, the temperature was then increased by 1ºC, at intervals of 3 minutes, until reaching 43ºC to melt any template partially annealed to the

U5 region of the primer. The reactions were then incubated at 72ºC for 15 min to allow extension by MyTaq of the region complementary to the U5 primer sequence (Figure S1, Step 3b), followed by 5 min at 95ºC to inactivate Exo1.

Form and extend loops and 1st-round PCR amplification. The fragments from Step 3 were subsequently cycled through denaturing/annealing/extension steps in the presence of MyTaq and the 1st-round PCR primer (RF2). Because the 3'-end of the fragment was engineered to bind to its complementary site in U5, hybridization will result in the formation of a loop (Figure S1, Step 4a). MyTaq will extend from the U5 priming site, through the end of the fragment, creating a loop of human genomic sequence with the HIV LTR duplicated on each end (Figure S1, Step 4b). Once the loop is formed and the LTR is duplicated, the RF2 primer can bind and begin amplification (Figure S1, Step 5).

These looping/1st-round PCR reactions were composed of 20uL of the product from Step 3, 1x MyTaq buffer, 30 pmol RF2 primer (5'-TCTGAGCCTGGGAGCTCTCTG-3') and 3.75 U of MyTaq in a total volume of 50 uL. Reactions were incubated for 2 min at 95ºC; followed by 10 cycles of 20 sec at 94ºC, 30 sec at 60ºC, and 2 min at 72ºC; then followed by 40 cycles of 10 sec at 92ºC, 15 sec at 65ºC, and 2 min at 72ºC; and a final 5 min incubation at 72º. Additional second and third round PCR reactions with nested LTR primers were performed to ensure a product that was specific, and could be visualized on a gel. Second round PCR mixtures contained 5 uL of the 1st round reaction, 30 pmol of RF1 (5'-TTAAGCCTCAATAAAGCTTGCCTTG-3'), 1x MyTaq buffer and 3.75 U MyTaq in 50 uL. Third round PCR mixtures contained 2 uL of the 2nd round reaction, 30 pmol of 1.U5 (5'-TCAAGTAGTGTGTGCCCGTCTGT-3'), 1x MyTaq buffer and 3.75 U MyTaq in 50 uL. Cycling parameters were the same for both: 2 min at 95ºC, followed by 34 cycles of 20 sec at 94º, 30 sec at 65º and 2 min at 72º, and a final 5 min incubation at 72ºC.

Identify chromosomal integration site. Third round-PCR products from Step 5 were identified by gel electrophoresis and sequenced using 2.U5 (5'-GTTGTGTGACTCTGGTAACTAGAGAT-3'), an U5-specific primer located downstream of the looping site. Sequences were edited using Sequencher v5.0.1 and the last 40 bases of the HIV 3' LTR sequence was used to pinpoint the site of provirus integration into the human genome (Figure S1, Step 6).

Design primers to integration sites. Nested primers were designed in the human sequence 3' to each integration site. T_ms ranged from 62° - 68°C (based on Integrated DNA Technologies' OligoAnalyzer; https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/), and when possible, the ΔG for the 3' end was less negative than -7.5 kcal/mol. If the integration happened to occur near or within a repetitive element, the primer was designed so that it did not overlap the element (Figure S1, Step 7). All primer sequences were analyzed using Primer-BLAST (51) to identify whether primers would likely amplify unintended regions of the human genome.(51) If necessary, the primers were revised to specifically amplify the intended target. Because each env amplification used a unique primer, there were no positive control DNA templates to optimize the PCR primers or conditions.

Amplification and sequencing of HIV env linked to integration sites. Forty-five uL of the saved linear extension reaction from Step 2 was combined with 90 pmol of env.6984F (5'-ACAGTACAATGTACACATGGAATTA-3'), 90 pmol of the outer integration site-specific primer, 200 nM dNTPs, 1x Advantage2 buffer and 0.5 uL Advantage2 50x polymerase in a total volume of 300 uL. This was split into three 100 uL aliquots and cycled as follows: 2 min at 95°, followed by 45 cycles of 10 sec at 94°C, 15 sec at 65°C, and 3 min at 68°C, and ending with 5 min at 72°C. The 2nd-round PCR contained 5 uL of 1st-round product, 15 pmol of env.6963F (5'-TGGAATTAGGCCAGTAGTATCAACTCA-3'), 15 pmol of the inner integration site-specific primer, 200 nM dNTPs, 1x Advantage2 buffer and 0.5 uL Advantage2 in a total volume of 50 uL (Figure S1, Step 8). Cycling parameters were the same as for 1st-round PCR, except the

number of cycles was decreased to 34. Positive reactions were detected by gel electrophoresis and the C2V5 region of HIV *env* was bi-directionally sequenced with the primers DR7 (*52*) and BH2 (*53*).

Laboratory procedures to reduce the risk of amplicon carryover contamination. We used a separate area within the laboratory to set-up 1st-round PCR without exposure to potentially contaminating PCR amplicons. Given that our method includes three rounds of PCR, a second area was designated solely to set-up 2nd-round PCR. The benchtop was covered with a disposable absorptive pad that was replaced frequently. Pipet tip boxes were assigned for exclusive use in this area and pipettors were wiped with soap and water prior to use.

Alignment of human sequences with integrated provirus. Bowtie 2 (*54*) was used to align all trimmed sequences against *human genome* version 19 (hg19). The top two alignments for each read that aligned at more than one location and whose alignment lengths were within one percent of each other were also aligned with BLAT (*55*). These alignments were mapped to known genes based on annotations from RefSeq genes (*56*) and UCSC known genes (*57*). For GO analysis and analysis involving comparison to known cancer genes, we only considered protein coding genes and characterized long non-coding RNAs. All integration sites were eventually mapped to a single chromosomal position.

Gene Ontology (GO) Analyses: Enrichment of GO terms for ontology biological process terms was tested with topGO package (*58*) in R using "classic" and elim" methods and Fisher statistics. To identify GO terms enriched in genes with HIV integration compared to all human genes, we required the GO terms to be significant (p-value < 0.05) by both "elim" and "classic" methods. To define cancer-related GO categories, we used the 43 GO terms defined across 10 hallmarks of cancer (*23, 24*) from a recent study (*25*).

Statistical Methods. For comparison of two-by-two contingency tables, we applied two-sided Fisher's exact tests (*58*) at the standard 5% type-I error rate. For evaluation of GO category enrichment of genes with integration sites in the three participants as compared with all genes in the human genome, we used the method described above. For evaluation of integration site enrichment for gene annotation features, which includes 1) the 43 hallmark cancer-related GO terms, tested individually and collectively as the inclusion of any of the terms, and 2) inclusion of the gene in the set of 1332 known cancer-related genes (*11-15*), we compared the fraction of (unique) integration sites that have the evaluated feature in each participant (pooled over the three time points) and in any participant (pooled over participants and time points) in our data to integration sites in Jurkat cells reported by Wang et al. (*20*). Specifically, the Jurkat cells data was used to construct a null distribution, as estimated by repeatedly uniformly sampling integration sites (with replacement) using a one-sided test at a 2.5% nominal type-I error rate. Individual GO annotations evaluated in this manner were corrected using the Holm-Bonferroni method (*59*) across the evaluated 43 hallmark GO terms. We repeated this analysis using only the subset of integration sites found multiple times (and therefore confirmed to be proliferating), pooled over participants. We also used the unique integration site samples to evaluate the null hypothesis that the number of unique integration sites per gene (among genes having any integration sites) is what would be expected if the observed unique sites followed the distribution found in the Jurkat cell data, and to evaluate the number of genes having integration sites observed in more than one of the three subjects. To evaluate trends over time in the percentage of unique integration sites that fall in genes having a particular feature (such as being a known cancer gene), we applied the Cochran-Armitage test (*45*) at a 5% nominal type-I error rate. All hypothesis tests were conducted using the R statistical computing language except as noted above, and p-values for individual analyses were not corrected for multiple comparisons except as noted above.

To evaluate whether the number of genes found to have insertion sites in more than one of the three subjects is significantly more than what would be expected by chance, given the number of genes found to have any integration sites in each subject, we conducted a one-sided hypothesis test (at nominal type-I error rate alpha = 0.025). We compared the observed number, 12, to a null distribution estimated by repeatedly counting the number of genes with at least two integrations across three sets of genes selected randomly from all genes observed to have any integrations in the Wang et al. dataset, with each gene set of size equal to the total number of genes for which any integration site was found in the corresponding subject. We sampled genes (without replacement) with probabilities proportional to the number of integration sites per gene observed in the Jurkat cell dataset.

This project was conceived and funding obtained by LMF, JIM; the project was designed by TAW, SM, JIM, LMF; specimens were collected by LMF, TAW; data were generated by SM, HCH, SS; data were analyzed by TAW, SM, KG, CYKC, PTE, JIM, LMF; and the manuscript was written by TAW, SM, KG, CYKC, PTE, JIM, LMF.

Supplemental Figures

Figure S1. Overview of HIV integration site by loop amplification (ISLA)

Figure S2. Phylogenetic relationships between HIV-1 *env* (C2V5 region) genes sampled from participant B1 through time**.** Phylograms generated and labeled as described in the legend of Figure 2. In cases where HIV C2V5 *env* sequences were generated from cells that shared a specific integration site, the *env* sequences (~625bp) were identical. In contrast, among proviruses integrated at different chromosomal sites, only 3 clusters of sequences shared identical C2V5 sequences. Two of these clusters were made up of pairs of two sequences each, and when the entire *C2env-nef*-3'LTR region (~2.8kb) was sequenced, these differed by 2 mismatches in one case and 39 mismatches plus 2 indels (insertion/deletion) in the second. The

third cluster consisted of 4 sequences, two identical in the *C2env-nef*-3'LTR region and integrated in the same position in the CCDC64 gene; whereas the other sequences differed from the identical pair by 36 bases plus one indel, and 3 mismatches, respectively.

Figure S3. Phylogenetic relationships between HIV-1 *env* (C2V5 region) genes sampled from participant R1 through time**.** Sequences organized as described in the legend for Figure 2.

Supplemental Tables

Table S1. Panel A. Genes with HIV integrations detected in multiple study participants, and detection of proliferating cells (≥2 cells with identical integration site). **Panel B.** Genes with proliferating cells detected within each participant. **Panel C.** Multiple integration sites within the same gene found in a participant, without demonstrated proliferation**.**

Table S2. Top20 Gene Ontology (GO) terms overrepresented in ART-suppressed study participants compared to those annotated in the human genome.

Table S3. Sequences of HIV-3'LTR and adjacent chromosome integration sites.

Step 1: Linear extension with 30 cycles (50uL)

env primer   *nef* primer   Integration site

HIV

LTR   LTR

chromosome

13uL for Steps 2-7

37uL for Step 8

single-stranded fragments

*nef*   LTR

env   LTR
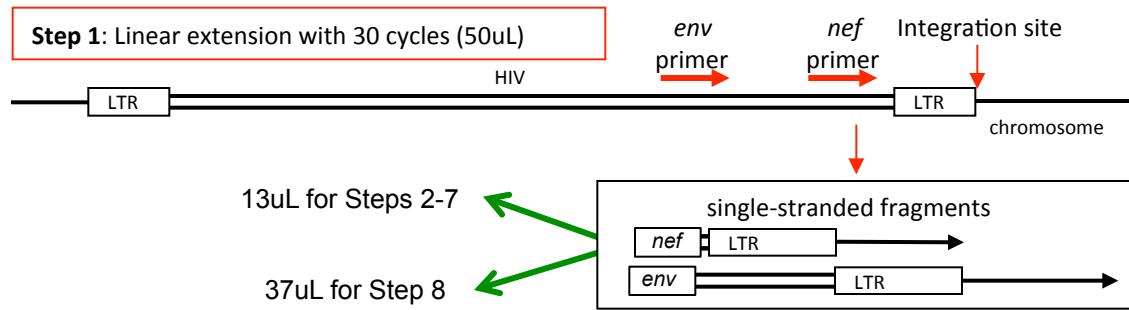
Step 2: Bind random decamers tailed with a U5-specific primer sequence and extend

Integration site   single-stranded fragment

LTR

decamer   U5 priming sequence
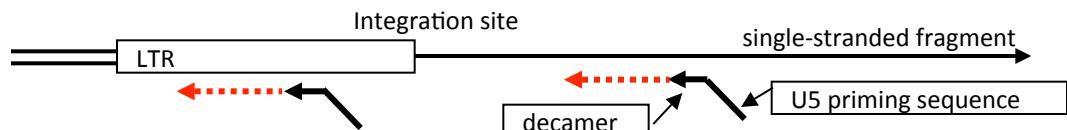
Step 3: (a) Digest single-stranded fragment with Exo I; (b) Fill with Taq

LTR   (a)

(b) Reverse/complement of U5 priming sequence
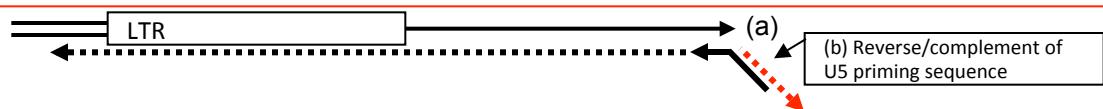
Step 4: Denature, anneal to form loops and extend

LTR

(a) Reverse/complement of U5 priming sequence forms loop

U5 priming site

LTR

(b) Final loop product

5'   U3   R   GCCCGTCTGTTGTGTGACTCTGGTAACTAGAGAT

3'   U3   R   CGGGCAGACAACACACTGAGACCATTGATCTCTA

U5 priming site

human sequence
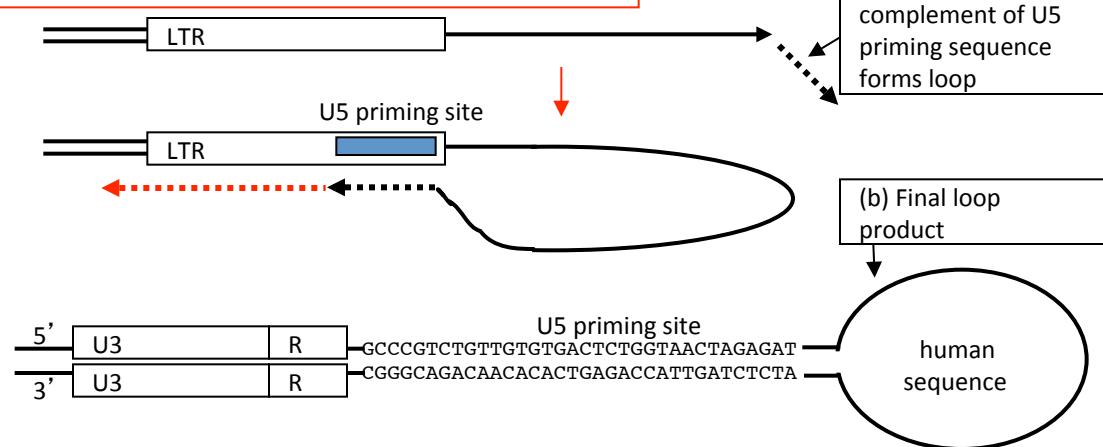
Step 5: Amplify with nested R-region primers

Integration site

5'   3'

RF.2 | RF.1 | 1.U5 | 2.U5   1.U5 | RF.1 | RF.2

Human sequence
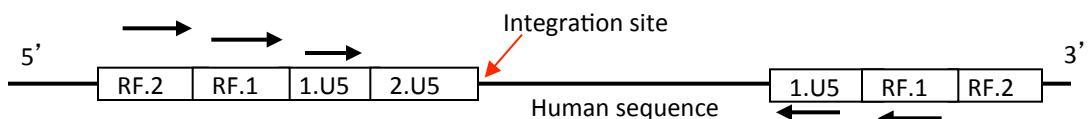
Step 6: Sequence the integration site and identify the chromosome

Step 7: Design nested primers downstream of the integration site

Step 8: Amplify with nested *env* and chromosome primers
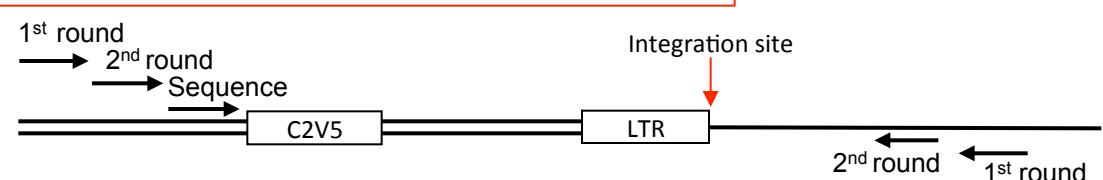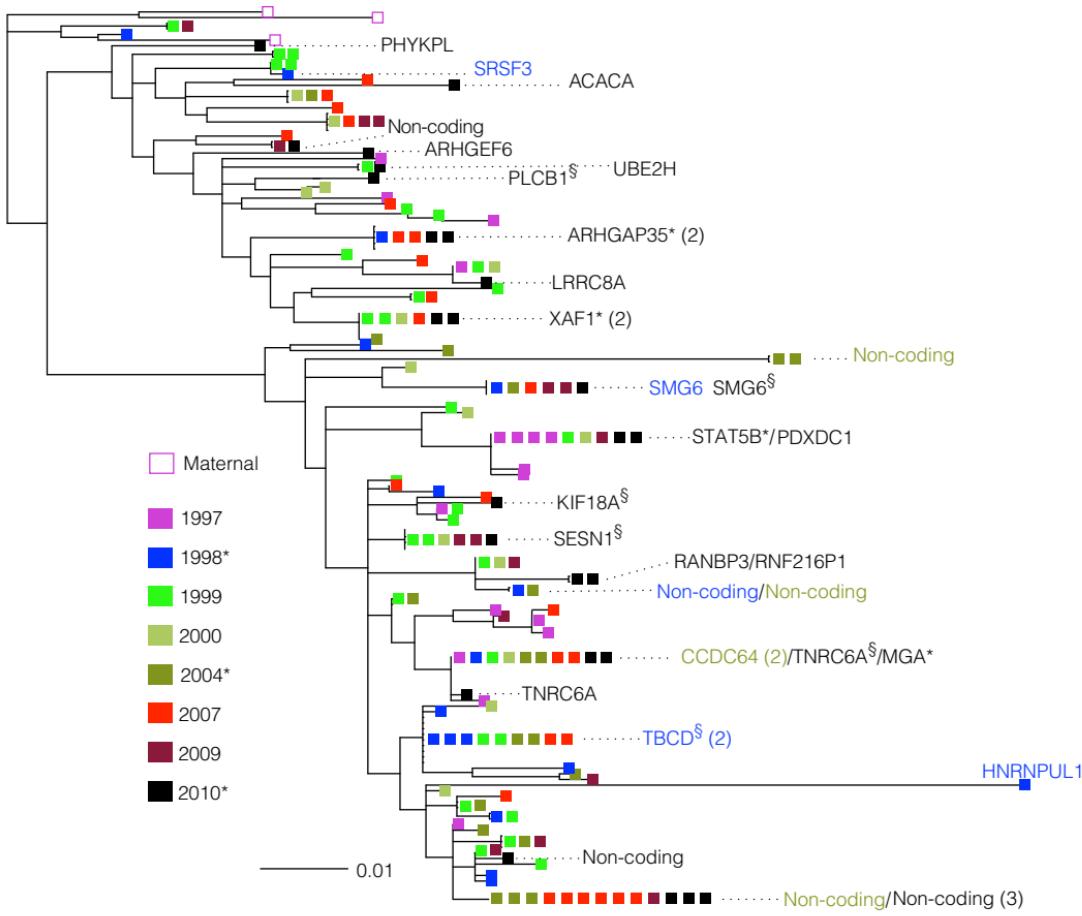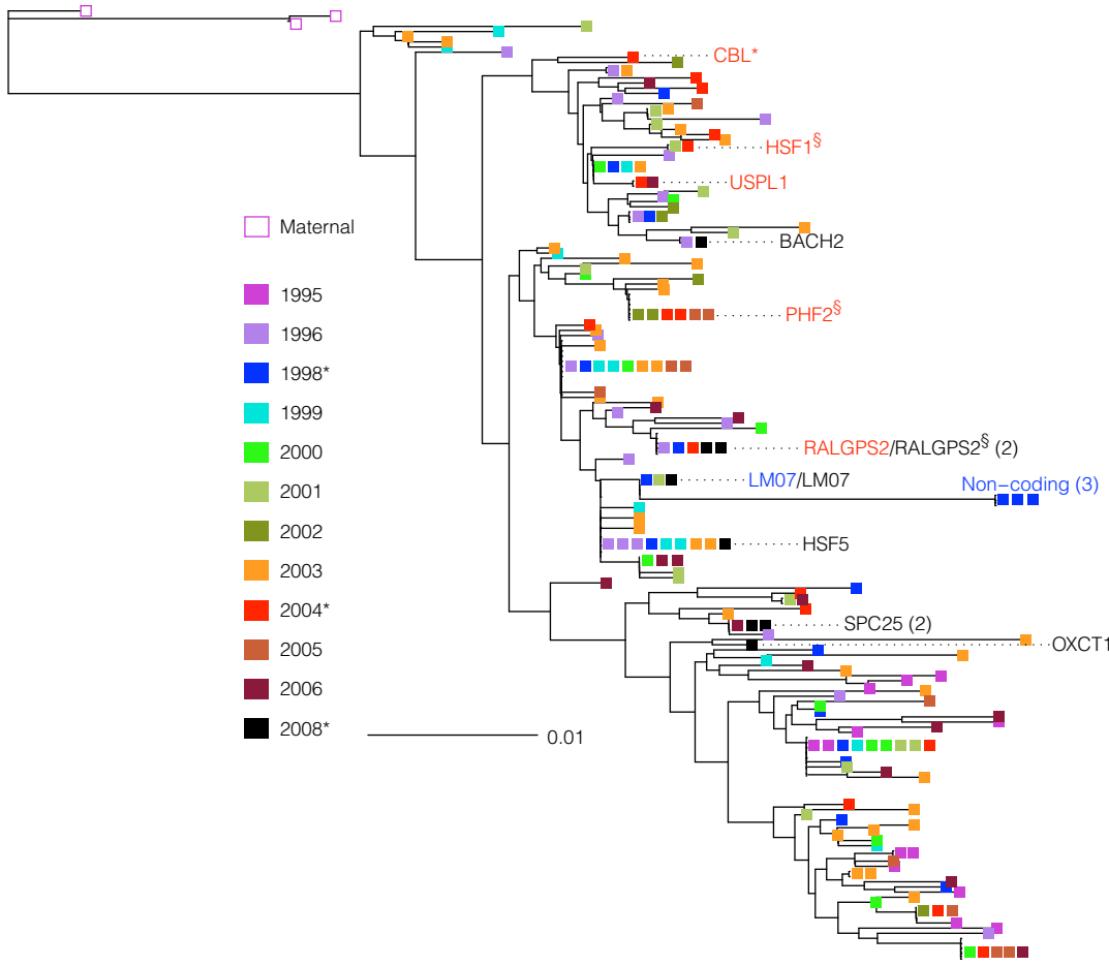
1st round

2nd round

Sequence

Integration site

C2V5   LTR

2nd round   1st round

PHYKPL

SRSF3

ACACA

Non-coding

ARHGEF6

PLCB1§   UBE2H

ARHGAP35* (2)

LRRC8A

XAF1* (2)

Non-coding

SMG6   SMG6§

STAT5B*/PDXDC1

KIF18A§

SESN1§

RANBP3/RNF216P1

Non-coding/Non-coding

CCDC64 (2)/TNRC6A§/MGA*

TNRC6A

TBCD§ (2)

HNRNPUL1

Non-coding

Non-coding/Non-coding (3)

☐ Maternal

■ 1997
■ 1998*
■ 1999
■ 2000
■ 2004*
■ 2007
■ 2009
■ 2010*

|——| 0.01

Maternal

1995
1996
1998*
1999
2000
2001
2002
2003
2004*
2005
2006
2008*

0.01

CBL*
HSF1§
USPL1
BACH2
PHF2§
RALGPS2/RALGPS2§ (2)
LM07/LM07   Non-coding (3)
HSF5
SPC25 (2)   OXCT1

Table S1, Panel A. Genes with HIV integrations detected in multiple study participants, and detection of proliferating cells (≥2 cells with identical integration site)

Gene:	Proliferation in 2 participants	Proliferation in 1 participant		No proliferation detected								
	BACH2	OXCT1	STAT5B*	ANKRD13C	C2CD3	CREBBP*	HNRNPUL1	IKZF3	KCTD13	MAPK1	ST8SIA4	TRAPPC10
Participant-B1												
# different integration sites			1			1	1		1	1		1
# times integration site detected			1			1	1		1	1		1
# same HIV C2V5 detected at other time points												
Total # times detected in gene			1			1	1		1	1		1
Participant-L1												
# different integration sites	7	1	3	1	1			1		1		1
# times integration site detected	4,2,1,1,1,1,1	1	1,1,1	1	1			1		1		1
# same HIV C2V5 detected at other time points												
Total # times detected in gene	11	1	3	1	1	1	1	1	1	1		1
Participant-R1												
# different integration sites	2	2		1	1	1	1	1	1		1	
# times integration site detected	2,2	1,1		1	1	1	1	1	1		1	
# same HIV C2V5 detected at other time points	1											
Total # times detected in gene	5	2		1	1	1	1	1	1		1	

* Indicates gene associated with cancer

Table S1, Panel B.   Genes with proliferating cells detected within each participant

Gene:	ARHGAP35*	CCDC64	MOB1A*	MGA*	PDE3B	PRKCH	SESN	SMG6	SRSF3	TBCD	TNRC6A	UBE2H	WDR26	XAF1*
Participant-B1														
# different integration sites	1	1	1	1	1	1	1	2	1	1	1	1	1	1
# times integration site detected	2,2	4,1	2	1,2	2	2	1	1,9	2,1,1	5,1,1	1,2,2	1,2	3,2	3,3,3
# same HIV C2V5 detected at other time points	3						5	4		7				4
Total # times detected in gene	7	5	2	3	2	2	6	14	4	14	5	3	5	13

Gene:	ACSF3	APOBEC3C	GPC3*	MDC1	MKL2	RNF217
Participant-L1						
# different integration sites	1	1	1	1	3	1
# times integration site detected	3	1,1,2	2	7,16,9	2,1,1	2
# same HIV C2V5 detected at other time points	2	11		25		
Total # times detected in gene	5	15	2	57	4	2

Gene:	CBL*	HSF1	HSF5	LMO7	OXCT1	PHF2	RALGPS2	SPC25	USPL1	ZHX3
Participant-R1										
# different integration sites	1	1	2	1	1	1	1	1	1	1
# times integration site detected	2	2	1,1	1,1	1,1	5	1,4	4	2	3
# same HIV C2V5 detected at other time points		1	8	1		4	2	1	1	
Total # times detected in gene	2	3	10	3	2	9	7	5	3	3

* Indicates gene associated with cancer

Table S1, Panel C.   Multiple integration sites within the same gene found in a participant, without demonstrated proliferation

Participant/Gene:	**B1**	GCN1L1	NSMCE2	**R1**	GCN1L1	NSMCE2
# different integration sites		2	2		2	2
# times integration site detected		1,1	1,1		1,1	1,1
Total # times detected in gene		2	2		2	2

* Indicates gene associated with cancer

Table S2. Top-20 Gene Ontology (GO) terms overrepresented in ART-suppressed study participants compared to those annotated in the human genome

GO Term^	Biological Function	Annotated Genes with given GO Term	Expected number of Genes	Genes with HIV integration	p value (elim)
GO:0000278*	mitotic cell cycle	824	13.38	31	0.00016
GO:0007173†	epidermal growth factor receptor signaling pathway	222	3.61	14	0.00019
GO:0051726†	regulation of cell cycle	725	11.77	25	0.00032
GO:0006491	N-glycan processing	9	0.15	3	0.00033
GO:0045944ø	positive regulation of transcription from RNA polymerase II promoter	764	12.41	25	0.00069
GO:0019048ø	virus-host interaction	364	5.91	15	0.00092
GO:0036092*	phosphatidylinositol-3-phosphate biosynthetic process	14	0.23	3	0.00135
GO:0006511	ubiquitin-dependent protein catabolic process	420	6.82	16	0.00143
GO:0048015*	phosphatidylinositol-mediated signaling	206	3.35	10	0.00202
GO:0032008*	positive regulation of TOR signaling	16	0.26	3	0.00203
GO:0033962	cytoplasmic mRNA processing body assembly	16	0.26	3	0.00203
GO:0009790	embryo development	980	15.92	30	0.00208
GO:0007219*	Notch signaling pathway	141	2.29	8	0.00214
GO:0043603†	cellular amide metabolic process	34	0.55	4	0.00215
GO:0045646	regulation of erythrocyte differentiation	34	0.55	4	0.00215
GO:0048011	neurotrophin TRK receptor signaling pathway	281	4.56	12	0.0022
GO:0022402†	cell cycle process	1074	17.44	37	0.00225
GO:0045070ø	positive regulation of viral genome replication	17	0.28	3	0.00243
GO:0051294	establishment of spindle orientation	17	0.28	3	0.00243
GO:0061003	positive regulation of dendritic spine morphogenesis	5	0.08	2	0.00254

^ The top 20 biologic functions enriched among the genes with HIV integration, ranked according to p-values generated by "elim" method and Fisher statistics

† Statistically significant (p<0.025) after controlling for multiple comparisons

* Includes human cancer associated genes

ø Expected to have a strong influence on viral life cycle

References

1. J. D. Siliciano, J. Kajdas, D. Finzi, T. C. Quinn, K. Chadwick, J. B. Margolick, C. Kovacs, S. J. Gange, R. F. Siliciano, Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med.* **9**, 727–728 (2003). Medline doi:10.1038/nm880

2. N. H. Tobin, G. H. Learn, S. E. Holte, Y. Wang, A. J. Melvin, J. L. McKernan, D. M. Pawluk, K. M. Mohan, P. F. Lewis, J. I. Mullins, L. M. Frenkel, Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: Expression of archival virus and replication of virus. *J. Virol.* **79**, 9625–9634 (2005). Medline doi:10.1128/JVI.79.15.9625-9634.2005

3. A. Sigal, J. T. Kim, A. B. Balazs, E. Dekel, A. Mayo, R. Milo, D. Baltimore, Cell-to-cell spread of HIV permits ongoing replication despite antiretroviral therapy. *Nature* **477**, 95–98 (2011). Medline doi:10.1038/nature10347

4. S. Letendre, J. Marquie-Beck, E. Capparelli, B. Best, D. Clifford, A. C. Collier, B. B. Gelman, J. C. McArthur, J. A. McCutchan, S. Morgello, D. Simpson, I. Grant, R. J. Ellis; CHARTER Group, Validation of the CNS Penetration-Effectiveness rank for quantifying antiretroviral penetration into the central nervous system. *Arch. Neurol.* **65**, 65–70 (2008). Medline doi:10.1001/archneurol.2007.31

5. C. V. Fletcher, K. Staskus, S. W. Wietgrefe, M. Rothenberger, C. Reilly, J. G. Chipman, G. J. Beilman, A. Khoruts, A. Thorkelson, T. E. Schmidt, J. Anderson, K. Perkey, M. Stevenson, A. S. Perelson, D. C. Douek, A. T. Haase, T. W. Schacker, Persistent HIV-1 replication is associated with lower antiretroviral drug concentrations in lymphatic tissues. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2307–2312 (2014). Medline doi:10.1073/pnas.1318249111

6. J. R. Bailey, A. R. Sedaghat, T. Kieffer, T. Brennan, P. K. Lee, M. Wind-Rotolo, C. M. Haggerty, A. R. Kamireddi, Y. Liu, J. Lee, D. Persaud, J. E. Gallant, J. Cofrancesco Jr., T. C. Quinn, C. O. Wilke, S. C. Ray, J. D. Siliciano, R. E. Nettles, R. F. Siliciano, Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *J. Virol.* **80**, 6441–6457 (2006). Medline doi:10.1128/JVI.00591-06

7. N. Chomont, M. El-Far, P. Ancuta, L. Trautmann, F. A. Procopio, B. Yassine-Diab, G. Boucher, M. R. Boulassel, G. Ghattas, J. M. Brenchley, T. W. Schacker, B. J. Hill, D. C. Douek, J. P. Routy, E. K. Haddad, R. P. Sékaly, HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009). Medline doi:10.1038/nm.1972

8. T. A. Wagner, J. L. McKernan, N. H. Tobin, K. A. Tapia, J. I. Mullins, L. M. Frenkel, An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J. Virol.* **87**, 1770–1778 (2013). Medline doi:10.1128/JVI.01985-12

9. T. Ikeda, J. Shibata, K. Yoshimura, A. Koito, S. Matsushita, Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J. Infect. Dis.* **195**, 716–725 (2007). Medline doi:10.1086/510915

10. H. Imamichi, V. Natarajan, J. W. Adelsberger, C. A. Rehm, R. A. Lempicki, B. Das, A. Hazen, T. Imamichi, H. C. Lane, Lifespan of effector memory CD4+ T cells determined by replication-incompetent integrated HIV-1 provirus. *AIDS* 1 (2014). Medline doi:10.1097/QAD.0000000000000223

11. S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, P. A. Futreal, COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39** (Database), D945–D950 (2011). Medline doi:10.1093/nar/gkq929

12. M. S. Lawrence, P. Stojanov, C. H. Mermel, J. T. Robinson, L. A. Garraway, T. R. Golub, M. Meyerson, S. B. Gabriel, E. S. Lander, G. Getz, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014). Medline doi:10.1038/nature12912

13. M. Zhao, J. Sun, Z. Zhao, TSGene: A web resource for tumor suppressor genes. *Nucleic Acids Res.* **41** (D1), D970–D976 (2013). Medline doi:10.1093/nar/gks937

14. C. Kandoth, M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, L. Ding, Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013). Medline doi:10.1038/nature12634

15. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz Jr., K. W. Kinzler, Cancer genome landscapes. *Science* **339**, 1546–1558 (2013). Medline doi:10.1126/science.1235122

16. H. Fan, C. Johnson, Insertional oncogenesis by non-acute retroviruses: Implications for gene therapy. *Viruses* **3**, 398–422 (2011). Medline doi:10.3390/v3040398

17. A. R. Schröder, P. Shinn, H. Chen, C. Berry, J. R. Ecker, F. Bushman, HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529 (2002). Medline doi:10.1016/S0092-8674(02)00864-4

18. F. Bushman, M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, C. Hoffmann, Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3**, 848–858 (2005). Medline doi:10.1038/nrmicro1263

19. R. S. Mitchell, B. F. Beitzel, A. R. Schroder, P. Shinn, H. Chen, C. C. Berry, J. R. Ecker, F. D. Bushman, Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLOS Biol.* **2**, e234 (2004). Medline doi:10.1371/journal.pbio.0020234

20. M. J. Soto, A. Peña, F. G. Vallejo, A genomic and bioinformatics analysis of the integration of HIV in peripheral blood mononuclear cells. *AIDS Res. Hum. Retroviruses* **27**, 547–555 (2011). Medline

21. G. P. Wang, A. Ciuffi, J. Leipzig, C. C. Berry, F. D. Bushman, HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17**, 1186–1194 (2007). Medline doi:10.1101/gr.6286907

22. D. Persaud, G. K. Siberry, A. Ahonkhai, J. Kajdas, D. Monie, N. Hutton, D. C. Watson, T. C. Quinn, S. C. Ray, R. F. Siliciano, Continued production of drug-sensitive human immunodeficiency virus type 1 in children on combination antiretroviral therapy who have undetectable viral loads. *J. Virol.* **78**, 968–979 (2004). Medline doi:10.1128/JVI.78.2.968-979.2004

23. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57–70 (2000). Medline doi:10.1016/S0092-8674(00)81683-9

24. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011). Medline doi:10.1016/j.cell.2011.02.013

25. C. L. Plaisier, M. Pan, N. S. Baliga, A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. *Genome Res.* **22**, 2302–2314 (2012). Medline doi:10.1101/gr.133991.111

26. Y. Han, K. Lassen, D. Monie, A. R. Sedaghat, S. Shimoji, X. Liu, T. C. Pierson, J. B. Margolick, R. F. Siliciano, J. D. Siliciano, Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J. Virol.* **78**, 6122–6133 (2004). Medline doi:10.1128/JVI.78.12.6122-6133.2004

27. Y. C. Ho, L. Shan, N. N. Hosmane, J. Wang, S. B. Laskey, D. I. Rosenbloom, J. Lai, J. N. Blankson, J. D. Siliciano, R. F. Siliciano, Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013). Medline doi:10.1016/j.cell.2013.09.020

28. S. Swaminathan, C. Huang, H. Geng, Z. Chen, R. Harvey, H. Kang, C. Ng, B. Titz, C. Hurtz, M. F. Sadiyah, D. Nowak, G. B. Thoennissen, V. Rand, T. G. Graeber, H. P. Koeffler, W. L. Carroll, C. L. Willman, A. G. Hall, K. Igarashi, A. Melnick, M. Müschen, BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint. *Nat. Med.* **19**, 1014–1022 (2013). Medline doi:10.1038/nm.3247

29. R. Roychoudhuri, K. Hirahara, K. Mousavi, D. Clever, C. A. Klebanoff, M. Bonelli, G. Sciumè, H. Zare, G. Vahedi, B. Dema, Z. Yu, H. Liu, H. Takahashi, M. Rao, P. Muranski, J. G. Crompton, G. Punkosdy, D. Bedognetti, E. Wang, V. Hoffmann, J. Rivera, F. M. Marincola, A. Nakamura, V. Sartorelli, Y. Kanno, L. Gattinoni, A. Muto, K. Igarashi, J. J. O'Shea, N. P. Restifo, BACH2 represses effector programs to stabilize T$_{reg}$-mediated immune homeostasis. *Nature* **498**, 506–510 (2013). Medline doi:10.1038/nature12199

30. S. Tsukumo, M. Unno, A. Muto, A. Takeuchi, K. Kometani, T. Kurosaki, K. Igarashi, T. Saito, Bach2 maintains T cells in a naive state by suppressing effector memory-related genes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 10735–10740 (2013). Medline doi:10.1073/pnas.1306691110

31. M. Kuwahara, J. Suzuki, S. Tofukuji, T. Yamada, M. Kanoh, A. Matsumoto, S. Maruyama, K. Kometani, T. Kurosaki, O. Ohara, T. Nakayama, M. Yamashita, The Menin-Bach2 axis is critical for regulating CD4 T-cell senescence and cytokine homeostasis. *Nat. Commun.* **5**, 3555 (2014). 10.1038/ncomms4555 Medline doi:10.1038/ncomms4555

32. J. Liu, A. B. Sørensen, B. Wang, M. Wabl, A. L. Nielsen, F. S. Pedersen, Identification of novel Bach2 transcripts and protein isoforms through tagging analysis of retroviral integrations in B-cell lymphomas. *BMC Mol. Biol.* **10**, 2 (2009). 10.1186/1471-2199-10-2 Medline doi:10.1186/1471-2199-10-2

33. Y. Han, M. Wind-Rotolo, H. C. Yang, J. D. Siliciano, R. F. Siliciano, Experimental approaches to the study of HIV-1 latency. *Nat. Rev. Microbiol.* **5**, 95–106 (2007). Medline doi:10.1038/nrmicro1580

34. M. K. Lewinski, D. Bisgrove, P. Shinn, H. Chen, C. Hoffmann, S. Hannenhalli, E. Verdin, C. C. Berry, J. R. Ecker, F. D. Bushman, Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. *J. Virol.* **79**, 6610–6619 (2005). Medline doi:10.1128/JVI.79.11.6610-6619.2005

35. L. Holmfeldt, L. Wei, E. Diaz-Flores, M. Walsh, J. Zhang, L. Ding, D. Payne-Turner, M. Churchman, A. Andersson, S. C. Chen, K. McCastlain, J. Becksfort, J. Ma, G. Wu, S. N. Patel, S. L. Heatley, L. A. Phillips, G. Song, J. Easton, M. Parker, X. Chen, M. Rusch, K. Boggs, B. Vadodaria, E. Hedlund, C. Drenberg, S. Baker, D. Pei, C. Cheng, R. Huether, C. Lu, R. S. Fulton, L. L. Fulton, Y. Tabib, D. J. Dooling, K. Ochoa, M. Minden, I. D. Lewis, L. B. To, P. Marlton, A. W. Roberts, G. Raca, W. Stock, G. Neale, H. G. Drexler, R. A. Dickins, D. W. Ellison, S. A. Shurtleff, C. H. Pui, R. C. Ribeiro, M. Devidas, A. J. Carroll, N. A. Heerema, B. Wood, M. J. Borowitz, J. M. Gastier-Foster, S. C. Raimondi, E. R. Mardis, R. K. Wilson, J. R. Downing, S. P. Hunger, M. L. Loh, C. G. Mullighan, The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat. Genet.* **45**, 242–252 (2013). Medline doi:10.1038/ng.2532

36. I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, M. A. Nowak, Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18545–18550 (2010). Medline doi:10.1073/pnas.1010978107

37. L. Josefsson, S. Palmer, N. R. Faria, P. Lemey, J. Casazza, D. Ambrozak, M. Kearney, W. Shao, S. Kottilil, M. Sneller, J. Mellors, J. M. Coffin, F. Maldarelli, Single cell analysis of lymph node tissue from HIV-1 infected patients reveals that the majority of CD4$^+$ T-cells contain one HIV-1 DNA molecule. *PLOS Pathog.* **9**, e1003432 (2013). 10.1371/journal.ppat.1003432 Medline doi:10.1371/journal.ppat.1003432

38. J. A. Anderson, N. M. Archin, W. Ince, D. Parker, A. Wiegand, J. M. Coffin, J. Kuruc, J. Eron, R. Swanstrom, D. M. Margolis, Clonal sequences recovered from plasma from patients with residual HIV-1 viremia and on intensified antiretroviral therapy are identical to replicating viral RNAs recovered from circulating resting CD4+ T cells. *J. Virol.* **85**, 5220–5223 (2011). Medline doi:10.1128/JVI.00284-11

39. H. P. Mok, S. Javed, A. Lever, Stable gene expression occurs from a minority of integrated HIV-1-based vectors: Transcriptional silencing is present in the majority. *Gene Ther.* **14**, 741–751 (2007). Medline doi:10.1038/sj.gt.3302923

40. K. Imai, H. Togami, T. Okamoto, Involvement of histone H3 lysine 9 (H3K9) methyltransferase G9a in the maintenance of HIV-1 latency and its reactivation by BIX01294. *J. Biol. Chem.* **285**, 16538–16545 (2010). Medline doi:10.1074/jbc.M110.103531

41. J. Friedman, W. K. Cho, C. K. Chu, K. S. Keedy, N. M. Archin, D. M. Margolis, J. Karn, Epigenetic silencing of HIV-1 by the histone H3 lysine 27 methyltransferase enhancer of Zeste 2. *J. Virol.* **85**, 9078–9089 (2011). Medline doi:10.1128/JVI.00836-11

42. P. G. Eipers, J. F. Salazar-Gonzalez, C. D. Morrow, HIV gene expression from intact proviruses positioned in bacterial artificial chromosomes at integration sites previously identified in latently infected T cells. *Virology* **410**, 151–160 (2011). Medline doi:10.1016/j.virol.2010.11.001

43. S. Sherrill-Mix, M. K. Lewinski, M. Famiglietti, A. Bosque, N. Malani, K. E. Ocwieja, C. C. Berry, D. Looney, L. Shan, L. M. Agosto, M. J. Pace, R. F. Siliciano, U. O'Doherty, J. Guatelli, V. Planelles, F. D. Bushman, HIV latency and integration site placement in five cell-based models. *Retrovirology* **10**, 90 (2013). Medline doi:10.1186/1742-4690-10-90

44. L. M. Frenkel, Y. Wang, G. H. Learn, J. L. McKernan, G. M. Ellis, K. M. Mohan, S. E. Holte, S. M. De Vange, D. M. Pawluk, A. J. Melvin, P. F. Lewis, L. M. Heath, I. A. Beck, M. Mahalanabis, W. E. Naugler, N. H. Tobin, J. I. Mullins, Multiple viral genetic analyses detect low-level human immunodeficiency virus type 1 replication during effective highly active antiretroviral therapy. *J. Virol.* **77**, 5721–5730 (2003). Medline doi:10.1128/JVI.77.10.5721-5730.2003

45. P. Armitage, Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386 (1955). doi:10.2307/3001775

46. K. A. Clouse, D. Powell, I. Washington, G. Poli, K. Strebel, W. Farrar, P. Barstad, J. Kovacs, A. S. Fauci, T. M. Folks, Monokine regulation of human immunodeficiency virus-1 expression in a chronically infected human T cell clone. *J. Immunol.* **142**, 431–438 (1989). Medline

47. T. M. Folks, K. A. Clouse, J. Justement, A. Rabson, E. Duh, J. H. Kehrl, A. S. Fauci, Tumor necrosis factor alpha induces expression of human immunodeficiency virus in a chronically infected T-cell clone. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2365–2368 (1989). Medline doi:10.1073/pnas.86.7.2365

48. M. D. Daniel *et al*., Simian immunodeficiency virus from African green monkeys. *J. Virol.* **62**, 4123–4128 (1988). Medline

49. R. Kikukawa, Y. Koyanagi, S. Harada, N. Kobayashi, M. Hatanaka, N. Yamamoto, Differential susceptibility to the acquired immunodeficiency syndrome retrovirus in cloned cells of human leukemic T-cell line Molt-4. *J. Virol.* **57**, 1159–1162 (1986). Medline

50. A. G. Rodrigo, P. C. Goracke, K. Rowhanian, J. I. Mullins, Quantitation of target molecules from polymerase chain reaction-based limiting dilution assays. *AIDS Res. Hum. Retroviruses* **13**, 737–742 (1997). Medline doi:10.1089/aid.1997.13.737

51. J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T. L. Madden, Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012). 10.1186/1471-2105-13-134 Medline doi:10.1186/1471-2105-13-134

52. S. L. Liu, T. Schacker, L. Musey, D. Shriner, M. J. McElrath, L. Corey, J. I. Mullins, Divergent patterns of progression to AIDS after infection from the same source: Human

immunodeficiency virus type 1 evolution and antiviral responses. *J. Virol.* **71**, 4284–4295 (1997). Medline

53. M. Altfeld, M. M. Addo, R. Shankarappa, P. K. Lee, T. M. Allen, X. G. Yu, A. Rathod, J. Harlow, K. O'Sullivan, M. N. Johnston, P. J. Goulder, J. I. Mullins, E. S. Rosenberg, C. Brander, B. Korber, B. D. Walker, Enhanced detection of human immunodeficiency virus type 1-specific T-cell responses to highly variable regions by using peptides based on autologous virus sequences. *J. Virol.* **77**, 7330–7340 (2003). Medline doi:10.1128/JVI.77.13.7330-7340.2003

54. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). Medline doi:10.1038/nmeth.1923

55. W. J. Kent, BLAT: The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002). Medline doi:10.1101/gr.229202

56. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI Reference Sequence project: Update and current status. *Nucleic Acids Res.* **31**, 34–37 (2003). Medline doi:10.1093/nar/gkg111

57. F. Hsu, W. J. Kent, H. Clawson, R. M. Kuhn, M. Diekhans, D. Haussler, The UCSC Known Genes. *Bioinformatics* **22**, 1036–1046 (2006). Medline doi:10.1093/bioinformatics/btl048

58. A. Alexa, J. Rahnenführer, T. Lengauer, Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006). Medline doi:10.1093/bioinformatics/btl140

59. S. Holm, A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6**, 65–70 (1979).