

RESEARCH ARTICLE

HIV LATENCY

Specific HIV integration sites are linked to clonal expansion and persistence of infected cells

F. Maldarelli,^{1*} X. Wu,^{2*} L. Su,² F. R. Simonetti,^{1,3} W. Shao,² S. Hill,¹ J. Spindler,¹ A. L. Ferris,¹ J. W. Mellors,⁴ M. F. Kearney,¹ J. M. Coffin,⁵ S. H. Hughes^{1†}

The persistence of HIV-infected cells in individuals on suppressive combination antiretroviral therapy (cART) presents a major barrier for curing HIV infections. HIV integrates its DNA into many sites in the host genome; we identified 2410 integration sites in peripheral blood lymphocytes of five infected individuals on cART. About 40% of the integrations were in clonally expanded cells. Approximately 50% of the infected cells in one patient were from a single clone, and some clones persisted for many years. There were multiple independent integrations in several genes, including *MKL2* and *BACH2*; many of these integrations were in clonally expanded cells. Our findings show that HIV integration sites can play a critical role in expansion and persistence of HIV-infected cells.

HIV replication is suppressed with combination antiretroviral therapy (cART), but infected cells persist in patients and are a critical obstacle to curing HIV infection (1, 2). Analysis of HIV populations in vivo shows that after long-term suppressive cART, genetically identical HIV variants emerge (3, 4). The source and mechanisms involved in the emergence of these identical variants are not understood. One possibility is that the identical variants arise from cells that have clonally expanded. Because HIV DNA integrates at many sites in the human genome, the site of integration can be used to identify clonally expanded cells that arose from a single infected progenitor.

Clonal expansion of HIV-infected cells in patients

We analyzed the integration sites in peripheral blood mononuclear cells (PBMCs) or CD4⁺ T cells from patients on prolonged cART obtained by negative selection using a previously described technique (5–7). DNA from PBMCs or CD4⁺ T cells was randomly sheared to ~400-base pair (bp) fragments, and linker-mediated polymerase chain reaction was used to selectively amplify fragments that contained viral/host DNA junctions (8). Both ends of the amplified junction fragments were sequenced on the Il-

lumina platform (San Diego, CA) to determine the viral/host junctions and the breakpoints in the host DNA. The sequence of the viral/host junction identifies the exact position and orientation in which the HIV DNA was integrated. The breakpoints in the host DNA can be used to identify the integration sites in clonally expanded cells. If several cells with the same integration site are present, shearing their DNA will give rise to multiple fragments in which the integration site is the same, but the host DNA breakpoints differ.

Integration site analysis was performed by using PBMCs or CD4⁺ T cells from five patients (table S1). A diverse population of viruses was present in each patient either before therapy or shortly after cART was initiated; however, after prolonged cART (mean duration of treatment, 11.7 years), identical viral sequences, defined at either the RNA or the DNA level, emerged (Fig. 1 and fig. S1). In total, 2410 integration sites were mapped; these represented 1632 different integration events. 1388 integration sites (57%) were detected once, and 1022 sites (43%) were associated with more than one host DNA breakpoint, revealing that a large fraction of the infected cells are from expanded clones (Fig. 2, tables S2 and S3, and fig. S2). We validated the method for identifying expanded clones by making two completely independent libraries from cells from the last time point from patient 1; the same highly expanded clones were identified in both libraries. In some cases, clonal expansion was extensive. For example, in patient 3 the initial analysis found the same site in the *HORMAD2* gene in 62 of 317 integrations, but because of overlaps in the breakpoint analysis, this figure is an underestimate. We also estimated the fraction of the total integration

sites that were derived from this expanded clone (8). This analysis implied that ~58% of all the HIV-infected cells in this patient were derived from a single infected cell.

Among the five patients studied, HIV integration sites were found in 985 different genes (fig. S3). Most of the genes had only a single integrant (659 integrants, 67%); these integrants were not shown to be from clonally expanded cells. The remaining 326 genes (33%) either had a single integrant in a cell that underwent clonal expansion (126 genes) or had multiple integrants (200 genes), some of which (59 genes; 30%) were in cells that underwent clonal expansion (fig. S3). Approximately 70% (21 of 29) of the genes with multiple integrants in highly expanded clones (listed in table S4) are known to be directly involved in the regulation of cell growth. HIV proviruses (the integrated form of retroviral DNA is called a provirus) were also found in intergenic regions of the genome and in sites that could not be mapped to a single site in the human genome (ambiguous sites); some of the cells with such integrants were also highly expanded (fig. S2).

Persistence of clones of HIV-infected cells

Longitudinal sampling revealed that some of the clones that emerged on cART persisted for many years. Patient 1 had at least 13 different clones that were present in samples taken 11.2 years apart, and 11 other clones were present in samples taken 6.6 years apart (Table 1). A number of these persistent clones had integration sites in genes known to be associated with cell growth (*STAT5B*, *PARP8*, and *DDX6*), mitosis (*PKP4* and *MAP4*), or both. In patient 3, we analyzed a small set of integration sites (47 total sites) from a sample taken before the patient started cART. Most of the integration sites from this pre-cART sample were present only once, an observation that is consistent with the short half-life of infected cells during uncontrolled HIV replication (9). However, one site was from a clonally expanded cell, showing that there is clonal expansion in the absence of cART. Others have reported that there are HIV-infected clones that can persist for prolonged periods in patients on cART (10, 11). Our data show that the prolonged persistence of expanded clones is common and frequently associated with specific integrations in genes involved in controlling cell growth and division. Although there was variation in integration sites among patients, all five patients showed evidence of clonal expansion of infected cells, even in the smallest data sets (35 and 46 distinct integration sites) (table S2).

Integration sites associated with clonal expansion or persistence of infected cells

Two genes with remarkable patterns of HIV integration were identified in patient 1 (Fig. 3). In the data set obtained from CD4⁺ T cells after

¹HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702, USA. ²Leidos Biomedical Research, Frederick, MD 21702, USA. ³Department of Biomedical and Clinical Sciences L. Sacco, University of Milan, 20122 Milan, Italy. ⁴Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA. ⁵Department of Molecular Biology and Microbiology, Tufts University, Boston, MA 02111, USA. *These authors contributed equally to this work. †Corresponding author. E-mail: hughesst@mail.nih.gov

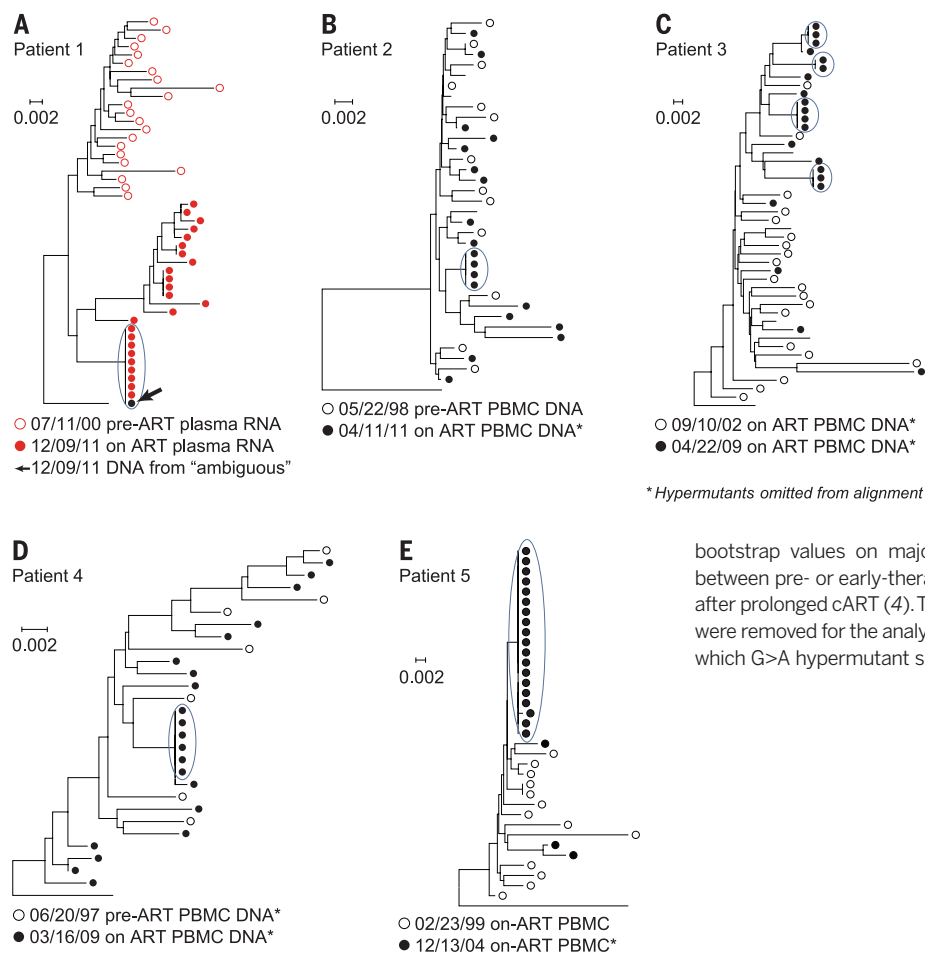
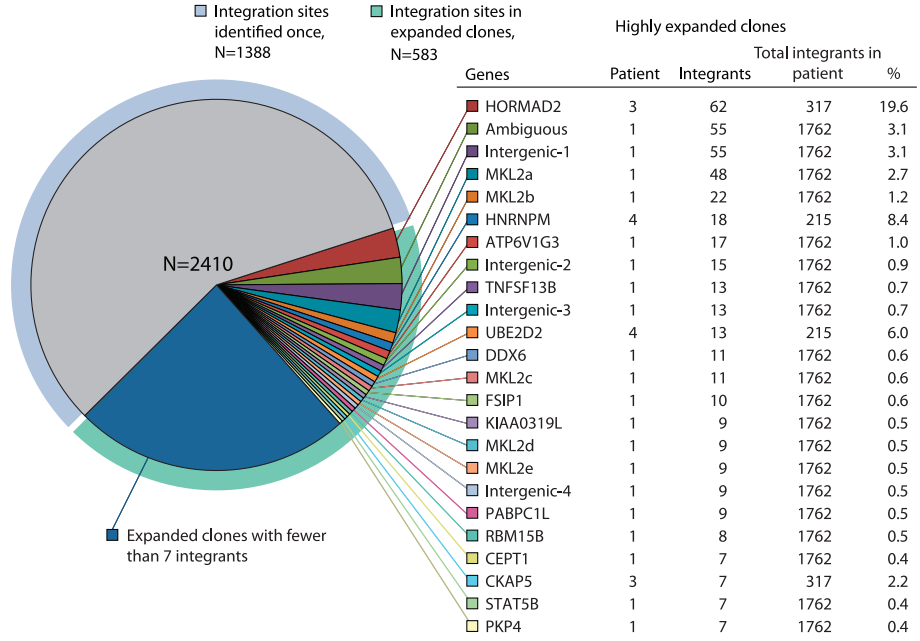


Fig. 1. Long-term cART reveals the presence in patients, at both the RNA and the DNA level, of HIV genomes that have identical sequences. (A to E) Single-genome sequences were obtained from plasma virion RNA or from PBMC DNA prior to or shortly after the initiation of cART, and after long-term cART. Sequences were aligned by using Clustal W (Clustal is maintained at the Conway Institute, Dublin, Ireland), and neighbor-joining trees were rooted on the consensus sequence of subgroup B HIV (8). Virus populations from pretherapy samples, or samples taken shortly after the initiation of cART, are shown as open circles (red, virion RNA; black, PBMC DNA); populations after prolonged cART are shown as solid circles (red, virion RNA; black, PBMC DNA). (A) In patient 1, the black dot (indicated with the arrow) represents DNA sequences from the provirus in a clone of expanded cells, whose integration site could not be mapped (Fig. 2, “ambiguous”). Short branches and low bootstrap values on major nodes of the trees support a lack of divergence between pre- or early-therapy sequences and populations of identical sequences after prolonged cART (4). To avoid distorting the trees, all hypermutant sequences were removed for the analysis shown in the figure. Asterisk denotes the trees from which G>A hypermutant sequences were removed.

Fig. 2. Distribution of integration sites in the five patients. A total of 2410 integration sites were obtained from PBMCs or negatively selected CD4⁺ cells from the five patients. Genes in which we isolated a particular integration site seven or more times are shown. The table shows which patient harbored each of these expanded clones, how many times the different integrations sites were isolated, and the fraction of the infected cells the expanded clone represents in the patient. The MKL2 clones marked MKL2a, -b, -c, and -d correspond to the integration sites marked a, b, c, and d in Fig. 3. One provirus in a highly expanded clone in patient 1 was in a sequence that could not be unambiguously assigned in the human genome (denoted “ambiguous”). Our standard analysis (presented here) underestimated the fraction of the infected cells in patient 3 that had an integration site in HORMAD2 (8).



11.4 years of cART, there were 11 distinct integration sites in intron 6 of *MKL2* (intron 6 is ~3.5 kb), more than half of which were in clonally expanded cells; some of these cells were highly expanded (Figs. 2 and 3). There were also four nearby integration sites in intron 4, and none in any other part of *MKL2* (Fig. 3A). All 15 of these proviruses were integrated in the same transcriptional orientation as the host gene. Thus, ~7% of the infected cells in this patient had proviruses in a region that constitutes a very small fraction ($\sim 2 \times 10^{-6}$) of the human genome. In the same data set, there were 15 independent integration sites—also in the same transcriptional orientation as the host gene—in introns 4 and 5 of *BACH2* (Fig. 3B); two additional integration sites in *BACH2* were identified in earlier samples from this patient.

For comparison, we analyzed two large HIV integration site libraries made from acutely infected HeLa cells (~250,000 sites) and human CD34⁺ hematopoietic stem cells (~150,000 sites). The frequencies of HIV integration in *MKL2* and *BACH2* in cells from patient 1 were much greater than in HeLa or CD34⁺ cells. In cells from patient 1, integrations in *MKL2* were 7% of the total integrations, compared with 0.03% of total integrations in HeLa and CD34⁺ cells. Similarly, integrations in *BACH2* in cells from patient 1 were 1.5% of the total integrations, compared with 0.002% in HeLa cells and 0.01% in CD34⁺ cells. There was no preference for integrations in specific introns in

these genes in HeLa cells or CD34⁺ cells. Nor was there any indication, in either library, of preferential integration in one orientation in *MKL2* or *BACH2*. Across the entire patient data set (table S3), there was a weak, but significant preference for integrants in genes to be in an orientation opposite to the direction in which the gene is transcribed. Of the 1313 integrants that were in genes, 594 were in the same transcriptional orientation as the gene, and 719 were in the opposite orientation ($P = 0.02$, Fisher's exact test). These findings led us to conclude that the HIV DNA insertions in *MKL2* and *BACH2* were selected, after integration, because they altered the level of expression of the *MKL2* and *BACH2* proteins or gave rise to the expression of altered forms of the proteins, and that these alterations affected the expansion and survival of the infected cells. In the case of *BACH2*, all of the integrations were upstream of the initiation site for translation, which is consistent with the integrations altering the level of expression of *BACH2*. The integrations in *MKL2* were in introns that were between two coding exons, and these integrations were more likely to have affected the structure of the protein. Both of these mechanisms have been seen with other types of retroviruses and are known to be involved in oncogenic transformation in animals (12).

In all, we found that out of the 985 genes in which there were integrations in the patients, there were 200 genes that had multiple inde-

pendent integrations; 59 of these were associated with expanded clones. The genes that had at least three independent integrants are shown in table S4; as mentioned earlier, many of these genes have roles in cell growth. In addition, there were integrations in more than one patient in more than 60% (18 of 29) of the genes listed in table S4. For example, a total of 10 independent integrants were found in *STAT5B* in four of the five patients; some of these integrants were in expanded clones. However, the proviruses integrated in *STAT5B* showed no orientation preference. Gene ontology analysis showed that the patient integration sites were enriched for genes in several pathways involved in cell growth. The HeLa and human CD34⁺ cell data sets (which were similar to each other) were not enriched for genes in these pathways (fig. S4). This analysis also showed that the patient data set was related to leukemia and Burkitt's lymphoma; the HeLa and human CD34⁺ data sets were not associated with any disease-related pathways.

Although as expected (13) most of the integration sites in the patients were in genes, 21% (509) were in intergenic regions or in sequences that could not be mapped to a single location in the human genome. Forty-four percent (226) of the integrants in the intergenic regions, or that could be mapped to a single site, were in cells that underwent clonal expansion, some of which were highly expanded (Fig. 2 and fig. S2). One of the proviruses in the two

Table 1. In patient 1, clones persist for many years.

Gene	Integration site	Detected in year*			Orientation†	Gene description
		0.2	4.8	11.4		
DDX6	chr11:118636877	+	+	+	Reverse	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
TNFSF13B	chr13:108927255	+	+	+	Forward	tumor necrosis factor (ligand) superfamily, member 13
PKP4	chr2:159370503	+	+	+	Forward	plakophilin 4
RBM15B	chr3:51434782	+	+	+	Reverse	RNA binding motif protein 15B
PARP8	chr5:50035007	+	+	+	Reverse	poly (ADP-ribose) polymerase family, member 8
ATP6V1G3	chr1:198506789	–	+	+	Forward	ATPase, H ⁺ transporting, lysosomal 13 kD, V1 subunit G3
SLC30A7	chr1:101445575	–	+	+	Forward	solute carrier family 30 (zinc transporter), member 7
ZCCHC11	chr1:52956520	–	+	+	Reverse	zinc finger, CCHC domain containing 11
NUMA1	chr11:71744577	–	+	+	Forward	nuclear mitotic apparatus protein 1
CTAGE5	chr14:39760983	–	+	+	Reverse	CTAGE family, member 5
PATL2	chr15:44962782	–	+	+	Forward	protein associated with topoisomerase II homolog 2 (yeast)
P4HB‡	chr17:79800929	–	+	+	Reverse	prolyl 4-hydroxylase, beta polypeptide
ZC3H4	chr19:47603701	–	+	+	Reverse	zinc finger CCCH-type containing 4
COG5	chr7:107182849	–	+	+	Reverse	component of oligomeric golgi complex 5
GIMAP6‡	chr7:150318109	–	+	+	Reverse	GTPase, IMAP family member 6
ZNF16	chr8:146170620	–	+	+	Forward	zinc finger protein 16
R3HDM2‡	chr12:57731640	+	–	+	Forward	R3H domain containing 2 (R3HDM2), mRNA
FSIP1	chr15:39975096	+	–	+	Forward	fibrous sheath interacting protein 1
NLR5	chr16:57030643	+	–	+	Forward	NLR family, CARD domain containing 5
STAT5B	chr17:40421711	+	–	+	Reverse	signal transducer and activator of transcription 5B
PABPC1L	chr20:43561100	+	–	+	Reverse	poly(A) binding protein, cytoplasmic 1-like
MAP4	chr3:47981132	+	–	+	Reverse	microtubule-associated protein 4
BTNL3‡	chr5:180401789	+	–	+	Forward	butyrophilin-like 3
SQLE	chr8:126012090	+	–	+	Forward	squalene epoxidase
CHD7	chr8:61594573	+	+	–	Reverse	chromodomain helicase DNA binding protein 7

*Relative to start of cART.

†The orientation of the provirus relative to the direction of transcription of the gene.

‡Nearest gene.

most highly expanded clones in patient 1 [each site was identified 55 times, which is an underestimate because of breakpoint overlaps (8)] was in an intergenic region; the other could not be mapped to a single site in the human genome (Fig. 2). The predominant virus in the plasma of patient 1 late in therapy was clonal (Fig. 1 and fig. S1) and was insensitive to a switch in cART (Fig. 1 and fig. S1), suggesting that it was produced by a clone of infected cells, rather than from infection of new cells. More than 1 kb of sequence in *gag-pro-pol*

from the RNA genome of this predominant virus exactly matched the sequence of the ambiguously mapped provirus, identifying this provirus as the source of the clonal viral RNA in the plasma (Fig. 1A, black arrow).

Discussion

Our results strongly imply that in at least some cases, sites of HIV integration play an important role in the expansion and persistence of infected cells in patients. This conclusion is particularly strong for the integrations into

specific introns of the *MKL2* and *BACH2* genes. The integrations in *MKL2* and *BACH2* that were linked to clonal expansion were in internal introns, and in the same transcriptional orientation as the genes in which they are inserted. Even setting aside the fact that that it is extremely unlikely that such a large fraction of the integrations would have occurred in these two small segments of the genome, the probability that all 33 of the integrations we saw in *BACH2* and *MKL2* in the patients (table S4) would have been in the same orientation as

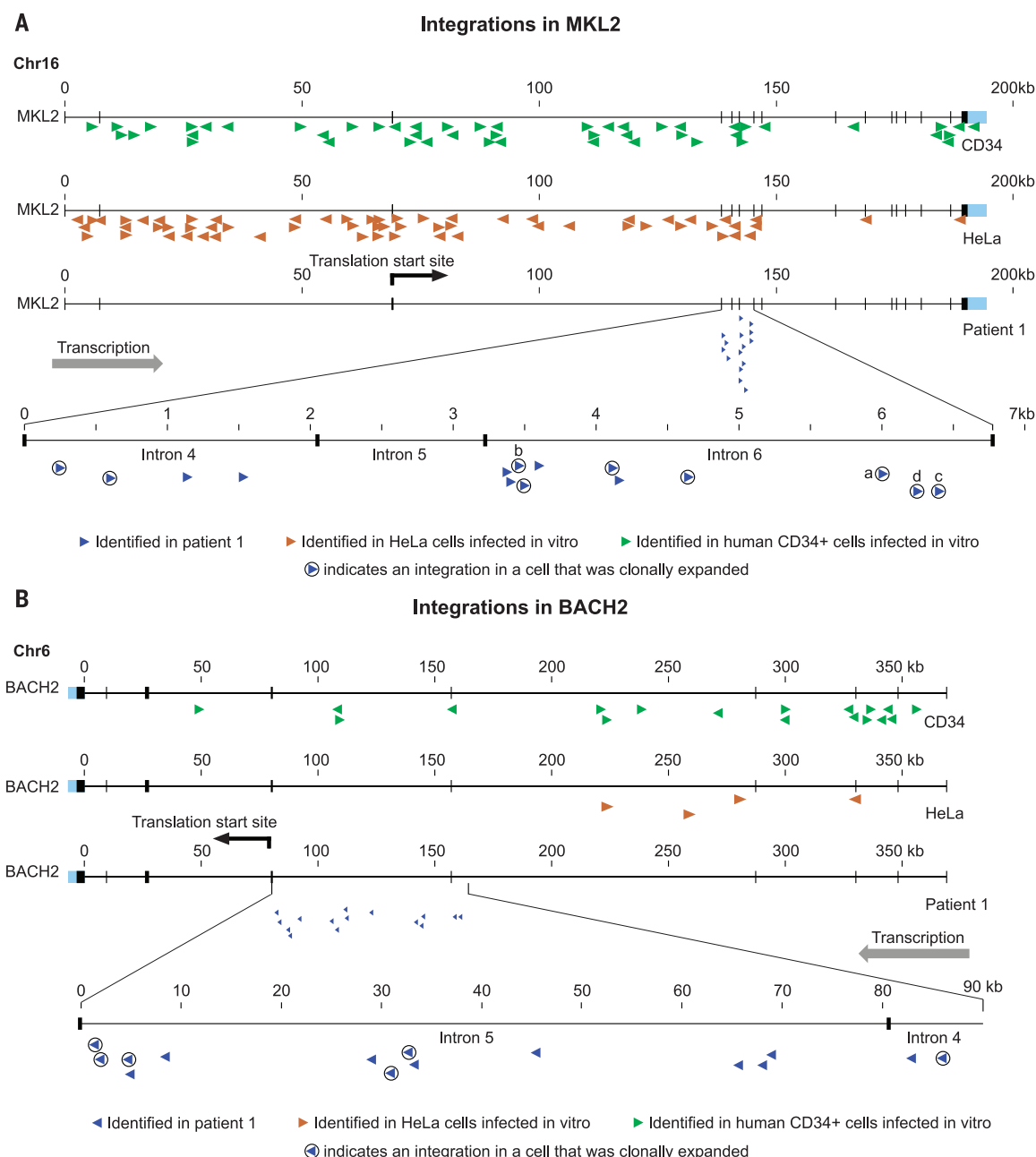


Fig. 3. Integration sites in the *MKL2* and *BACH2* genes in patient 1 after 11.4 years of cART. (A) There were 15 distinct integration sites (blue arrows) in a small region of the *MKL2* gene in patient 1. The arrows denote the transcriptional orientation of each provirus. The circled arrows indicate integration sites in clonally expanded cells. The arrows marked a, b, c, and d correspond to the clones marked MKL2a, -b, -c, and -d in Fig. 2. (B) There were 15 distinct integration sites in a small region of the *BACH2* gene in patient 1 (blue arrows). Some of these integration sites were in clonally expanded cells (circled arrows). (A) and (B) also show the HIV integration sites identified in the same two genes in acutely infected HeLa cells (total sites, 248,658; brown arrows) and CD34⁺ cells (total sites, 159,484; green arrows).

the genes is $\sim 10^{-10}$. In prior studies, a limited number of HIV integration sites were identified in patients (14–17), and HIV proviruses were found in intron 6 of *MKL2* and intron 5 of *BACH2*, in the same orientation that the genes are transcribed. However, in the published studies the integration sites in these genes were not linked to the clonal expansion of the infected cells (table S5). Both *BACH2* and *MKL2* are involved in the growth and development of cells, and *BACH2* is known to play a key role in T cell development (18). Both genes [and the *MKL2*-related gene *MKL1*, in which there were four independent integration sites, some of which were associated with clonally expanded cells in patient 1 (table S4)] have been implicated in human cancers (19–21), in which they were activated by DNA rearrangements that created gene fusions. The pattern of multiple integrations in *MKL2* and *BACH2* found in the patients cannot be the result of preferential integration because HIV integration is neither intron-specific nor orientation-specific (22). Thus, the only plausible explanation for the data that is in accord with the rules for HIV integration is that the cells with the integrations in *MKL2* and *BACH2* were selected after integration because the integrations in these genes contributed to the expansion and persistence of the host cells. This interpretation is supported, for *BACH2*, by a report showing that this gene is a target for retroviral insertional activation in mice infected with mouse leukemia virus (23).

Most of our analyses were performed by using cells from patients on long-term cART, which blocks the infection of additional cells but has no effect on cells that have already been infected (9). During untreated HIV infections, $\sim 10^9$ cells are infected daily. The vast majority (99%) of the newly infected cells die within 24 to 48 hours, and a substantial proportion of the remaining cells die within 2 to 4 weeks (24–26). Viremia decreases by 4 to 5 logs when patients undergo cART; however, the number of cells containing HIV DNA decreases by approximately 1 log (9), indicating that a substantial fraction ($\sim 10\%$) of the cells that were infected before the initiation of cART persist. Most of these long-lived infected cells contain proviruses that are obviously defective; however, $\sim 12\%$ of the proviruses appear to be functional, although only a small fraction of these apparently functional proviruses can be induced to make virus in ex vivo experiments (27). Cells infected with highly defective or fully latent proviruses that produce little or no viral protein may have a survival advantage relative to cells that produce virions because cells that express viral proteins are more likely to be lysed by HIV-specific cytotoxic T lymphocyte or be subject to cytopathic effects of the viral proteins. This same logic applies to HIV-infected cells that undergo clonal expansion. Although we have not yet shown that clonally expanded cells produce replication-competent HIV, we have shown that a highly expanded clone of cells does produce HIV virions in suf-

ficient quantity to cause viremia, which means that the selection against cells that produce viral proteins is not so strong that it prevents extensive clonal expansion of cells that express the viral proteins required to produce virions.

Our data show that many of the infected cells that persist have undergone clonal expansion; these clones were revealed but not created by cART. For some infected cell clones, it is likely that the integration site is only a passive marker of clonal expansions that are driven by another factor or factors, such as antigen stimulation or homeostatic proliferation signals (28). In contrast, we show here that some cells with HIV integration sites in specific genes are strongly selected because these integrations promote the survival and expansion of the infected cells. Although there are obvious similarities in the integration sites seen in the five patients, there is considerable heterogeneity from one patient to another, both in the extent of clonal expansion and in the genes in which proviruses are integrated in the clonally expanded cells (fig. S5). This complexity highlights the difficulty in attempting to extrapolate, from bulk HIV DNA quantification, the size and nature of the population of HIV proviruses that make up the reservoir that gives rise to HIV rebound after cessation of cART (28).

Our findings have relevance for three important areas. (i) To effectively target HIV persistence with the goal of achieving a cure, it will be important not only to suppress any replication of the virus, but also to block the expansion of infected cells. (ii) Although the HIV vectors used in gene therapy have safety features that the parental virus lacks, we now know that like many other retroviruses, HIV integration can lead to clonal expansion and persistence of infected cells. This discovery suggests that persons treated with HIV-based vectors should be carefully monitored for evidence of clonal expansion of vector-infected cells. (iii) We also suggest that it is time to reexamine the question of whether HIV integration can contribute to the development of malignancies. Although there are well-defined cancers in HIV-infected patients that are the result of uncontrolled expression of herpes viruses, there are reports of a small number of lymphomas with HIV proviruses integrated at defined sites; one lymphoma had a provirus integrated in *BACH2* (15, 29, 30). Despite these published reports, it is widely believed that HIV DNA is not detectable in most cancers from HIV-infected patients; however, the experiments supporting this belief are not well-documented in the literature. It is possible that prior attempts to detect HIV DNA in cancers examined only a very small portion of the HIV genome and, as such, missed HIV proviruses having large deletions; large deletions are a characteristic of the proviruses that cause mouse and avian tumors. Thus, our findings have important implications for designing and implementing strategies to eliminate persistent HIV infection, for the use of lentiviral vectors for gene therapy in human patients and, possibly, for the origin of some HIV-related malignancies.

REFERENCES AND NOTES

1. S. G. Deeks et al., *Nat. Rev. Immunol.* **12**, 607–614 (2012).
2. S. Palmer, L. Josefsson, J. M. Coffin, *J. Intern. Med.* **270**, 550–560 (2011).
3. J. R. Bailey et al., *J. Virol.* **80**, 6441–6457 (2006).
4. M. F. Kearney et al., *PLoS Pathog.* **10**, e1004010 (2014).
5. C. C. Berry et al., *Bioinformatics* **28**, 755–762 (2012).
6. N. A. Gillet et al., *Blood* **117**, 3113–3122 (2011).
7. S. S. De Ravin et al., *J. Virol.* (2014).
8. Materials and methods are available as supplementary materials on Science Online.
9. J. Coffin, R. Swanstrom, *Cold Spring Harb. Perspect. Med.* **3**, a012526 (2013).
10. H. Imamichi et al., *AIDS* (2014).
11. L. Josefsson et al., *Proc. Natl. Acad. Sci. U.S.A.* **110**, E4987–E4996 (2013).
12. N. Rosenberg, P. Jolicoeur, in *Retroviruses*, J. M. Coffin, S. H. Hughes, H. E. Varmus, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1997), pp. 475–585.
13. A. R. Schröder et al., *Cell* **110**, 521–529 (2002).
14. T. Ikeda, J. Shibata, K. Yoshimura, A. Koito, S. Matsushita, *J. Infect. Dis.* **195**, 716–725 (2007).
15. K. D. Mack et al., *J. Acquir. Immune Defic. Syndr.* **33**, 308–320 (2003).
16. Y. Han et al., *J. Virol.* **78**, 6122–6133 (2004).
17. H. Katano et al., *Microbes Infect.* **9**, 1581–1589 (2007).
18. G. Hu, J. Chen, *Nat. Commun.* **4**, 2830 (2013).
19. S. Kobayashi et al., *Genes Chromosomes Cancer* **50**, 207–216 (2011).
20. U. Flucke et al., *Histopathology* **62**, 925–930 (2013).
21. S. Muehlich et al., *Oncogene* **31**, 3913–3923 (2012).
22. R. Craigie, F. D. Bushman, *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012).
23. J. Liu et al., *BMC Mol. Biol.* **10**, 2 (2009).
24. D. D. Ho et al., *Nature* **373**, 123–126 (1995).
25. X. Wei et al., *Nature* **373**, 117–122 (1995).
26. J. M. Coffin, *Science* **267**, 483–489 (1995).
27. Y. C. Ho et al., *Cell* **155**, 540–551 (2013).
28. N. Chomont et al., *Nat. Med.* **15**, 893–900 (2009).
29. B. Shiramizu, B. G. Herndier, M. S. McGrath, *Cancer Res.* **54**, 2069–2072 (1994).
30. B. G. Herndier et al., *Blood* **79**, 1768–1774 (1992).

ACKNOWLEDGMENTS

The authors are indebted to the study participants and to the clinical staff of the National Institute of Allergy and Infectious Diseases/Critical Care Medicine Department clinic who cared for them. We thank C. Lane, H. Malech, H. Imamichi, S. Matsushita, and L. Frenkel for stimulating discussions. We are grateful to J. Meyer and A. Kane for help with the figures and T. Burdette for help in preparing the manuscript. The data presented in this work is tabulated in the main paper and in the supplementary materials. The integration sites are compiled in table S3; the data can also be accessed using the National Center for Biotechnology Information accession no. PRJNA241020. Funding for this research was provided with Federal funds from the National Cancer Institute, an NIH Bench to Bedside award (F.M.), and by funds from the National Cancer Institute under contract HSSN261200800001E (X.W. and L.S.). J.M.C. was supported by a Research Professorship from the American Cancer Society with additional support from the F. M. Kirby Foundation and by funding from the National Cancer Institute (Leidos contract 25XS119). J.W.M. was supported by funding from the National Cancer Institute (Leidos contract 25XS119). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6193/179/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S5
References (31–35)

1 April 2014; accepted 13 June 2014
Published online 26 June 2014;
10.1126/science.1254194