The Role of HIV Integration Sites in Extensive Clonal Expansion of Infected Cells in Patients

F. Maldarelli[1], X. Wu[2], L. Su[2], W. Shao[3], S. Hill[1], F. Simonetti[1], J. Spindler[1], M. Kearney[1], J. Coffin[4], S. H. Hughes[1]

[1]HIV Drug Resistance Program, Center for Cancer Research, National Cancer Institute, Frederick, MD; [2]Laboratory of Molecular Technology, Leidos Biomedical Research, Inc., Frederick National Laboratory, Frederick MD; [3]Advanced Biomedical Computing Center, Information Systems Program, Leidos Biomed, FNL; [4]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston MA

Background

Despite successful suppression of HIV by combination antiretroviral therapy (cART), infected cells persist for many years in patients. Long-term control of viremia by cART reveals clonal viral genomes in the blood of most patients, implying clonal expansion of some HIV-infected cells. The mechanisms driving clonal expansion are unknown but are central to HIV persistence. To address this issue, we determined the distribution of proviral integration sites in 5 patients on successful cART using a specific and highly sensitive amplification strategy that yields large numbers of virus-host junctions, identifies their exact location, and measures the relative degree of clonal expansion.

Methodology

DNA was prepared from PBMCs from 5 patients, randomly sheared, and linker-mediated PCR was used to selectively amplify both the 5′ and 3′-LTR integration site junctions, whose sequences, as well as those of the sheared breakpoints in the host DNA, were determined by Illumina paired-end sequencing. Clonal expansion of infected cells was demonstrated by the presence of DNA fragments with exactly the same integration site and different host DNA breakpoints.

Results

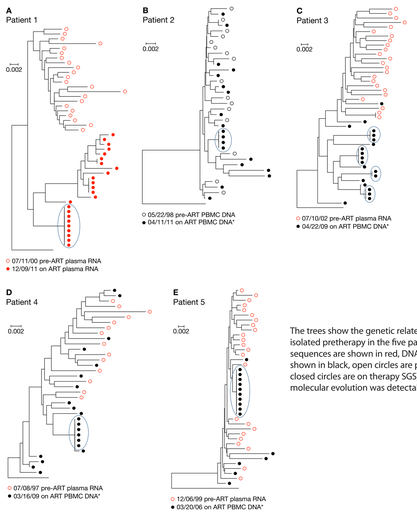
Analysis of integration site libraries comprising >2000 independent integration events revealed clonal expansion of HIV-infected cells in all patients studied. In one patient, about half of the infected PBMCs were derived from a single infected cell. Infected clones persisted in patients for at least 11 years. Integrations in the same orientation in a specific intron of two different genes (MKL2 and BACH2) were seen in independent clones from more than one patient. In one patient, we detected 669 unique sites, of which 14 were in different expanded clones derived from independent integration events in a single 3.5 kb intron of MKL2, and 12 were in a single intron of BACH2. These data show that the expression of these genes must have been affected by these integrations in ways that made a critical contribution to the clonal expansion and/or persistence of the infected cells. Both these genes have been linked to the control of cell growth and human cancers. Other genes associated with cell growth (e.g., PAK2, STAT5B, and CYTH1) in which HIV DNA integration may have contributed to clonal expansion and persistence, were also identified in more than one patient.

Table 1. Patient Demographic Data

Patient	Age at entry (y)	Gender	Race/ Ethnicity	CD4 (cells/μl)	CD4%	HIV RNA (Log_{10} copies/ml plasma)	Current ARV regimen
1*	46.5	M	Black	22	3	5.4	TDF+FTC+RTG
2	24.6	F	Black	243	13	3.9	TDF+FTC+NVP
3	51.6	M	Black	24	2	5.6	TDF+FTC+r/ATZ
4	39.2	M	White	924	39	4.2	ABC+3TC+NVP
5	43.5	M	Black	17	2	5.4	TDF+FTC+NVP

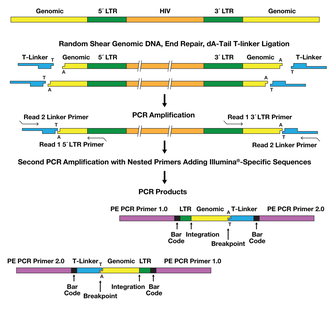
*deceased

Figure 1. A comparison of the pretherapy HIV-1 sequence diversity with the diversity after long-term cART in the five patients.



The trees show the genetic relatedness of virus isolated pretherapy in the five patients (RNA sequences are shown in red, DNA sequences are shown in black, open circles are pre-therapy SGS, closed circles are on therapy SGS). Little or no molecular evolution was detectable during therapy.

Figure 2. Methods used to selectively amplify the host/virus DNA junctions from patient DNA.

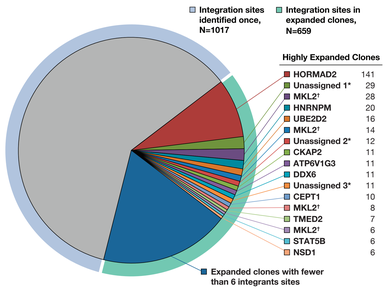


DNA was isolated from cells from the patient's blood and sheared to approximately 300-500 BP fragments (Covaris). The ends of the sheared DNAs were repaired, and a single dA added to the 3′ ends. This allows the selective ligation of a linker that is used to amplify segments that contain the ends of the viral DNA and the adjacent host sequences. Nested primers were used to increase the selective amplification of the desired fragments and to add sequences that were used for Illumina sequencing. Both ends of the amplified segments were sequenced. Differences in the breakpoints of sheared ends of the adjacent host sequences were used to show when the same integration site was present more than once in the original patient DNA.

Table 2. There is extensive clonal expansion of HIV-infected cells in five patients.

Patient	Years on therapy	Distinct integration sites	Number of expanded clones	Percent of expanded clones	Total number of sites in expanded clones	Percent of total sites in expanded clones
1	0.2	254	22	8.7	69	27.2
	4.8	80	11	13.8	27	33.8
	11.4	944	99	10.5	375	39.7
2	12.9	31	2	6.5	4	12.9
3	0	55	2	3.6	4	7.3
	7.2	355	12	3.4	176	49.6
4	12.2	228	19	8.3	83	36.4
5	14.5	125	10	8	24	19.2

Figure 3. HIV-infected cells undergo clonal expansion.

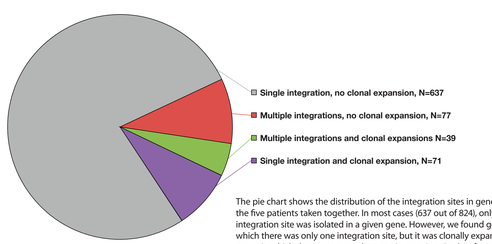


The pie chart shows the fraction of integration sites obtained once and the sites isolated more than once (those that derive from clonally expanded cells). The genes in which the viral DNA is integrated in the most highly amplified clones are shown, as is the relative degree of the amplification, based on how often we recovered the integration sites.

† These are independent integrations into the same intron of MLK2.

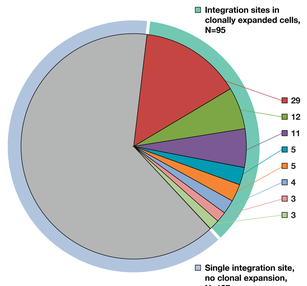
* Unassigned 1, 2, and 3 are distinct sequences.

Figure 4. Genes with integrations in patients.



The pie chart shows the distribution of the integration sites in genes for the five patients taken together. In most cases (637 out of 824), only one integration site was isolated in a given gene. However, we found genes in which there was only one integration site, but it was clonally expanded (71), genes in which there was more than one integration site, but for which we have no evidence of clonal expansion (77), and genes in which there were multiple integrations, at least some of which were clonally expanded (39).

Figure 5. Integrations that are not in any known gene.



Total integration sites that are not in genes	222
Total single integration sites that are not in genes	127
Total integration sites in expanded clones that are not in genes	95
Number of different integrations sites in expanded clones that are not in genes	18

Although most of the HIV integration sites that were identified from the patient samples were in genes, some were not. This pie chart shows what portion of the integration sites that were not in genes that were obtained only once, and what portion was associated with clonal amplification of the infected cells.

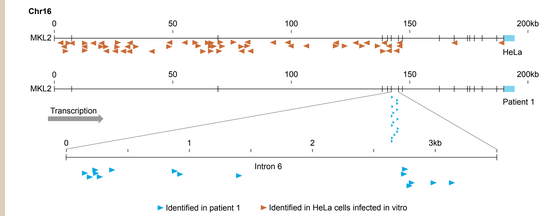
Table 3. Clones persist for >11 years in patient 1.

Gene with integration	Minimum persistence (y)	Gene description
FSIP1	11.2	fibrous sheath interacting protein 1
PARP8	11.2	poly (ADP-ribose) polymerase family, member 8
DDX6	11.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
STAT5B	11.2	signal transducer and activator of transcription 5B
PKP4	11.2	plakophilin 4
CTAGE5	11.2	CTAGE family, member 5 pseudogene; CTAGE family member; CTAGE family, member 4; CTAGE family, member 5
PATL2	11.2	protein associated with topoisomerase II homolog 2 (yeast)
MAP4	6.6	microtubule-associated protein 4
RBM15B	6.6	RNA binding motif protein 15B
NLRC5	6.6	NLR family, CARD domain containing 5
ATP6V1G3	6.6	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G3
PABPC1L	6.6	poly(A) binding protein, cytoplasmic 1-like
TNFSF13	6.6	TNFSF12-TNFSF13 readthrough transcript; tumor necrosis factor (ligand) superfamily, member 12; tumor necrosis factor (ligand) superfamily, member 13
ZC3H4	6.6	zinc finger CCCH-type containing 4
NAIF1	6.6	nuclear apoptosis inducing factor 1
ZCCHC11	4.8	zinc finger, CCHC domain containing 11
ZNF16	4.8	zinc finger protein 16

Table 4. Genes with ≥3 HIV integrations that are associated with clonal expansion.

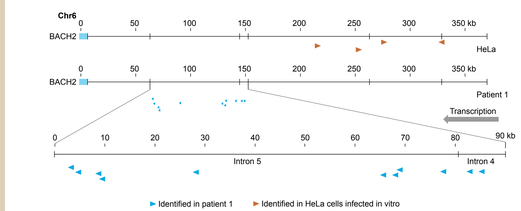
Gene	Patients with integrants	Distinct intragenic integrations (N)	Total integration sites per gene (N)	Gene designation
MKL2	1, 4	17	80	MKL/myocardin-like protein 2 Transcriptional coactivator of serum response factor
BACH2	1, 3	13	17	BTB and CNC homology 1, basic leucine zipper transcription factor 2
STAT5B	1, 3, 4, 5	6	13	Signal transducer and activator of transcription 5B
NFATC3	1, 4, 5	5	9	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3
PAK2	1	4	8	p21 protein (Cdc42/Rac)-activated kinase 2
ATM	1, 3	4	5	Ataxia telangiectasia mutated Serine Theronine Kinase
HORMAD2	1, 3	3	143	HORMA domain containing 2
DDX6	1, 3	3	13	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6
NSD1	1, 5	3	9	Nuclear receptor binding SET domain protein 1
MKL1	1	3	7	Megakaryoblastic leukemia (translocation) 1 transcription interacting protein
USP4	4	3	6	Ubiquitin-specific peptidase 4 (proto-oncogene)
BTBD9	1	3	5	BTB (POZ) domain containing 9
TAOK1	1	3	5	TAO kinase 1

Figure 6. Integrations in MKL2 in patient 1.



The figure shows a map of the MKL2 gene and compares the integration sites obtained in the analysis of a large library of integration sites from HeLa cells infected in vitro (~248,000 independent integration sites) with the integration sites obtained from an analysis of the integration sites in patient 1. Note that, while the in vitro library has integrations scattered throughout the gene, in the patient, all of the integration sites are in the same intron, and all are in the same orientation.

Figure 7. Integrations in BACH2 in patient 1.



The overall form of this figure is similar to Figure 6. It shows the integration sites in BACH2 in both the HeLa library and patient 1. Again, all of the integrants in the patient are in the same intron, and in the same orientation. The HeLa library has relatively few integrants in BACH2, presumably because BACH2 is not expressed in HeLa cells. We analyzed a large library of HIV integration sites in CD34+ cells. There are a number of integrants throughout the gene, in both orientations in that library (data not shown).

MKL2 and BACH2 are involved in human cancer.

MKL2

- *The transcriptional coactivators megakaryoblastic leukemia 1/2 mediate the effects of loss of the tumor suppressor deleted in liver cancer 1.* Muehlich S., Hampl V., Khalid S., Singer S., Frank N., Breuhahn K., Gudermann T., Prywes R. **Oncogene**. 2012; 31(35): 3913-23
- *Megakaryoblastic leukemia-1/2, a transcriptional co-activator of serum response factor, is required for skeletal myogenic differentiation.* Selvaraj A., Prywes R. **J. Biol. Chem**. 2003; 278(43): 41977-87.
- *Presence of C11orf95-MKL2 fusion is a consistent finding in chondroid lipomas: a study of eight cases.* Flucke U., Tops B.B., de Saint Aubain Somerhausen N., Bras J., Creytens D.H., Küsters B., Groenen P.J., Verdijk M.A., Suurmeijer A.J., Mentzel T. **Histopathology**. 2013; 62(6): 925-30.

BACH2

- *A BACH2 link between pre-B cell receptor checkpoint and pre-B cell ALL.* Ye BH, Mai Y. **Cancer Cell**. 2013; 24(3): 282-4.
- *Identification of IGHCδ-BACH2 fusion transcripts resulting from cryptic chromosomal rearrangements of 14q32 with 6q15 in aggressive B-cell lymphoma/leukemia.* Kobayashi S., Taki T., Chinen Y., Tsutsumi Y., Ohshiro M., Kobayashi T., Matsumoto Y., Kuroda J., Horiike S., Nishida K, Taniwaki M. **Genes Chromosomes Cancer**. 2011; 50(4): 207-16.

Conclusions

- There is extensive clonal expansion of HIV-infected cells in each of the first 5 patients we analyzed; in one patient a single clone represented approximately half the infected cells in the blood.
- Clonally expanded cells can persist in patients for >11 years.
- In a substantial number of instances, the integration site plays an important role in the clonal expansion and/or persistence of the clone.

Prepared by Scientific Publications, Graphics & Media, Frederick National Laboratory - 20180