Only a Few HIV-1 Integration Sites Confer Growth Advantage to Infected Cells *in Vivo*

0363

John M. Coffin[1], Michael J. Bale[2], Daria Wells[3], Shuang Guo[3], Brian Luke[3] Jennifer Zerbato[4,5], Michelle Sobelewski[4], Twan Sia[2,6], Wei Shao[3], Xiao-Lin Wu[3], Frank Maldarelli[2], Mary F. Kearney[2], John W. Mellors[4], Stephen H. Hughes[2]

[1]Tufts University, Boston MA, [2]National Cancer Institute, Frederick MD, [3]Leidos Biomedical Inc., Frederick MD, [4]University of Pittsburgh, Pittsburgh PA,
[5]Present address: The Peter Doherty Institute for Infection and Immunity, Melbourne, Australia Present Address: Swarthmore College, Swarthmore PA

BACKGROUND

HIV persists during antiretroviral therapy (ART) because of the survival of infected cells that are descendants of a tiny fraction of the CD4+T cells infected prior to ART initiation. We and others previously reported that there is selection *in vivo* for clones of cells with proviruses integrated in several specific genes, based on analyzing small numbers of integration sites.

METHODS

We compared about 380,000 integration sites in PBMC infected *ex vivo* to about 52,000 integration sites obtained from 32 individuals on suppressive ART for >1 year, of which about 31,000 were unique. 16,000 (13,000 unique) sites were obtained from 28 pretherapy donors. The datasets were compared to look for evidence of selection *in vivo*, and to infer the underlying mechanisms.

Table1. Datasets Analyzed for this study

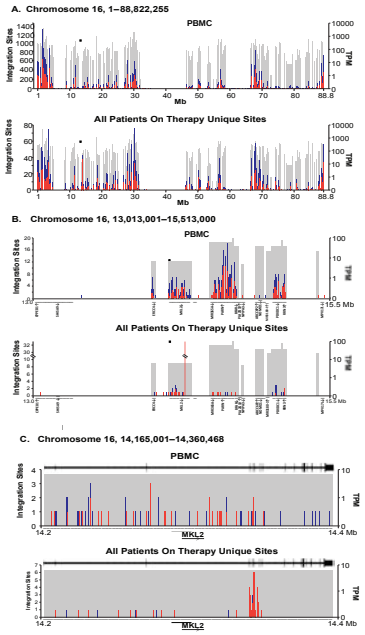
Dataset	Length (Mb) or Number of sites (% of total)		
	Total	In Genes	In Expressed Genes
Genome	3080 Mb	1171 Mb (38%)	688 Mb (22%)
Ex vivo infected PBMC	384,551	323529 (84%)	318,597 (83%)
Pretherapy Unique Sites	13,324	10,540 (79%)	10,173 (76%)
On ART, Unique Sites	33,336	25,804 (77%)	24,816 (74%)

What determines the distribution of HIV proviruses in people on ART?

- Initial distribution?
- Selection for/against proviruses in genes?
- Clonal amplification?

Table 2. Genes showing selection of provirus integration on ART

Gene Name	Unique IS in Genes	IS in PBMC	Enrichment Probability	Provirus Orientation: with/against	Orientation Probability
All Genes	25,804	323,587		10426/12774	$3.0X10^{-50}$
STAT5B	268	562	$<1X10^{-50}$	197/71	$2.2X10^{-15}$
BACH2	98	132	$1.2X10^{-43}$	71/20	$2.6X10^{-8}$
MKL2	49	69	$7.9X10^{-16}$	40/6	$1.3X10^{-7}$
MKL1	85	331	$1X10^{-20}$	53/30	$3.6X10^{-3}$
IL2RB	30	68	$7.9X10^{-7}$	17/9	$4.7X10^{-2}$
MYB	11	31	$6X10^{-4}$	10/0	$9.8X10^{-4}$



Multiscale Distribution of Integration Sites

Mean expression
IS per bin orientation same as chromosome
IS per bin opposite of chromosome
Region enlarged in next panel

Figure 2. Distribution of integration sites. This figure shows the distribution of HIV integration sites in the indicated chromosome regions of chromosome 16 divided into 250 bins. Genes, with their name and transcriptional orientation, are shown on the X-axis in B and C. Their extent and relative expression (in TPM) are shown by by the grey boxes. The numbers of integration sites in each bin are shown by the vertical bars. The analysis was run on at 3 different bin sizes in panels A, B and C. **A.** Whole chromosome (ca 88 Mb; 352 kb/bin). **B.** A 2 Mb segment (8 kb/bin), centered on the MKL2 gene. **C.** The MKL2 gene (ca 200 kb; 800 bp/bin).

RESULTS

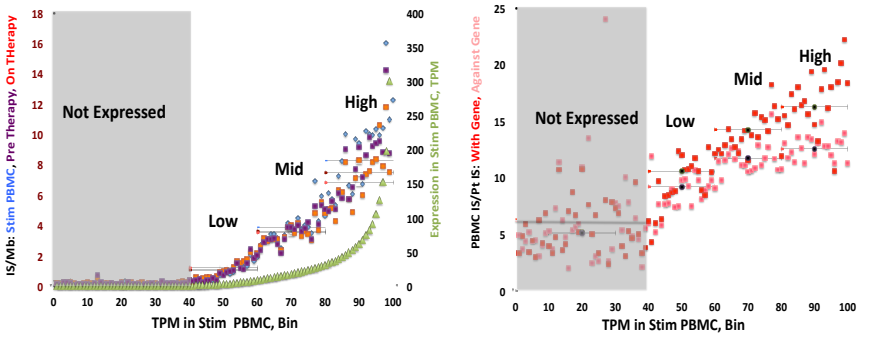


Figure 1. Proviral site selection and counterselection are related to gene expression. Stimulated PBMC were HIV-infected were used for RNA-seq analysis, and all genes ranked by expression in transcripts per million (TPM, green). HIV integration sites in genes in the PBMC (blue), from pre-therapy HIV-infected donors (plum) or on therapy donors (red) were divided into 100 bins with ca. 200 genes in each bin. In both panels, unique integration site data are plotted against the ranked TPM. Horizontal bars show the same data divided into 3 expression levels (low, >0-10 TPM; mid,>10-30 TPM; high, >30 TPM) **A.** Integration data for each bin are presented as density (unique sites/kb). **B.** Integration site densities, plotted, as the ratio, for each bin, of *ex vivo* integration sites to donor PBMC integration sites for the on ART samples, separated by the orientation of the provirus relative to the host gene. Light colors indicate proviruses opposite the orientation of the gene, darker colors in the same orientation.

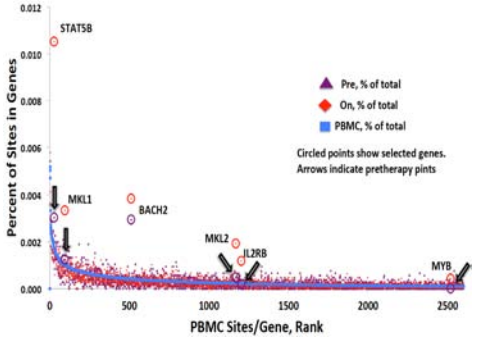


Figure 3. Comparison of integration frequency in genes that are good targets for HIV. The percentage of unique sites in individual genes is plotted, for each dataset, using the distribution in *ex vivo*-infected PBMC, ranked from highest to lowest. The circles show the data for the six selected genes in table 2.

CONCLUSIONS

The primary determinant of the distribution of proviruses in people on ART is the initial distribution, which is subsequently modified by selection against proviruses in expressed genes; more strongly for proviruses in the same orientation as the gene. Proviruses integrated in the sense integration in any one of only 6 genes can enhance cell expansion and/or survival; however, this percentage comprises only a small fraction of the cells that expand and are unlikely to be of major importance to HIV-1 persistence. Other mechanisms driving clonal expansion, for example immune signaling, are more important.

ACKNOWLEDGMENTS

We thank Connie Kinna, Anne Arthur, Valerie Turnquist, Terri Burdette, and Susan Toms for administrative support. We are grateful to Allen Kane for help with the figures.

Funding

This project has been funded with federal funds from the National Cancer Institute, in part by contract HHSN261200800001E to Leidos Biomedical, Inc. JMC was a Research Professor of the American Cancer Society and was supported by Subcontract I3XS110 and JWM was supported by Subcontract 12XS547 to the University of Pittsburgh from Leidos Biomedical research, Inc.

HIV Dynamics and Replication Program
National Cancer Institute at Frederick

NIH NATIONAL CANCER INSTITUTE

Tufts UNIVERSITY