JOURNAL OF VIROLOGY, Oct. 2005, p. 12199–12204 Vol. 79, No. 19
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.19.12199–12204.2005

Simian Immunodeficiency Virus Integration Preference Is Similar to That of Human Immunodeficiency Virus Type 1

Bruce Crise,[1] Yuan Li,[1] Chiuchin Yuan,[1] David R. Morcock,[1] Denise Whitby,[1] David J. Munroe,[2] Larry O. Arthur,[1] and Xiaolin Wu[2]*

AIDS Vaccine Program[1] and Laboratory of Molecular Technology,[2] Scientific Application International Corporation, National Cancer Institute at Frederick, Frederick, Maryland 21702

Received 16 May 2005/Accepted 5 July 2005

Simian immunodeficiency virus (SIV) is a useful model for studying human immunodeficiency virus (HIV) pathogenesis and vaccine efficacy. As with all other retroviruses, integration is a necessary step in the replication cycle of SIV. The location of the retrovirus integration site is known to impact on viral gene expression, establishment of viral latency, and other aspects of the replication cycle of a retrovirus. In this study, 148 SIV provirus integration sites were sequenced and mapped in the human genome. Our analysis showed that SIV integration, like that of HIV type 1 (HIV-1), exhibited a strong preference for actively transcribed regions in the genome (A. R. Schroder et al., Cell 110:521–529, 2002) and no preference for the CpG islands or transcription start sites, in contrast to observations for murine leukemia virus (X. Wu et al., Science 300:1749–1751, 2003). The parallel integration target site preferences of SIV and HIV-1 suggest that these lentiviruses may share similar mechanisms for target site selection and that SIV serves as an accurate model of HIV-1 with respect to integration.

All retroviruses require integration into the host cell genome for completion of the replication cycle of the virus. The integration process is catalyzed by virally encoded integrase, and much has been learned about the biochemical steps involved (6). Integrase first catalyzes the removal of a dinucleotide from the 3′ ends of viral termini and then joins the processed viral ends to the target DNA in a concerted cleavage and ligation reaction. However, the process through which integration sites are selected by retroviruses remains poorly understood. Previous studies have shown that most of the host genome is accessible for retrovirus integration, but the target site selection is not totally random (6, 14, 29, 30, 41). Many large-scale surveys of genome-wide retrovirus integration sites have been reported recently (23, 26, 35, 43). One of the surprising findings is that retroviruses from diverse genera have different target site preferences even though their integrases share very similar biochemistry. Schroder et al. showed that human immunodeficiency virus type 1 (HIV-1) greatly prefers integrating into transcription units within the host genome (35). In contrast, Wu et al. showed that murine leukemia virus (MLV) prefers transcription start sites or CpG island regions (43). Yet another virus, avian sarcoma-leukosis virus (ASLV), has much weaker preferences for any of these locations (23, 26).

Integration has broad-ranging consequences for the host, including establishment of persistent or latent infections (16) and activation of oncogenes (15). In the case of HIV-1, viral latency is a significant hurdle to effective and durable treatment of acquired immunodeficiency (27). Simian immunodeficiency virus (SIV) infection of macaques serves as a model of AIDS that is amenable to a variety of experimental parameters while having many of the hallmarks of HIV-1 infection and pathogenesis, including the establishment of latently infected populations of cells (7, 12). Recent publications highlight the unique integration profile of HIV-1 relative to murine and avian retroviruses (2, 26, 43), and it was of interest to determine if the preference for actively transcribed regions of the genome is unique to HIV-1 or is common to other lentiviruses. Hematti et al. reported that SIV-based gene therapy vectors strongly favor transcription units and gene-dense regions in the rhesus monkey genome (11). However the study was focused on the integration sites in circulating blood cells derived from SIV-transduced $CD34^+$ hematopoietic stem cells after long-term transplantation (6 months), which may exert selective pressure on subpopulations of the transduced stem cells. The study presented here characterized the integration sites of SIV in human tissue culture cells by acute infection with a replication-competent SIV clone. The acute infection should have minimal selective effects for any subpopulation of integration sites and should thus represent the true integration preference of SIV. We found that SIV shared very similar target preferences with HIV-1, which prefers to integrate into actively transcribed genes.

MATERIALS AND METHODS

Cloning of SIV integration junction site sequences. To study the target site selection of SIV integration, the human lymphoid cell line CEMx174 (31) was infected with SIV. Virus was produced by transfecting 293 cells with a proviral clone encoding SIV(mne) clone 8 (18). Cell-free transfection supernatant was used to infect CEMx174 cells, which were subsequently cultured to establish a chronically infected culture. At the time of DNA extraction, approximately 60% of cells were infected, as determined by flow cytometry using intracellular internal Gag staining. Integration junction site sequences were amplified and cloned by the linker-mediated PCR method as described by Wu et al. (43). Briefly, genomic DNA was extracted using a DNAeasy kit (QIAGEN, CA) and then digested overnight with the restriction enzyme MseI. Digestion of human genomic DNA with this enzyme is predicted to generate DNA fragments of

* Corresponding author. Mailing address: Laboratory of Molecular Technology, Scientific Application International Corporation—Frederick, National Cancer Institute at Frederick, 915 Toll House Ave., Frederick, MD 21701. Phone: (301) 846-7677. Fax: (301) 846-6100. E-mail: forestwu@mail.nih.gov.

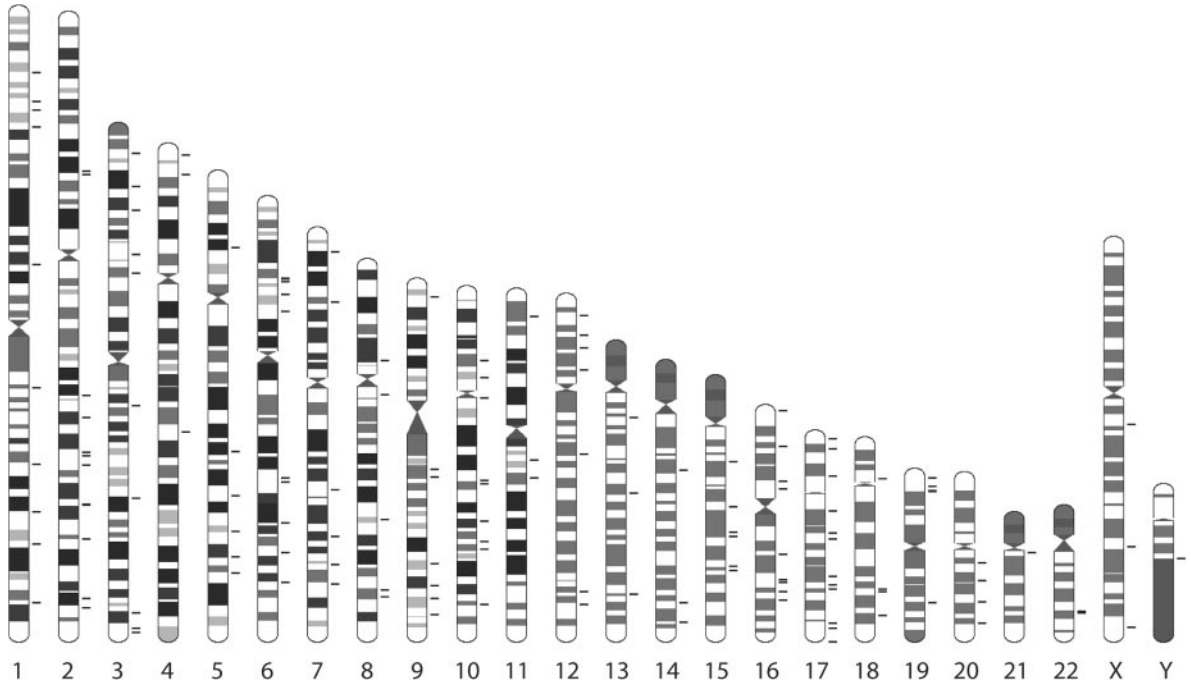

FIG. 1. Integration site location and distribution on human chromosomes. Sites of SIV integration are marked to the right of the chromosomes with cytogenetic bands.

approximately 70 base pairs on average. The digested DNA was ligated to a double-stranded linker of GGATTTGCTGGTGCAGTACAGGCCTTAAGAG GGAC and *p*-TAGTCCCTCTTAAGGCCT$_{C7}$. The ligated DNA was digested with HincII to reduce the number of linkers ligated to the internal proviral sequence. The first round of PCR amplification was carried out with provirus- and linker-specific primers CCTCTTCAATAAAGCTGCCTT and GGATTTG CTGGTGCAGTACAG, respectively. A second round of PCR was then performed with provirus- and linker-specific primers AAGTAAGCCAGTGTGTG CTCCCATC and AGTACAGGCCTTAAGAGGGA, respectively. The amplified PCR product was cloned into pCR2.1-TOPO, followed by transformation of *Escherichia coli* Top10f. Clones were sequenced on an ABI 3700 sequencer.

Mapping of SIV integration sites in the human genome. SIV integration sites were mapped to the human genome as described previously (43). Raw sequences were trimmed with a customized computer program to remove all vector elements, the remaining portion of the sequence was reviewed, and only sequences with high-quality scores were kept. The junction sequence next to the 3′ long terminal repeat (LTR) end sequence was compared to the human genome by using the BLAT program on the UCSC website (http://genome.ucsc.edu/). Authentic integration sites used in the analysis were defined as (i) containing sequence contiguous from the nested primer to the end of the 3′ LTR (CA) and the linker sequence; (ii) matching a genomic location, starting immediately (within 3 bases) after the end of the 3′ LTR; (iii) showing 95% or higher identity to the genomic sequence over the high-quality sequence region; and (iv) matching no more than one genomic locus with 95% or higher identity. A total of 148 unique SIV integration sites were mapped within the human genome (Hg16, July 2003 freeze; http://genome.ucsc.edu/) and used for the analysis in this study.

Microarray analysis of gene expression. To study the level of mRNA expression in host cells via microarray analysis, CEMx174 cells were either infected with SIV(mne) clone 8 or left uninfected. Cells were cultured until approximately 60% of the cells exposed to SIV were infected as judged by internal Gag staining, at which time total cellular RNA was isolated using an RNeasy kit (QIAGEN, CA). The RNA was labeled and hybridized to Affymetrix Human Genome U133 plus 2.0 arrays according to the manufacturer's protocol. Probe intensity was determined with the Affymetrix GCOS 1.1 software. All arrays were normalized to a median signal level of 150. RNA isolation and hybridization were done in triplicate with infected and parallel uninfected cultures.

Nucleotide sequence accession numbers. Nucleotide sequences of junction sites were deposited in GenBank (accession numbers AY679815 to AY680027).

RESULTS

Cloning and mapping of SIV integration sites in the human genome. To study the target site selection of SIV integration, the human lymphoid cell line CEMx174 (31) was infected with SIV(mne) clone 8. At the time of genomic DNA extraction, 60% of the cells were infected as judged by intracellular Gag staining. Integration junction site sequences were amplified with linker-mediated PCR and cloned as described previously (43). Junction site sequences were mapped to the human genome with the BLAT program using the stringent criteria for authentic integration sites as previously described (43). A total of 148 unique SIV integration sites were mapped within the human genome (Hg16, July 2003 freeze; http://genome.ucsc.edu/). The chromosomal locations of the integration sites are shown in Fig. 1. Integration sites were observed on every chromosome.

SIV greatly prefers genes as integration targets in the human genome. As was proposed early in the research on retrovirus integration, actively transcribed regions are favored targets for retrovirus integration (32, 43). Of the 148 unique SIV integration sites we mapped, 74% (110/148) were located within RefSeq genes, which are a set of well-annotated genes based on mRNA records (20). Computer simulation was used to evaluate whether this is statistically different from random integration. To do this, we generated 148 integration sites randomly in the human genome and then calculated the percentage of random integration sites within RefSeq genes. This process was performed 1,000 times. Figure 2 shows a comparison of values determined for the 1,000 random sets and the mapped SIV integrations. The mean frequency of random integration sites within RefSeq genes was 34%, similar to the

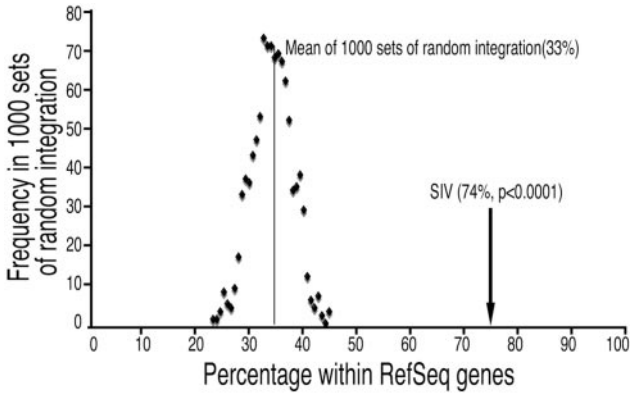

FIG. 2. RefSeq genes are preferred as integration targets by SIV. The percentage of SIV integrations in RefSeq genes is shown as a black arrow. The distribution of integrations in RefSeq genes for 1,000 sets (148 each) of computer-simulated random integrations is shown. The mean value of integration sites in RefSeq genes from random integration is 33%, and the value for SIV is 74%.

estimated RefSeq content of the human genome (38). This rate of random integration is significantly less than the 74% observed for integration of SIV into RefSeq genes in CEMx174 cells ($P < 0.0001$ by t test). We further analyzed the orientations of the SIV proviruses in relation to the directions of the genes in which they resided. We did not observe any orientation preference for SIV proviruses that landed inside genes.

SIV shows no preference for CpG islands or transcription start sites of genes. MLV favors integration at or near promoter regions of genes as defined by CpG islands or transcription start sites of genes (43). In contrast, HIV-1 shows no preference for CpG islands. To investigate whether SIV integration was similar to that of either MLV or HIV, we analyzed the integration sites of SIV for their proximity to CpG islands and transcription start sites in conjunction with the Wu et al. data set (43) and the Schroder et al. data set (35). Only 9% (13/148) of SIV integration sites landed within the ±5-kb region of CpG islands documented in the human genome. This was not statistically different from random integration (8%) or the HIV-1 (12%; 39/334) in vivo integrations from the Schroder et al. data set. However, this value was statistically lower than that for MLV, for which 28% of the integration sites landed in the same ±5-kb window (43). In corroboration of this, we found that only 4% (6/148) of SIV integration sites landed within the ±5-kb region of transcription start sites of RefSeq genes, compared to 10% for HIV-1 and 24% for MLV (43) (Table 1).

SIV targets actively transcribed genes. It has been shown that HIV-1 and MLV prefer to target more actively transcribed genes (10, 23, 32, 35, 43). To investigate if this is the case for SIV integration, we generated gene expression profiles of the CEMx174 cells used for mapping SIV integrations. RNA isolations and microarray analyses of both infected and uninfected cells were conducted three independent times. Although not all target genes are expressed, as judged by an absent/present call with the Affymetrix GCOS software, an average of 80% of the target genes were "present" in CEMx174 cells in all six array experiments. This is statistically

TABLE 1. Frequency of integration sites in transcription units and transcription start sites

Location	Frequency of integration sites (%) at the indicated location			
	SIV (n = 146)	HIV-1[a] (n = 334)	MLV[b] (n = 902)	Random sites (n = 10,000)
In RefSeq genes	74	72	41	33
In CpG islands ± 5 kb	9	12	28	8
In transcription start sites of RefSeq genes ± 5 kb	4	10	24	3

[a] Data set from HIV-1 in vivo integration sites in SupT1 cells reanalyzed in human genome freeze Hg16 (35).
[b] Data set from MLV integration sites in HeLa cells reanalyzed in human genome freeze Hg16 (43).

higher than the 50% "present" rate for target genes of 10,000 in silico random sites ($P < 0.001$). In a comparison of probes representing the targeted genes to all probes on the gene chips (both infected and uninfected) (Fig. 3A), we found that the median expression level of SIV-targeted genes was fivefold higher (34 to 43 for all genes and 217 to 242 for SIV-targeted genes on all six arrays; $P < 0.0001$ by a Mann-Whitney test). In another analysis approach, all the genes were assigned to 1 of 10 "bins" based on expression level, with each bin containing an equal number of genes. The number of SIV-targeted genes within each bin was counted and plotted (Fig. 3B). The analysis was carried out separately for each of the Affymetrix chips (three for infected and three for uninfected cells). Again, genes targeted by SIV integration were shown to be more highly expressed. Interestingly, although SIV integration preferred actively transcribed genes, there was a decrease in the number of integrations found in the bin with the highest expression level (Fig. 3B, bins 10 and 9, respectively). This is very similar to what has been reported for HIV-1, which showed reduced integration in the most highly expressed category of genes analyzed (23).

Further characterization of the location of SIV integrations. HIV-1 integration sites in vitro exhibit a tendency for clustering, suggesting that hot spots for integration exist (35). The highest density of SIV integrations that we observed, three independent integration sites within a 575-kb stretch of genomic DNA, was considerably lower than that observed for HIV-1 integration in SupT1 cells. In contrast to our results with SIV, HIV-1 integration into one hot spot consisting of five independent integration sites within 2.4 kb was observed, while five other hot spots with three independent integration sites within 100 kb were identified (35). In studies of rhesus macaque $CD34^+$ hematopoietic stem cells transduced with SIV-based vectors and reintroduced in vivo, no hot spots for SIV integration were observed (11). This is consistent with our data, although the small SIV data set may limit our ability to detect hot spots. Alternatively, integration hot spots may be cell line specific, since no hot spots were observed for HIV-1 in HeLa cells (43).

There have been conflicting reports about HIV-1 integration into human repeated sequences, with some data indicating that Alu and LINE elements are favored by HIV-1 integration (36, 37). Our data showed that 63 out of 148 (43%) SIV integrations landed in repeat sequences within the human genome,

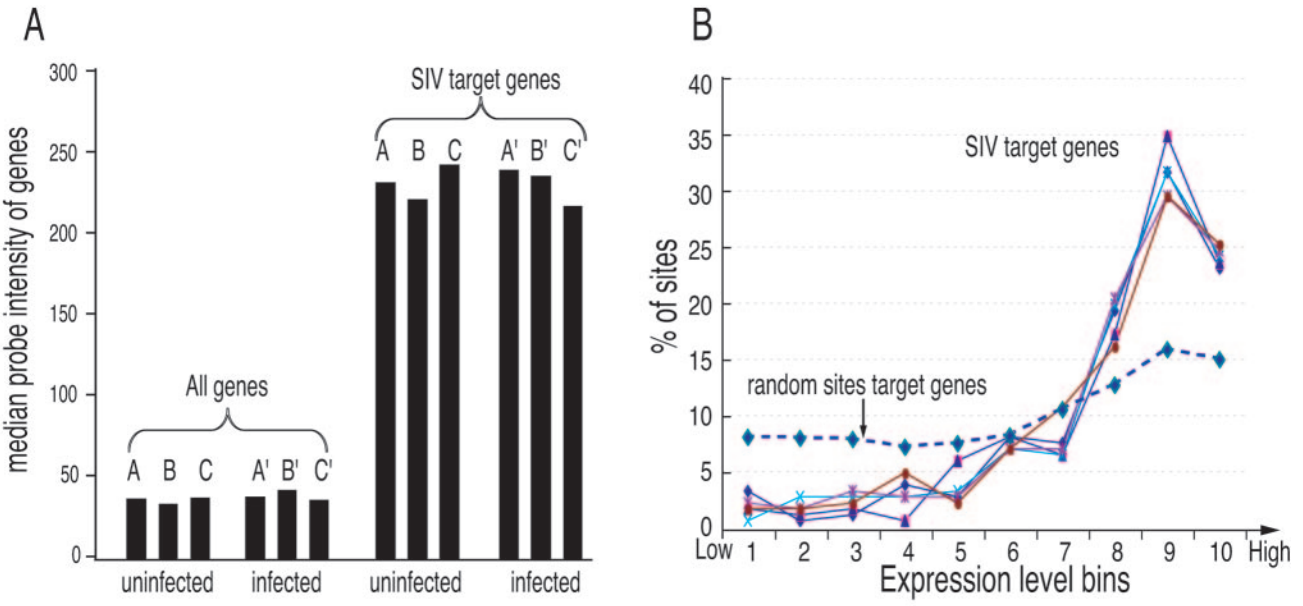

FIG. 3. Targeted genes are more actively expressed. (A) Targeted gene expression compared to expression of all genes on the chip. The median signal intensities of probes for all genes on the Affymetrix Hu133 chip are statistically lower than the median signal intensities of targeted genes (infected and uninfected, tested in triplicate; $P < 0.0001$ by t test). (B) Target gene expression levels by bin analysis. All genes on the chip were evenly divided into 10 bins based on expression levels; thus, each bin has 10% of total genes. The genes targeted by SIV and 10,000 in silico random integration sites were also placed in these 10 bins based on expression values. Lines of different colors represent results from each individual microarray experiment. Targeted genes were found more often in the high-expression-level bins than in the low-expression-level bins ($P < 0.0001$ by a chi-square test). However, the bin with the highest expression level does not hold most of the targeted genes.

statistically no different from the randomly generated integration site frequency of 48% (10,000 random integration sites; data not shown). Due to the limited number of integrations in the SIV data set, it was not feasible to extend our analysis to individual classes of repeated sequences.

DISCUSSION

Viral components of preintegration complexes (1, 4, 8, 22), including integrase, nucleocapsid protein (4, 28), and the attachment sites of the reverse-transcribed genome, contribute to the ability of reverse-transcribed DNA to integrate into the target genome. Additionally, chromatin structure has been suggested to influence integration (13, 24, 25, 30, 44). Cellular cofactors may also play an important role in retrovirus target site selection (2, 42). Viral proteins and the attachment sites of SIV and HIV-1 have a high degree of homology relative to other retroviruses, such as MLV. The genomes of natural hosts of SIV and HIV are very similar. Our hypothesis is that SIV and HIV will exhibit similar preferences for target site selection. Our results showed that SIV strongly favors transcribed regions as targets in the human genome, with 74% in the RefSeq genes. In comparison, 72% of the integration sites of HIV-1 from cultured lymphoid cells (43) were within RefSeq genes when reanalyzed in the same genome freeze (Hg16). The similar rates of SIV and HIV-1 integration into RefSeq genes indicate that these viruses may share similar mechanisms including cellular cofactors for target site selection at the global level.

The data presented here reinforce earlier findings that host gene transcription levels influence target site selection. Multiple modes of analysis indicated that integration was more likely in genes that were highly expressed (Fig. 3). This is consistent with the findings of site distribution studies of MLV (43) and HIV-1 (10, 23, 35). However, the relationship between integration frequency and transcription did not extend to the most highly expressed genes. Our findings corroborate those of Mitchell et al., who showed that while HIV-1 prefers active genes, the most highly expressed genes are not those most favored for integration (23). A possible explanation is that very strong transcription may inhibit integration, as reported for ASLV: the frequency of proviral integration into a reporter gene was lower when transcription was strongly induced (40). Another possible explanation is that very active genes are typically smaller than poorly expressed genes (5), and thus the target region of these genes becomes smaller, as proposed in the case of HIV-1 (17). Increased transcription of the genes targeted by integration complexes may reflect the global state of chromatin in gene-rich regions (38, 39) and not the level of transcription at the time when integration takes place.

Recent data (13, 21, 44) show that retroviral integration prefers weak palindromic consensus sequence elements at the target sites. SIV and HIV-1 target sites share similar palindromic sequences, providing additional evidence that comparable mechanisms of integration are used.

Schroder et al. reported that the ranking of HIV-1 target gene expression increased in infected cells versus uninfected cells (35). We did not observe increased transcription of SIV target genes in three independent infection experiments. These contrasting results may reflect differences in experimental design: Schroder et al. employed a short-term culture of

cells infected with a lentivirus vector, while the approach used here was to generate cultures chronically infected with an infectious virus.

During the course of our work, Hematti et al. reported that a replication-defective SIV vector virus (SIVmac1A11) exhibited similar integration preferences for transcribed regions in primate hematopoietic stem and progenitor cells (11). As reported recently in many cases (3, 9, 19, 33, 34), peripheral blood cells repopulated in gene therapy patients by using retrovirus vectors may represent clonal expansion of certain subpopulations of the transplanted hematopoietic stem cells. Although Hematti et al. mapped integrations within in vivo-selected peripheral blood mononuclear cells derived from $CD34^+$ lymphocytes, whereas our study used an acutely infected $CD4^+$ lymphoid line, both analyses showed SIV integration to have a strong preference for transcription units. This study expands the similarity we observed for SIV and HIV-1 integration in the human genome to the rhesus macaque genome, suggesting that the preference for genes by SIV integration is a cross-species phenomenon.

ACKNOWLEDGMENTS

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400 (article H.36 of the Prime Contract).

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.

We thank Vicky Coalter for expert technical assistance in conducting flow cytometric assays.

REFERENCES

1. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3:**469–478.
2. **Bushman, F. D.** 2003. Targeting survival: integration site selection by retroviruses and LTR-retrotransposons. Cell **115:**135–138.
3. **Calmels, B., C. Ferguson, M. O. Laukkanen, R. Adler, M. Faulhaber, H. J. Kim, S. Sellers, P. Hematti, M. Schmidt, C. von Kalle, K. Akagi, R. E. Donahue, and C. E. Dunbar.** 2 June 2005. Recurrent retroviral vector integration at the MDS1-EVI1 locus in non-human primate hematopoietic cells. Blood 10.1182/blood-2005-03-1115.
4. **Carteau, S., R. J. Gorelick, and F. D. Bushman.** 1999. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. J. Virol. **73:**6670–6679.
5. **Castillo-Davis, C. I., S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov.** 2002. Selection for short introns in highly expressed genes. Nat. Genet. **31:**415–418.
6. **Coffin, J. M., S. H. Hughes, and H. E. Varmus.** 1997. Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
7. **Desrosiers, R. C.** 1990. The simian immunodeficiency viruses. Annu. Rev. Immunol. **8:**557–578.
8. **Fujiwara, T., and R. Craigie.** 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. USA **86:**3065–3069.
9. **Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer, and M. Cavazzana-Calvo.** 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science **302:**415–419.
10. **Han, Y., K. Lassen, D. Monie, A. R. Sedaghat, S. Shimoji, X. Liu, T. C. Pierson, J. B. Margolick, R. F. Siliciano, and J. D. Siliciano.** 2004. 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.
11. **Hematti, P., B. K. Hong, C. Ferguson, R. Adler, H. Hanawa, S. Sellers, I. E. Holt, C. E. Eckfeldt, Y. Sharma, M. Schmidt, C. von Kalle, D. A. Persons, E. M. Billings, C. M. Verfaillie, A. W. Nienhuis, T. G. Wolfsberg, C. E. Dunbar, and B. Calmels.** 2004. Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol. **2:**e423.
12. **Hirsch, V. M., and J. D. Lifson.** 2000. Simian immunodeficiency virus infection of monkeys as a model system for the study of AIDS pathogenesis, treatment, and prevention. Adv. Pharmacol. **49:**437–477.
13. **Holman, A. G., and J. M. Coffin.** 2005. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. Proc. Natl. Acad. Sci. USA **102:**6103–6107.
14. **Kitamura, Y., Y. M. Lee, and J. M. Coffin.** 1992. Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. Proc. Natl. Acad. Sci. USA **89:**5532–5536.
15. **Kung, H. J., and P. K. Vogt.** 1991. Current topics in microbiology and immunology, vol. 171. Retroviral insertion and oncogene activation. Springer-Verlag, Heidelberg, Germany.
16. **Lassen, K., Y. Han, Y. Zhou, J. Siliciano, and R. F. Siliciano.** 2004. The multifactorial nature of HIV-1 latency. Trends Mol. Med. **10:**525–531.
17. **Lewinski, M. K., D. Bisgrove, P. Shinn, H. Chen, C. Hoffmann, S. Hannenhalli, E. Verdin, C. C. Berry, J. R. Ecker, and F. D. Bushman.** 2005. Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J. Virol. **79:**6610–6619.
18. **Li, Y., W. J. Bosche, G. F. Heidecker, R. E. Benveniste, R. J. Gorelick, and B. J. Crise.** 2004. Complete plasmid sequence of SIV(Mne) clone 8. GenBank accession number AY817672.
19. **Li, Z., J. Dullmann, B. Schiedlmeier, M. Schmidt, C. von Kalle, J. Meyer, M. Forster, C. Stocking, A. Wahlers, O. Frank, W. Ostertag, K. Kuhlcke, H. G. Eckert, B. Fehse, and C. Baum.** 2002. Murine leukemia induced by retroviral gene marking. Science **296:**497.
20. **Maglott, D. R., K. S. Katz, H. Sicotte, and K. D. Pruitt.** 2000. NCBI's LocusLink and RefSeq. Nucleic Acids Res. **28:**126–128.
21. **Maxfield, L. F., C. D. Fraize, and J. M. Coffin.** 2005. Relationship between retroviral DNA-integration-site selection and host cell transcription. Proc. Natl. Acad. Sci. USA **102:**1436–1441.
22. **Miller, M. D., C. M. Farnet, and F. D. Bushman.** 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J. Virol. **71:**5382–5390.
23. **Mitchell, R. S., B. F. Beitzel, A. R. Schroder, P. Shinn, H. Chen, C. C. Berry, J. R. Ecker, and F. D. Bushman.** 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. **2:**E234.
24. **Muller, H. P., P. M. Pryciak, and H. E. Varmus.** 1993. Retroviral integration machinery as a probe for DNA structure and associated proteins. Cold Spring Harb. Symp. Quant. Biol. **58:**533–541.
25. **Muller, H. P., and H. E. Varmus.** 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:**4704–4714.
26. **Narezkina, A., K. D. Taganov, S. Litwin, R. Stoyanova, J. Hayashi, C. Seeger, A. M. Skalka, and R. A. Katz.** 2004. Genome-wide analyses of avian sarcoma virus integration sites. J. Virol. **78:**11656–11663.
27. **Persaud, D., Y. Zhou, J. M. Siliciano, and R. F. Siliciano.** 2003. Latency in human immunodeficiency virus type 1 infection: no easy answers. J. Virol. **77:**1659–1665.
28. **Poljak, L., S. M. Batson, D. Ficheux, B. P. Roques, J. L. Darlix, and E. Kas.** 2003. Analysis of NCp7-dependent activation of HIV-1 cDNA integration and its conservation among retroviral nucleocapsid proteins. J. Mol. Biol. **329:**411–421.
29. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes in vitro. EMBO J. **11:**291–303.
30. **Pryciak, P. M., and H. E. Varmus.** 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell **69:**769–780.
31. **Salter, R. D., D. N. Howell, and P. Cresswell.** 1985. Genes regulating HLA class I antigen expression in T-B lymphoblast hybrids. Immunogenetics **21:**235–246.
32. **Scherdin, U., K. Rhodes, and M. Breindl.** 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. **64:**907–912.
33. **Schmidt, M., D. A. Carbonaro, C. Speckmann, M. Wissler, J. Bohnsack, M. Elder, B. J. Aronow, J. A. Nolta, D. B. Kohn, and C. von Kalle.** 2003. Clonality analysis after retroviral-mediated gene transfer to $CD34^+$ cells from the cord blood of ADA-deficient SCID neonates. Nat. Med. **9:**463–468.
34. **Schmidt, M., S. Hacein-Bey-Abina, M. Wissler, F. Carlier, A. Lim, C. Prinz, H. Glimm, I. Andre-Schmutz, C. Hue, A. Garrigue, F. Le Deist, C. Lagresle, A. Fischer, M. Cavazzana-Calvo, and C. von Kalle.** 2005. Clonal evidence for the transduction of $CD34^+$ cells with lymphomyeloid differentiation potential and self-renewal capacity in the SCID-X1 gene therapy trial. Blood **105:**2699–2706.
35. **Schroder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:**521–529.
36. **Stevens, S. W., and J. D. Griffith.** 1994. Human immunodeficiency virus type

1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. Proc. Natl. Acad. Sci. USA **91:**5557–5561.

37. **Stevens, S. W., and J. D. Griffith.** 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. J. Virol. **70:**6459–6462.
38. **Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al.** 2001. The sequence of the human genome. Science **291:**1304–1351.
39. **Versteeg, R., B. D. van Schaik, M. F. van Batenburg, M. Roos, R. Monajemi, H. Caron, H. J. Bussemaker, and A. H. van Kampen.** 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res. **13:**1998–2004.
40. **Weidhaas, J. B., E. L. Angelichio, S. Fenner, and J. M. Coffin.** 2000. Relationship between retroviral DNA integration and gene expression. J. Virol. **74:**8382–8389.
41. **Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin.** 1994. Distribution of targets for avian retrovirus DNA integration in vivo. Genes Dev. **8:**1473–1487.
42. **Wu, X., and S. M. Burgess.** 2004. Integration target site selection for retroviruses and transposable elements. Cell. Mol. Life Sci. **61:**2588–2596.
43. **Wu, X., Y. Li, B. Crise, and S. M. Burgess.** 2003. Transcription start regions in the human genome are favored targets for MLV integration. Science **300:**1749–1751.
44. **Wu, X., Y. Li, B. Crise, S. M. Burgess, and D. J. Munroe.** 2005. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. J. Virol. **79:**5211–5214.