

NOTES

Integration Site Choice of a Feline Immunodeficiency Virus Vector

Yubin Kang,¹# Christopher J. Moressi,²# Todd E. Scheetz,^{3,4,5} Litao Xie,¹ Diane Thi Tran,¹
Thomas L. Casavant,^{2,4,5} Prashanth Ak,⁶ Craig J. Benham,⁶ Beverly L. Davidson,⁷
and Paul B. McCray, Jr.^{1,*}

Program in Gene Therapy, Department of Pediatrics,¹ Department of Electrical and Computer Engineering,² Department of Ophthalmology,³ Department of Biomedical Engineering,⁴ Center for Bioinformatics and Computational Biology,⁵ and Department of Internal Medicine,⁷ University of Iowa, Iowa City, Iowa 52242, and Genome Center, University of California, Davis, California 95616-8816⁶

Received 9 April 2006/Accepted 30 May 2006

We mapped 226 unique integration sites in human hepatoma cells following gene transfer with a feline immunodeficiency virus (FIV)-based lentivirus vector. FIV integrated across the entire length of the transcriptional units. Microarray data indicated that FIV integration favored actively transcribed genes. Approximately 21% of FIV integrations within transcriptional units occurred in genes regulated by the LEDGF/p75 transcriptional coactivator. DNA in regions of FIV insertion sites exhibited a “bendable” structure and a pattern of duplex destabilization favoring strand separation. FIV integration preferences are more similar to those of primate lentiviruses and distinct from those of Moloney murine leukemia virus, avian sarcoma leukosis virus, and foamy virus.

Retroviral gene transfer holds promise as a treatment for genetic diseases. However, integration carries a risk of insertional mutagenesis, as occurred in a trial for X-linked severe combined immunodeficiency (8). This result focused this field towards understanding integration site preferences for all classes of vectors under development. Sequencing of the human genome facilitated large-scale studies of integration patterns for human immunodeficiency virus type 1 (HIV-1) (17, 19), simian immunodeficiency virus (SIV) (6), murine leukemia virus (MLV) (26), avian sarcoma leukosis virus (ASLV) (17), and foamy virus (FV) (23). The differences in integration preferences between vectors of primate lentiviruses, MLV, ASLV, and FV suggest that it is inadvisable to assume that all vectors integrate with the same pattern.

Nonprimate lentiviral vectors derived from FIV were developed as an alternative to HIV vectors (18) and transduce a broad range of cells and tissues (3, 7, 10–13, 15, 16, 20, 21, 25). As a model for integration following hepatocyte transduction (12, 13, 21), we cloned, sequenced, and mapped FIV vector integration sites in human hepatoma cells. Vesicular stomatitis virus G protein-pseudotyped FIV encoding enhanced green fluorescence protein (eGFP) driven by a cytomegalovirus promoter/enhancer was produced as described previously (11, 25). HepG2 cells (HB-8065; ATCC, Rockville, MD) were transduced at a multiplicity of infection of ~1. We assayed eGFP expression by using a fluorescence-activated cell sorter at various intervals posttransduction. Following an initial de-

cline, the percentage of eGFP-expressing cells remained stable for the interval between 16 days and 1 month posttransduction (~5%). We isolated genomic DNA from non-selected cells 16 days following transduction and cloned FIV integration sites using adaptor-ligated PCR (Genome-Walker kit; BD Biosciences, Palo Alto, CA) as reported previously (19, 26).

We mapped 226 distinct FIV integration events to all chromosomes except chromosome 21 (Fig. 1) (see http://genome.uiowa.edu/SupplementaryData/Kang_JVI2006.html). We asked whether integration favored transcriptional units, defining an integration site as residing within a gene if it occurred between the transcriptional start and stop sites of one of the 23,874 human Reference Sequence (RefSeq) genes (May 2004 UCSC human reference sequence; NCBI Build 35). Of 226 integration sites, 179 (79%) occurred in genes. This percentage increased to 87% if all human mRNAs in GenBank were used. Published retrovirus integration data were downloaded and processed using our analysis pipeline to provide comparable data sets. The frequency of FIV integration in genes was similar to the frequencies reported for HIV and SIV (71% and 76%, respectively) (6, 19) and higher than those for MLV (45%) (26), foamy virus (31%) (23), and random integration (32.2%) (Table 1) (26). Of 179 integrations in RefSeq genes, 172 were in introns and only 7 were in exons or open reading frames (Table 1) (see http://genome.uiowa.edu/SupplementaryData/Kang_JVI2006.html). We identified a single integration hot spot with two events on chromosome 4 at position 123.54 Mb and only one integration event on gene-rich chromosome 19, which had significant affinity for HIV integration (19). A karyotype of the HepG2 cell line revealed that chromosome 19 had a normal size and banding pattern; however, one copy of chromosome 21 was missing (data not shown). FIV integration

* Corresponding author. Mailing address: Department of Pediatrics, 240G EMRB, University of Iowa College of Medicine, Iowa City, IA 52242. Phone: (319) 335-6844. Fax: (319) 335-6925. E-mail: paul-mccray@uiowa.edu.

These authors contributed equally to this work.

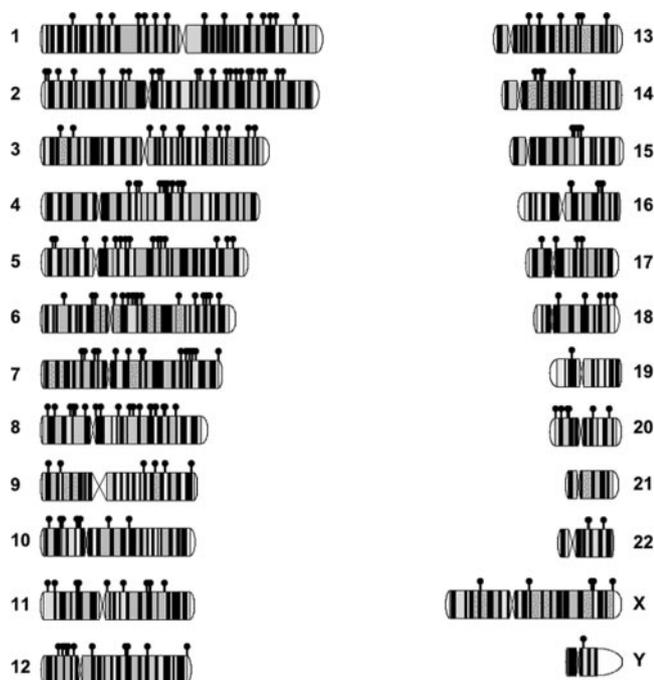


FIG. 1. FIV integration sites on human chromosomes. The unique FIV integration sites on all human chromosomes are shown. Each black dot represents one unique integration site. Some distinct integration sites appear to overlap due to their close proximity in the genome.

occurred along the entire length of RefSeq transcripts, and transcriptional start regions were not favored targets (Fig. 2).

The HepG2 cell transcriptional profile was analyzed using published Affymetrix data sets (22) (see http://genome.uiowa.edu/SupplementaryData/Kang_JVI2006.html). Of the 179 RefSeq genes targeted for integration, 163 were represented in the data set. The median expression value for all GeneChip array probes was 41.4, and the median expression value for the best probe in a RefSeq gene was 57.1. In contrast, the median expression value for the best probe in a RefSeq gene with an integration event was 124. For the 163 genes represented on the array with integration events, there was an approximately twofold increase in expression compared to that of the set of all RefSeqs and an approximately threefold increase in expression

TABLE 1. Comparison of integration site preferences for RefSeq genes among different viral transduction vectors

Viral vector	% of integration events in RefSeq genes ^a	Reference
FIV ^b	79	This study
HIV	71	19
SIV	76	6
MLV	45	26
ASLV	44	17
FV	31	23
Random	32.2	26

^a For accurate comparison, each data set was downloaded and analyzed with our integration mapping pipeline using the UCSC May 2004 human reference sequence (NCBI Build 35).

^b Of 226 integration sites, 179 occurred in genes.

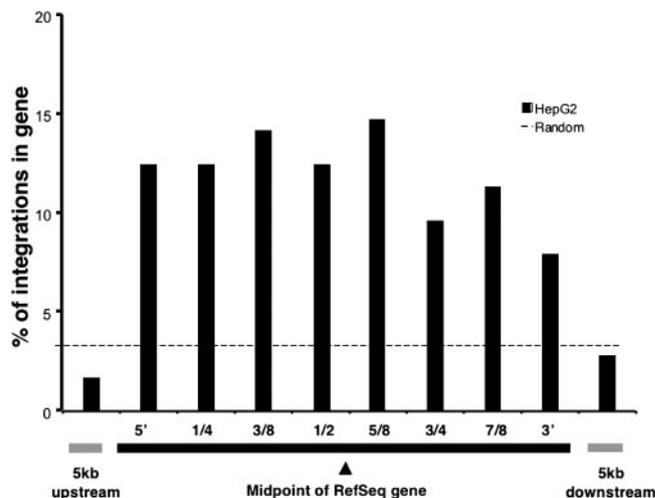


FIG. 2. Relationship of FIV integration to RefSeq genes. The cloned sequences were analyzed as described in the text. Regardless of size, the RefSeq genes with FIV integration events were divided into eight equal portions. The percentages of integrations occurring in each portion are shown. Integration events occurring within 5 kb upstream and 5 kb downstream of the gene are also shown. The dotted line represents the predicted pattern for random integration.

compared to that of all probes. Expression was also analyzed using the expressed sequence tag collection in the human UniGene set to determine which genes are transcribed in liver cells. Of the genes that could be correlated to NCBI's UniGene, approximately 95% (170/179) exhibited some level of expression in liver cells. These findings support the notion that FIV integrates into transcriptionally active regions of chromatin.

Recent studies indicate that LEDGF/p75 in host cells physically interacts with the integrase protein in HIV and FIV preintegration complexes, directing the complex to specific regions of chromosomal DNA and acting as a tethering protein (4, 14). Interestingly, HIV integration sites correlated with LEDGF/p75-regulated genes, and RNA interference knock-down of LEDGF/p75 expression changed the HIV integration pattern away from LEDGF/p75-regulated genes (5). This approach identified 1,849 LEDGF/p75-regulated Entrez genes on the Affymetrix U133 Plus2 chip (8.7% of the 21,311 Entrez genes represented on the U133 Plus2 chip). Using these same microarray data sets (5), we asked whether FIV integration sites shared a similar preference for LEDGF/p75-regulated genes. As with HIV integrations, the FIV vector integrated at a significantly higher frequency within genes regulated by LEDGF/p75 than within genes not LEDGF/p75 regulated. Of 179 FIV integrations, 38 (21.2%) occurred in LEDGF/p75-regulated genes ($P < 10^{-7}$ versus random integrations within genes). Of interest, the fraction of FIV integrants that mapped within LEDGF/p75-regulated genes was significantly higher ($P < 0.05$) than that observed for HIV integrants (~14%). This suggests that FIV may have a greater tendency than HIV to integrate in LEDGF/p75-regulated genes. These data support the idea that LEDGF/p75 is an important modulator of FIV and HIV lentiviral target site selection.

The specific nucleotide composition was analyzed for sequence-specific motifs or base preferences within 20 bases on

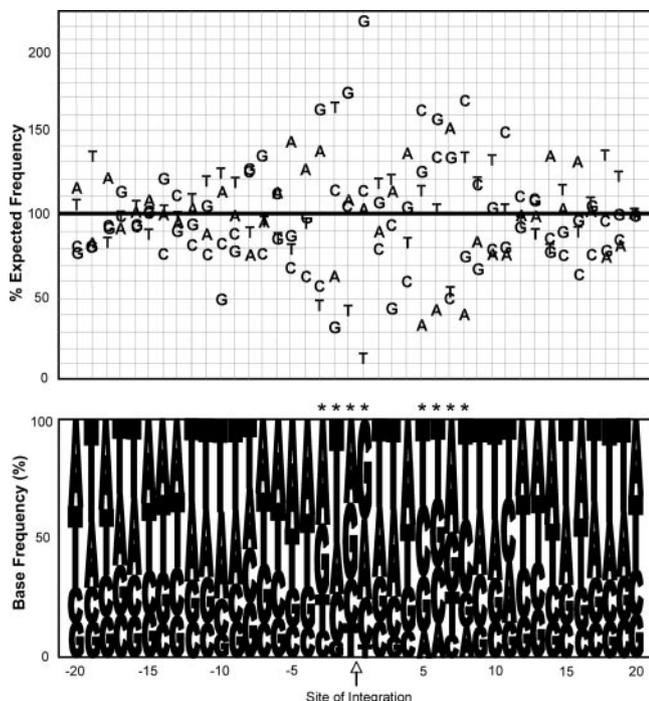


FIG. 3. Base preferences in regions immediately surrounding FIV integration sites. The upper panel shows the frequency of individual bases in a 20-bp window flanking the site of integration (indicated by an arrow below the bottom panel). The y axis indicates the frequency of each base as a percentage of the expected frequency. The lower panel presents the frequency of representation of bases at each position -20 and $+20$ nucleotides (relative to the 5' long-terminal-repeat end) from the site of integration. Asterisks indicate statistical significance at a P of <0.001 following a Bonferroni correction for multiple comparisons.

either side of integration sites (9, 27) (see http://genome.uiowa.edu/SupplementaryData/Kang_JVI2006.html). This analysis identified marked divergence from the expected frequencies of specific bases in the regions. Most interesting was an increase in the prevalence of G's at the sites of integration (Fig. 3). At several positions (from -3 to $+1$ and from $+5$ to $+8$) the deviations from predicted base preferences were very significant ($P < 10^{-7}$). The overall consensus sequences at these positions are RWRR and YKRY (using standard International Union of Biochemistry codes). While FIV nucleotide preferences shared a symmetry pattern similar to those reported for HIV, SIV, ASLV, and MLV (9, 27), the FIV preferences differed in primary sequence from HIV, MLV, SIV, and ASLV (9, 27). These results suggest that each retrovirus may have distinct, but not absolute, base preferences at integration sites.

We examined 20-base flanking regions on either side of the FIV insertion sites for physical properties, including B-DNA twist, A-philicity, DNA bending, and protein-induced deformability by using previously reported methods (24). Three sets of randomly selected control sequences were similarly analyzed and contrasted with FIV integration regions. No significant differences were observed in the physical properties of B-DNA twist, A-philicity, or protein-induced deformability. However, predicted DNA bendability adjacent to FIV insertion sites deviated significantly from the DNA bendability of the controls

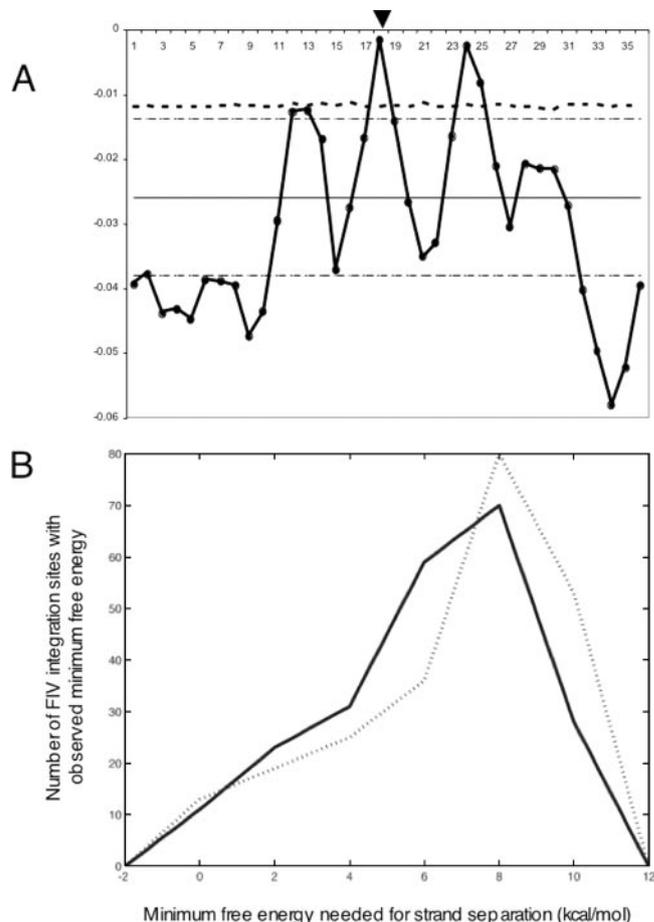


FIG. 4. (A) Bendability of DNA near sites of FIV integration. DNA flanking the point of integration 20 bp up- and downstream of the FIV integration site was analyzed for DNA bendability (thick black line with filled circles) and compared to control DNA samples as described in the text and at http://genome.uiowa.edu/SupplementaryData/Kang_JVI2006.html. Position 18 corresponds to the FIV insertion site (arrowhead). Randomly selected control sequences were used to calculate the mean (black line) ± 2 standard deviations (thin dashed lines) of values to characterize the significance of the variance. The threshold for determining significant bendability ($P < 0.01$) was calculated using a Monte Carlo simulation with 10,000 iterations (thick dashed line). (B) Distribution of the minimum free energies needed for strand separation in a 300-bp window at the FIV insertion sites (solid line) and control sites (dotted line). We assessed the differences at a significance level of 0.05 and obtained a P value for each test.

(Fig. 4A, position 18). In addition, deviations were observed at 6-bp intervals on either side of the integration site, specifically at positions 12, 18, and 24. A Monte Carlo simulation was performed to assess the significance of the observed differences in bendability. The threshold that only 1% of observations are expected to exceed ($P < 0.01$) is indicated in Fig. 4A. Bendability values at positions 18 and 24 both significantly exceed that threshold, and bendability at position 12 nearly exceeds that threshold. DNA bending models the ability of DNA to modify the depth and width of the major and minor grooves and may correlate with accessibility for binding by proteins or protein/DNA complexes. An additional analysis of physical properties was performed to compare the minimum free energies needed for strand separation at integration sites

to those at control sites (1, 2) (see http://genome.uiowa.edu/SupplementaryData/Kang_JVI2006.html). This analysis revealed that a significantly destabilized region (P values in the range of 10^{-3} to 10^{-4}) occurs roughly within 150 bp on either side of the insertion points, and a second destabilized region occurs 600 to 800 bp toward the 5' end of the insertion site (Fig. 4B). The two median distributions differ by 1.4 kcal/mol, which makes strand separation at the insertion sites 9.5 times more frequent at equilibrium. The probability of this pattern arising by chance is 4.57313×10^{-4} (Kolmogorov-Smirnoff test). These results suggest that FIV integration sites have distinct physical properties.

The current study provides novel insights into FIV integration site preferences. Overall, FIV integration preferences were similar to those reported for HIV (19) and SIV (6) and distinct from those reported for MLV, ASLV (17, 26), and FV (23). We speculate that the primary DNA sequences and physical properties at FIV target sites provide accessible regions for specific protein-protein and protein-DNA interactions that favor integration.

We thank Patrick L. Sinn for valuable discussion. We thank Dan Voytas for helpful discussions and for providing software for analysis of physical properties of genomic DNA. We thank Mirosalva Kaloper for technical help with computations. We thank Shivanand Patil and the UI Chromosome lab for karyotyping the HepG2 cell line.

This work was supported by the NIH (grants HL-75363, HL-79023, and HL-51670 to P.B.M.; grant GM-68903 to C.J.B.), the National Hemophilia Foundation (Y.K.), the Hemophilia Association of New York (Y.K.), a pilot and feasibility grant from the Center for Gene Therapy for Cystic Fibrosis (Y.K.), a career development award from Research to Prevent Blindness (T.E.S.), and the National Science Foundation (grant DBI-0416764 to C.J.B.).

REFERENCES

1. Benham, C. J. 1992. Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.* **225**:835–847.
2. Benham, C. J. 1993. Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc. Natl. Acad. Sci. USA* **90**:2999–3003.
3. Brooks, A. I., C. S. Stein, S. M. Hughes, J. Heth, P. B. McCray, Jr., S. L. Sauter, J. C. Johnston, D. A. Cory-Slechta, H. J. Federoff, and B. L. Davidson. 2002. Functional correction of established central nervous system deficits in an animal model of lysosomal storage disease with feline immunodeficiency virus-based vectors. *Proc. Natl. Acad. Sci. USA* **99**:6216–6221.
4. Busschots, K., J. Vercammen, S. Emiliani, R. Benarous, Y. Engelborghs, F. Christ, and Z. Debyser. 2005. The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J. Biol. Chem.* **280**:17841–17847.
5. Ciuffi, A., M. Llano, E. Poeschla, C. Hoffmann, J. Leipzig, P. Shinn, J. R. Ecker, and F. Bushman. 2005. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* **11**:1287–1289.
6. Crise, B., Y. Li, C. Yuan, D. R. Morcock, D. Whithy, D. J. Munroe, L. O. Arthur, and X. Wu. 2005. Simian immunodeficiency virus integration preference is similar to that of human immunodeficiency virus type 1. *J. Virol.* **79**:12199–12204.
7. Derksen, T. A., S. L. Sauter, and B. L. Davidson. 2002. Feline immunodeficiency virus vectors. Gene transfer to mouse retina following intravitreal injection. *J. Gene Med.* **4**:463–469.
8. Hacey-Bey-Abina, S., C. von Kalle, M. Schmidt, F. Le Deist, N. Wulfraat, E. McIntyre, I. Radford, J. L. Valleval, C. C. Fraser, M. Cavazzana-Calvo, and A. Fischer. 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **348**:255–256.
9. 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.
10. Hughes, S. M., F. Moussavi-Harami, S. L. Sauter, and B. L. Davidson. 2002. Viral-mediated gene transfer to mouse primary neural progenitor cells. *Mol. Ther.* **5**:16–24.
11. Johnston, J. C., M. Gasmil, L. E. Lim, J. H. Elder, J.-K. Yee, D. J. Jolly, K. P. Campbell, B. L. Davidson, and S. L. Sauter. 1999. Minimum requirements for efficient transduction of dividing and nondividing cells by feline immunodeficiency virus vectors. *J. Virol.* **73**:4991–5000.
12. Kang, Y., C. S. Stein, J. A. Heth, P. L. Sinn, A. K. Penisten, P. D. Staber, K. L. Ratliff, H. Shen, C. K. Barker, I. Martins, C. M. Sharkey, D. A. Sanders, P. B. McCray, Jr., and B. L. Davidson. 2002. In vivo gene transfer using a nonprimate lentiviral vector pseudotyped with Ross River virus glycoproteins. *J. Virol.* **76**:9378–9388.
13. Kang, Y., L. Xie, D. T. Tran, C. S. Stein, M. Hickey, B. L. Davidson, and P. B. McCray, Jr. 2005. Persistent expression of factor VIII in vivo following nonprimate lentiviral gene transfer. *Blood* **106**:1552–1558.
14. Llano, M., M. Vanegas, O. Fregoso, D. Saenz, S. Chung, M. Peretz, and E. M. Poeschla. 2004. LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* **78**:9524–9537.
15. Loewen, N., M. P. Fautsch, M. Peretz, C. K. Bahler, J. D. Cameron, D. H. Johnson, and E. M. Poeschla. 2001. Genetic modification of human trabecular meshwork with lentiviral vectors. *Hum. Gene Ther.* **12**:2109–2119.
16. Lotery, A. J., T. A. Derksen, S. R. Russell, R. F. Mullins, S. Sauter, L. M. Affatigato, E. M. Stone, and B. L. Davidson. 2002. Gene transfer to the nonhuman primate retina with recombinant feline immunodeficiency virus vectors. *Hum. Gene Ther.* **13**:689–696.
17. Mitchell, R. S., B. F. Beitzel, A. R. W. 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. [Online.]
18. Poeschla, E. M., F. Wong-Staal, and D. L. Looney. 1998. Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors. *Nat. Med.* **4**:354–357.
19. 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.
20. Sinn, P. L., S. L. Sauter, and P. B. McCray, Jr. 2005. Gene therapy progress and prospects: development of improved lentiviral and retroviral vectors—design, biosafety, and production. *Gene Ther.* **12**:1089–1098.
21. Stein, C. S., Y. Kang, S. L. Sauter, K. Townsend, P. Staber, T. A. Derksen, I. Martins, J. Qian, B. L. Davidson, and P. B. McCray, Jr. 2001. In vivo treatment of hemophilia A and mucopolysaccharidosis type VII using nonprimate lentiviral vectors. *Mol. Ther.* **3**:850–856.
22. Tachibana, K., Y. Kobayashi, T. Tanaka, M. Tagami, A. Sugiyama, T. Katayama, C. Ueda, D. Yamasaki, K. Ishimoto, M. Sumitomo, Y. Uchiyama, T. Kohro, J. Sakai, T. Hamakubo, T. Kodama, and T. Doi. 2005. Gene expression profiling of potential peroxisome proliferator-activated receptor (PPAR) target genes in human hepatoblastoma cell lines inducibly expressing different PPAR isoforms. *Nucl. Recept.* **3**:3.
23. Trobridge, G. D., D. G. Miller, M. A. Jacobs, J. M. Allen, H. P. Kiem, R. Kaul, and D. W. Russell. 2006. Foamy virus vector integration sites in normal human cells. *Proc. Natl. Acad. Sci. USA* **103**:1498–1503.
24. Vigdal, T. J., C. D. Kaufman, Z. Izsvak, D. F. Voytas, and Z. Ivics. 2002. Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. *J. Mol. Biol.* **323**:441–452.
25. Wang, G., V. A. Slepishkin, J. Zabner, S. Keshavjee, J. C. Johnston, S. L. Sauter, D. J. Jolly, T. W. Dubensky, Jr., B. L. Davidson, and P. B. McCray, Jr. 1999. Feline immunodeficiency virus vectors persistently transduce nondividing airway epithelia and correct the cystic fibrosis defect. *J. Clin. Invest.* **104**:R55–R62.
26. 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.
27. 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.