SUPPLEMENTARY MATERIAL

Supporting Data

SUPPORTING DATA

I. Tumor development in hemophilia B mice was not associated with vector integration.

In the course of necropsy performed on all vector-treated mice, tumor development was identified in the liver and spleen of two mice, one ICLV (20 μg $p24^{gag}$, mouse #46)-treated and one IDLV (250 μg $p24^{gag}$, mouse #58)-treated mouse (**Supplementary Figure S1**). To investigate if vector integration was involved with the development of the aforementioned tumors, we employed qPCR to analyze vector copy number (VCN) in tumor DNA and in DNA extracted from surrounding normal tissues, as well as from spleen and liver tissues obtained from tumor-free animals (in the relevant cohort). As shown in **Supplementary Table S1**, VCN in tumor DNA was significantly lower than the VCN obtained from all normal tissues. These findings indicate that vector-induced insertional mutagenesis was not the cause of the aforesaid tumor development.

II. A novel S1/B1-qPCR methodology enables accurate quantification of HIV-1 vector integration to host chromatin in the presence of episomal vector genomes.

In an attempt to quantify HIV-1 integration in mouse and rat cell lines *in vitro,* Tervo *et al.* and Goffinet *et al.* employed B1- and BC-based qPCR assays.[1, 2] These approaches are based on two rounds of PCR amplification. The first round of PCR is based on a set of two uniquely designed primers. One is an anchor fusion primer whose 3' end is directed to vector sequence. The 5' end of this primer contains a DNA sequence of the lambda phage (lambda), which is unrelated to either the host or the viral genome.[3] The second primer employed in this reaction is directed to a DNA sequence within the mouse B1 element. It is assumed that the amplification products of the first round of PCR include double-stranded (DS) B1 DNA, as well as DNAs comprising vector and mouse sequences within and in proximity to the B1 element. Samples (~10%) of these PCR products serve as templates in a nested-PCR reaction (second round) in which one primer is directed to the non-vector region in the fusion primer (employed in the first round of PCR) while the other primer is homologous to DNA sequences in the viral LTR. Employing this technique, Tervo *et al.* were able to quantify the rate of HIV-1 integration *in vitro*. However, the results of this study also demonstrated that significant background PCR amplification took place in the absence of the B1 primers. These findings were attributed to the formation of single-stranded (SS) DNA products following linear amplification of viral DNA.[1] As shown in **Supplementary Figure S2**, integrated and especially episomal viral/vector genomes (linear, single- and double-LTR forms) serve as templates for synthesis of SS viral DNAs. We hypothesized that the magnitude and overall impact of this phenomenon on the results of the B1-qPCR would be enhanced in experimental settings in which the majority of the viral genomes are episomal forms, which cannot contribute to B1-dependent amplification. To test this hypothesis, B1-qPCR

analysis was carried out on integrated and nonintegrated HIV-1 vector DNAs (derived from C2C12 cells transduced with either illegitimately integrated or episomal IDLVs, respectively) either in the presence or absence of the B1 primers. Indeed, as shown in **Supplementary Figure S3**, the presence of the B1 primers had no effect on the efficiency of B1-qPCR amplification of nonintegrated HIV-1 vector DNA. In contrast to this, a dramatic ~64 fold reduction (an increase of 6.3 ΔC_t) in B1-qPCR amplification of integrated HIV-1 vector genomes was observed in the absence of the B1 primers. These findings indicated that the high level of episomal vector genomes in mouse liver following intraperitoneal administration of IDLVs renders the conventional B1-qPCR method unsuitable for quantifying low levels of integrase-independent integration of IDLVs *in vivo*. We hypothesized that eliminating the SS PCR products generated during the first round of PCR independently of the B1-primers would enhance the specificity and sensitivity of the B1-qPCR method. To test this hypothesis, mouse C2C12 cell line DNA samples containing either episomal vector forms, integrated vector genomes, or various mixtures of episomal and integrated vector genomes, were subjected to B1-qPCR analysis, during which DNA amplification products obtained in the first round of PCR were subjected to S1-nuclease treatment (specifically digests SS DNAs). As shown in **Supplementary Figure S4**, qPCR amplification was significantly inhibited in all S1 nuclease-treated samples. Most importantly, without the S1-treatment, the 10-fold dilution of integrated vector genome DNA had minimal or no effect on the kinetics of the B1-qPCR amplification (especially in the presence of episomal genomes). However, due to elimination of SS DNAs (generated independently of the B1 primers), a significant inhibition in overall B1-qPCR amplification was observed following dilution of integrated vector DNAs in all S1 nuclease-treated samples (including episome-containing samples). Note that in order to minimize the synthesis of SS DNA templates, the viral

anchor primer employed in the aforementioned first round of PCR was directed to sequences upstream to the viral PPT (rather than to sequences in the viral LTRs, as described by Tervo *et al.*). Furthermore, earlier studies[4, 5] demonstrated tissue specific patterns of HIV-1 vector integration. Since our studies focused on hepatic hFIX delivery, we sought to characterize the rate of illegitimate IDLV integration relative to the levels of ICLV integration in mouse liver. Therefore the standard curve employed in this study was based on liver genomic DNA of an ICLV-treated mouse.

Overall, three modifications were incorporated into the conventional B1-qPCR methodology[1]: a) a novel anchor primer directed to sequences upstream to the viral PPT, b) subjecting the products of the first round of PCR to S1-nuclease degradation, and c) employing an integrated vector genome standard curve based on liver genomic DNA of an ICLV-treated mouse. The modified S1/B1-qPCR method allows for the first quantification of the rate of low level illegitimate IDLV integration in the presence of episomal vector forms.

SUPPLEMENTARY FIGURES

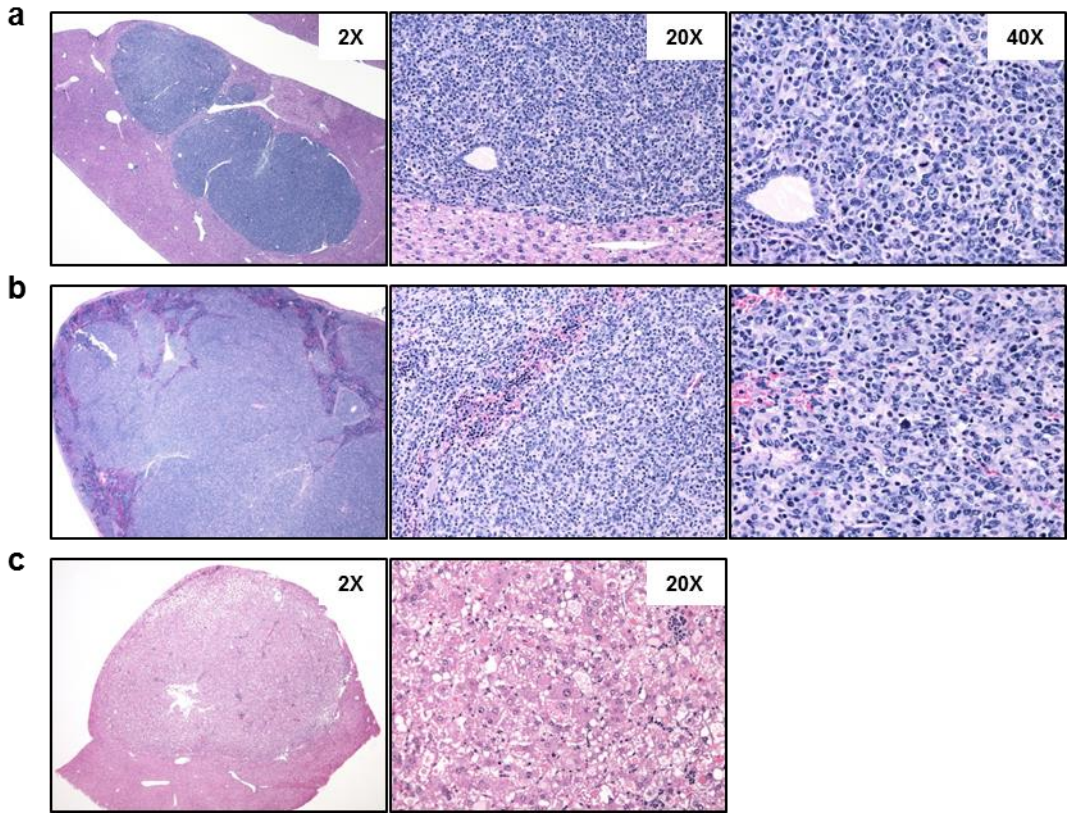

Figure S1 Histologic analysis of tumor tissues in mouse liver and spleen at one year post intraperitoneal administration of lentiviral vectors. Hematoxylin/eosin staining of tumor-containing liver (**a**, **c**) and spleen (**b**) tissues obtained from hemophilia B mouse (#46) intraperitonealy injected with ICLVs (20 µg $p24^{gag}$) (**a**, **b**) and from IDLV (250 µg $p24^{gag}$)-treated mouse (#58) (**c**). Histopathologic analysis of the aforementioned mouse organs contained histiocytic sarcoma (**a**, **b**) hepatocellular carcinoma (**c**) tumors.

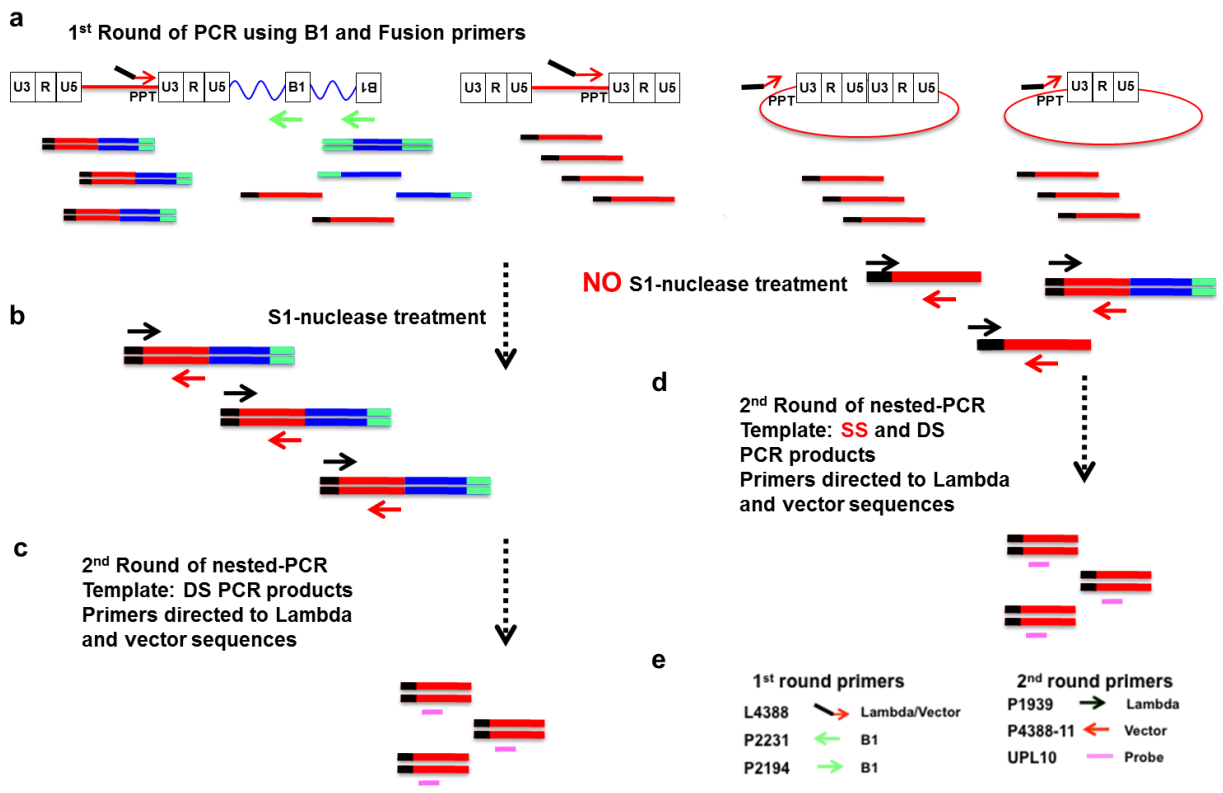

Figure S2 Optimization of the S1/B1-qPCR methodology.

(a) The first round of PCR amplification. Depiction of the DNA templates employed in the first round of amplification showing B1 elements and integrated vector genome in the context of host DNA, and extra-chromosomal vector genomes including linear and circular vector forms (either with a single or two LTRs). The fusion primer comprising a 3' sequence homologous to the vector DNA (red) and a 5' non-vector related sequence derived from the lambda phage[3] (black). The B1 element-directed primers are shown in green. The PCR products generated in the first round of PCR include double-stranded (DS) DNAs comprising host (blue) and vector (red) sequences, vector (red) and host (blue) single-stranded (SS) DNAs. The primers' parts in the PCR products are shown in black and green. (**b**) Elimination of SS DNA amplification products by S1-nuclease. (**c**) The second round of nested qPCR following S1-nuclease treatment. The

template DNAs employed in the nested qPCR include DS DNAs comprised of vector (red) and host (blue) sequences. The nested-qPCR primers are directed to the lambda phage portion of the fusion primer employed in the first round of PCR (black) and to a DNA sequence in the vector's LTR (red). The PCR amplification products of the nested qPCR are DS DNAs comprising the vector (red) sequence and the non-vector part in the fusion primer (black). The DNA probe (pink) employed in the nested qPCR is homologous to the vector sequence. (**d**) Nested-PCR amplification of DNA templates generated in the first round, which were not treated with S1-nuclease, including SS DNAs comprised of only vector DNA (red) and the non-vector part in the fusion primer (black). (**e**) Depiction of the probe and primers employed in the first and second rounds of S1/B1-qPCR.

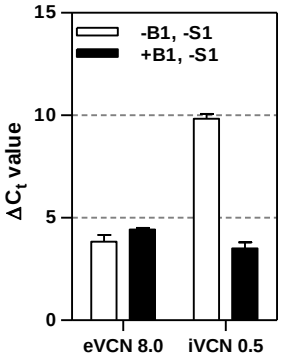

Figure S3 Significant background amplification of nonintegrating vector genomes by conventional nested B1-qPCR in the absence of the B1-directed primers.

DNA was extracted from IDLV-treated mouse C2C12 cells at 48 h post transduction (eVCN per cell = 8) or from the C2C12 cell line containing a heterogeneous population of illegitimately integrated IDLV genomes (iVCN per cell = 0.5). Conventional B1-qPCR was employed on the aforementioned DNA samples either in the presence or absence of B1-directed primers. ΔC_t was calculated as the difference between the C_t values obtained for the B1-qPCR and for the mouse GAPDH-qPCR, and served as a surrogate marker for vector integration frequency. Amplification of each sample was performed in triplicates; error bar is expressed as s.e.m.

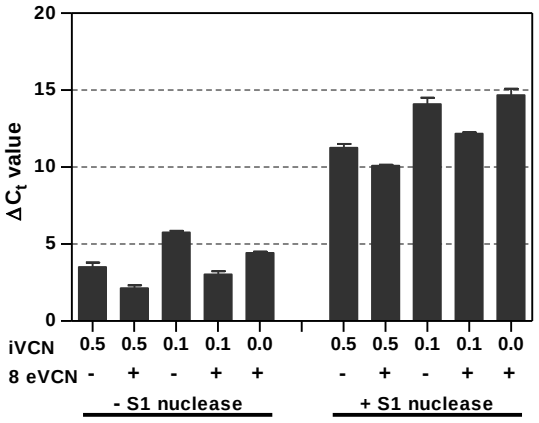

Figure S4 S1-nuclease treatment reduces background B1-qPCR amplification of nonintegrating HIV-1 vectors.

Mouse DNA samples containing 0.5 or 0.1 integrated vector genomes per cell, either with or without episomal vector genomes (eVCN per cell = 8), as well as mouse DNA samples containing only episomal vector genomes, were subjected to conventional B1-qPCR or to a modified procedure in which the amplification products generated during the first round of PCR were treated with S1-nuclease prior to their use in the nested qPCR. ΔC_t was calculated as the difference between the C_t values obtained for the B1-qPCR and for the mouse GAPDH-qPCR, and served as a surrogate marker for vector integration frequency. Amplification of each sample was performed in triplicates; error bar is expressed as s.e.m.

Table S1 Vector copy number (VCN) in tumor and in normal liver and spleen tissues

Total VCN	**Mouse #41**	**Mouse #47**	**Mouse #51**	**Mouse #46**	**Average**
Liver tissue	0.30	0.40	0.12	0.29	0.28
Spleen tissue	0.22	0.55	0.49	0.18	0.36
Tumor tissue				0.05*****	
Total VCN	**Mouse #67**	**Mouse #69**		**Mouse #58**	**Average**
Liver tissue	9.84	11.47		3.17	8.16
Tumor tissue				0.52*****	

VCN in DNA extracted from normal and tumor tissues was determined by a multiplex qPCR. The analyzed samples included tumor nodules (histiocytic sarcoma) and their surrounding tissues obtained from the liver and spleen of hemophilia B mouse (#46) injected with ICLVs (20 μg $p24^{gag}$). Reference DNA samples were extracted from liver and spleen tissues from tumor-free mice treated with a similar dose of ICLVs (#41, 47, and 51). Average VCN in non-tumor tissues was calculated. Also, VCN was determined in DNA samples extracted from a tumor (hepatocellular carcinoma) and its surrounding liver tissue at one year post IDLV administration (250 μg $p24^{gag}$) to mouse #58. VCN was determined in liver DNA obtained from tumor-free mice administered with a similar dose of IDLVs (#67 and 69). Average VCN in non-tumor tissues was calculated.

***** = marked decrease in total vector copy number.

SUPPLEMENTARY MATERIALS AND METHODS

Cell lines. HepG2 cells were maintained in Eagle's Minimal Essential Medium (Cellgro). 293T, Flp9, and mouse C2C12 cell lines were grown in Dulbecco's Modified Eagle Medium (Mediatech, Herndon, VA). All culture mediums were supplemented with 10% fetal bovine serum and 1% Antibiotic-Antimycotic solution (Cellgro).

Determination of hFIX protein concentration and coagulation activity.

hFIX protein concentration was measured using a sandwich ELISA, modified based on the original method described previously.[6] Briefly, hFIX proteins were captured with mouse monoclonal IgG anti-human factor IX (Haematologic Technologies, Essex Junction, VT), which diluted to a final concentration of 2 µg/ml in 0.1M carbonate buffer. Subsequently, captured hFIX proteins were detected with sheep polyclonal IgG anti-human factor IX conjugated with horseradish peroxidase (Affinity Biologicals, Ontario, Canada), which diluted to a final concentration of 0.5 µg/ml in PBS containing 1% bovine serum albumin and 0.05% Tween 20. ABTS substrate (Roche) was added and incubated for 5 min; absorbance was measured at 405 nm wavelength. Standard curve, with a linear range of 0.98-125 ng/ml, was generated using pooled normal human reference plasma (George King Biomedical, Overland Park, KS).

Multiplex qPCR. To quantify vector genomes in HepG2 cell lines, multiplex qPCR was carried out with the use of two sets of primers/probe: (i) a vector primers/probe set which consisted of NotI794 left primer (5'- taagaccaccgcacagca-3'), right primer (5'- cacttctccaattgtccctca-3'), and Roche Universal Probe Library (UPL) #25 (Cat. No. 04686993001); (ii) a housekeeping gene primers/probe set which contained human GUSB left primer (5'-gaatggggtcgacacgcta-3'), right primer (5'-tggaccaggttgctgatgt-3'), and Roche probe (Cat. No. 05190525001). Note that the GUSB left primer was redesigned in this study. Each qPCR reaction was prepared at final

concentrations of 1x ABsolute Blue QPCR Rox Mix (Thermo Scientific), 0.9 µM forward and reverse primers, 0.1 µM probe, and 100 ng gDNA. All qPCR reactions were performed at 1 cycle of 50 °C/2 min, 1 cycle of 95 °C/15 min, and 40 cycles of 95 °C/15 sec and 60 °C/1 min on a 7300 real time PCR system (Applied Biosystems). Genomic DNAs extracted from Flp9 cells[7] bearing one copy of the lentiviral vector per diploid genome were employed to establish a standard curve. Data were normalized to human GUSB levels. To quantify total vector genomes in mouse tissues, multiplex qPCR was carried out using a vector primers/probe set as described above, together with the mouse GAPD primers/probe (Roche Cat. No. 05046211001). The standard curve was generated from C2C12 cells transduced with ICLVs carrying codon-optimized R338L/GFP-BSD (VCN per cell = 1).

S1/B1-qPCR. Integrated vector genomes were quantified using an S1/B1-qPCR, modified from a B1-based qPCR described previously.[1] In this study, anchor vector primer, L4388 left primer (5'-atgccacgtaagcgaaactgcctggctagaagcacaaga-3'), was redesigned to bind outside the LTR regions, upstream to the polypurine tract region (PPT), in order to prevent amplification of single-stranded vector DNAs from both LTRs. Additional primers employed in the modified S1/B1-qPCR included two B1-directed primers documented previously[1]: P2194 (5'-acagccagggctacacagag-3') and P2231 (5'- cctcccaagtgctgggattaaag-3'); mouse GAPD primers (Roche Cat. No. 05046211001) were also incorporated in the reaction. Each S1/B1-qPCR reaction was carried out in a 1x PCR master mix (Promega) using similar PCR conditions and slightly modified thermal profiles (1 cycle of 95 °C/8 min and 20 cycles of 95 °C/10 sec, 60 °C/10 sec, and 72 °C/170 sec), as detailed previously.[1] The first round PCR products were subjected to S1-nuclease treatment (Fermentas) conducted at 37 °C for 30 min followed by enzyme inactivation at 70 °C for 10 min. S1 nuclease-treated PCR products were purified using a PCR

purification kit (Qiagen) prior to being subjected to the nested qPCR, which was carried out with the use of Lambda (P1939) left primer (5'-atgccacgtaagcgaaact-3')[3], P4388-11 right primer (5'-aggtacctgaggtgtgactgg-3'), and Roche UPL #10 (Cat. No. 04685091001). Data were normalized to mouse GAPD levels, which were measured in a separate reaction. Liver DNA obtained from an ICLV (65 μg $p24^{gag}$)-treated mouse was employed to establish a standard curve.

Nonintegrating viral DNA samples used in **Supplementary Figures S3** and **S4** were extracted from C2C12 cells transduced with IDLV (pTK1335) for 48 h, containing 8 VCN per cell. Of note, the sequence of the amplified PCR products, generated by mouse GAPD primers employed in multiplex qPCR and S1/B1-qPCR, is identical to a DNA sequence in chromosome 1 of C57BL/6J mice (NCBI Reference Sequence: NC_000067.6). The sequence of PCR products is as follows; primer sequences are underlined.

5'-<u>ATGGTGAAGGTCGGTGTGA</u>ACGAATTTGGCCGTATTGGGCGCCTGGTCAC

CAGGGCTGCCATCT<u>GCAGTGGCAAAGTGGAGATT</u>-3'

SUPPLEMENTARY REFERENCES

1. Tervo, HM, Goffinet, C and Keppler, OT (2008). Mouse T-cells restrict replication of human immunodeficiency virus at the level of integration. *Retrovirology* **5**: 58.

2. Goffinet, C, Michel, N, Allespach, I, Tervo, HM, Hermann, V, Krausslich, HG *et al.* (2007). Primary T-cells from human CD4/CCR5-transgenic rats support all early steps of HIV-1 replication including integration, but display impaired viral gene expression. *Retrovirology* **4**: 53.

3. Brussel, A and Sonigo, P (2003). Analysis of early human immunodeficiency virus type 1 DNA synthesis by use of a new sensitive assay for quantifying integrated provirus. *J Virol* **77**: 10119-10124.

4. Bartholomae, CC, Arens, A, Balaggan, KS, Yanez-Munoz, RJ, Montini, E, Howe, SJ *et al.* (2011). Lentiviral vector integration profiles differ in rodent postmitotic tissues. *Mol Ther* **19**: 703-710.

5. Marshall, HM, Ronen, K, Berry, C, Llano, M, Sutherland, H, Saenz, D *et al.* (2007). Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS One* **2**: e1340.

6. Walter, J, You, Q, Hagstrom, JN, Sands, M and High, KA (1996). Successful expression of human factor IX following repeat administration of adenoviral vector in mice. *Proc Natl Acad Sci U S A* **93**: 3056-3061.

7. Bayer, M, Kantor, B, Cockrell, A, Ma, H, Zeithaml, B, Li, X *et al.* (2008). A large U3 deletion causes increased in vivo expression from a nonintegrating lentiviral vector. *Mol Ther* **16**: 1968-1976.