 Related Commentary, page 755

# The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy

Eugenio Montini,[1] Daniela Cesana,[1,2] Manfred Schmidt,[3] Francesca Sanvito,[4] Cynthia C. Bartholomae,[3] Marco Ranzani,[1,2] Fabrizio Benedicenti,[1] Lucia Sergi Sergi,[1] Alessandro Ambrosi,[2,5] Maurilio Ponzoni,[4] Claudio Doglioni,[2,4] Clelia Di Serio,[2,5] Christof von Kalle,[3] and Luigi Naldini[1,2]

[1]San Raffaele-Telethon Institute for Gene Therapy and [2]Vita Salute San Raffaele University, Milan, Italy. [3]National Center for Tumor Diseases, Heidelberg, Germany. [4]Department of Pathology, San Raffaele Hospital, Milan, Italy. [5]University Centre of Statistics for Biomedical Sciences, Vita Salute San Raffaele University, Milan, Italy.

**γ-Retroviral vectors (γRVs), which are commonly used in gene therapy, can trigger oncogenesis by insertional mutagenesis. Here, we have dissected the contribution of vector design and viral integration site selection (ISS) to oncogenesis using an in vivo genotoxicity assay based on transplantation of vector-transduced tumor-prone mouse hematopoietic stem/progenitor cells. By swapping genetic elements between γRV and lentiviral vectors (LVs), we have demonstrated that transcriptionally active long terminal repeats (LTRs) are major determinants of genotoxicity even when reconstituted in LVs and that self-inactivating (SIN) LTRs enhance the safety of γRVs. By comparing the genotoxicity of vectors with matched active LTRs, we were able to determine that substantially greater LV integration loads are required to approach the same oncogenic risk as γRVs. This difference in facilitating oncogenesis is likely to be explained by the observed preferential targeting of cancer genes by γRVs. This integration-site bias was intrinsic to γRVs, as it was also observed for SIN γRVs that lacked genotoxicity in our model. Our findings strongly support the use of SIN viral vector platforms and show that ISS can substantially modulate genotoxicity.**

## Introduction

Integrative viral vectors commonly used in gene therapy may trigger oncogenesis as a consequence of insertional mutagenesis (1, 2). Because γ-retroviral vector (γRV) gene transfer into hematopoietic stem/progenitor cells (HSPCs) caused an unexpectedly high frequency of clonal proliferation and overt leukemia in clinical trials (3, 4), safer vectors and stringent preclinical safety assays (5, 6) are urgently needed to overcome this major hurdle.

We previously tested the oncogenic potential of prototypical murine leukemia virus–derived (MLV-derived) γRV and HIV-derived lentiviral vector (LV) using an in vivo genotoxicity assay based on transduction and transplantation of tumor-prone *Cdkn2a*$^{-/-}$ murine HSPCs (7). The *Cdkn2a* locus has a central role in regulating senescence and preventing cell transformation caused by aberrant oncogene expression. Because *Cdkn2a* inactivation synergizes with several types of cancer-promoting lesions, *Cdkn2a*$^{-/-}$ mice have been invaluable in insertional mutagenesis studies for identifying cancer genes, many of which are highly relevant in human oncogenesis. The relevance of the *CDKN2A* pathway in human tumor suppression is well documented because of its frequent inactivation in almost all types of human cancer (8, 9). Moreover, 2 X-linked SCID (X-SCID) patients affected by γRV-induced leukemia from 2 independent clinical trials had lost expression of the *CDKN2A* locus as a secondary mutation (10, 11). These findings indicate that *CDKN2A* plays a role also in the pathophysiology of human leukemias triggered by γRV insertions in clinical trials and further validate the choice of this model to assess vector genotoxicity.

In our previous study (7), γRV treatment triggered a dose-dependent acceleration of tumor onset in transplanted mice, whereas LV did not. Because the LV tested differed from γRV in both the molecular design (12, 13) and the integration site selection (ISS) (14–17), the relative contribution of these features to the lower genotoxicity of LV remained unknown. This may be due to the self-inactivating (SIN) long terminal repeat (LTR) (18) coupled to a moderately active promoter in the internal position. Indeed, the strong transcriptional enhancers of the LTR have long been associated with the transcriptional activation of oncogenes in γ-retrovirus–induced tumors (19, 20). A SIN LTR modification of γRV coupled to placement of retroviral enhancer/promoter within the vector reduced genotoxicity only 2-fold in vitro (21) and was insufficient to enhance safety in an in vivo model (22). The combination

**Authorship note:** Eugenio Montini and Daniela Cesana contributed equally to this work.

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Nonstandard abbreviations used:** CIS, common integration site; GO, Gene Ontology; HSPC, hematopoietic stem/progenitor cell; IPA, Ingenuity Pathways Analysis; ISS, integration site selection; LAM-PCR, linear amplification–mediated PCR; *Lin*$^{-}$, lineage marker negative; LV, lentiviral vector; MLV, murine leukemia virus; PGK, phosphoglycerate kinase; Q-PCR, quantitative PCR; Q–RT-PCR, quantitative RT-PCR; RTCGD, Retrovirus Tagged Cancer Gene Database; γRV, γ-retroviral vector; SB, Sleeping Beauty; SF, spleen focus–forming virus; SIN, self inactivating; TSS, transcription start sites; TU, transducing unit(s); VCN, vector copy number; VSV.G, vesicular stomatitis virus G protein.

**Citation for this article:** *J. Clin. Invest.* **119**:964–975 (2009). doi:10.1172/JCI37630.

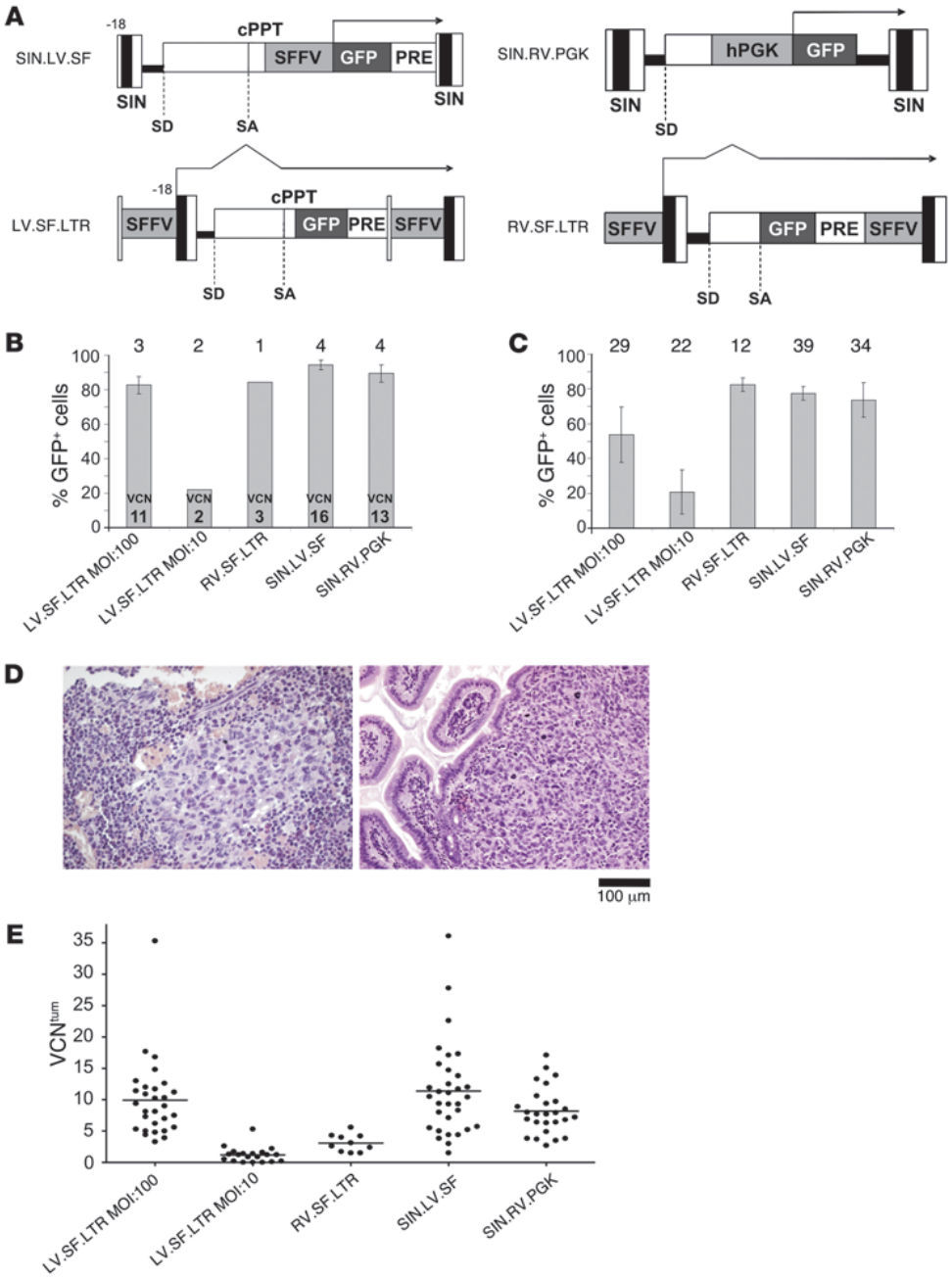


**Figure 1**
Transduction of *Cdkn2a*$^{-/-}$ HSPC by chimeric vectors and tumor development in transplanted mice. (**A**) Scheme of the proviral forms of the vectors tested. SIN, SIN LTR with deletion of the U3 region; SFFV, enhancer/promoter of the SF U3 LTR; PGK, promoter of the human phosphoglycerate kinase gene; SD and SA, viral splice donor/acceptor sites; cPPT, central polypurine tract; PRE, posttranscriptional regulatory element from the woodchuck hepatitis virus. Transgene transcripts are indicated by arrows. (**B**) Percentage of GFP$^+$ (mean ± SD) *lin*$^-$*Cdkn2a*$^{-/-}$ cells transduced in vitro with the indicated vectors 6 days after transduction. Number of experiments indicated on top. The average VCN per cell measured by Q-PCR 14 days after transduction is indicated inside the bar. (**C**) Percentage of GFP$^+$ cells (mean ± SD) in the blood of transplanted mice at 6–8 weeks after transplant. Number of mice indicated on top. (**D**) Representative H&E-stained sections of BM (left panel) and gut (right panel) from a myeloid tumor in the RV.SF.LTR group. Original magnification, ×20. Scale bar: 100 μm. (**E**) VCN distribution in tumor-infiltrated (VCN$^{tum}$) tissue (BM, spleen or thymus) for each mouse analyzed (dots) among the different treatment groups. Horizontal line represents the average VCN for each group.

of strong transcriptional enhancers with the γRV preference for integrating close to promoters may imply a higher risk of altering gene expression as compared with that of LV (16, 17, 23, 24). Moreover, the putative γRV integration bias for "hot spots" or gene classes associated with cell growth and cancer may further increase the oncogenic risk of insertional mutagenesis (7, 25–29).

In order to identify the most relevant features responsible for the different genotoxicity of γRV and LV, we swapped genetic features between these vectors and tested the genotoxicity of a panel of chimeric vectors on tumor-prone HSPCs in vivo. Using this strategy, we were able to identify transcriptionally active LTR as the major determinant of genotoxicity and validate the improved safety conferred by SIN LTR design in both vector platforms. Interestingly, however, ISS, when active LTR were present, modulated genotoxicity to an unanticipated extent.

## Results

*Vector construction and testing in tumor-prone HSC transplantation.* The panel of chimeric and parental vectors tested is shown in Figure 1A. We challenged the previously reported safety of LV by introducing the strong spleen focus–forming virus (SF) retroviral enhancer/promoter (30) in the U3 region of the LTR (LV.SF.LTR) and comparing this vector to a γRV carrying SF LTRs (RV.SF.LTR) (31). Similarly, we tested the oncogenic potential of a γRV with SIN LTRs carrying the moderately active human phosphoglycerate kinase (PGK) promoter in internal position (SIN.RV.PGK) (13). To address the position dependence of strong enhancer/promoters in genotoxicity, we placed the SF sequence in an internal position within the vector (SIN.LV.SF) and compared it with its SF.LTR counterpart. Finally, to assess the impact of promoter strength, we compared SIN LV with internal SF and PGK promoter. All vec-

**Table 1**
Tumor phenotype incidence in each group of transplanted mice

| Group | *n* | Lymphoid | Myeloid | *P* value vs. mock[A] | Guts with myeloid infiltrate | *P* value vs. mock[B] |
|---|---|---|---|---|---|---|
| LV.SF.LTR, MOI = 10 | 20 | 7 (35%) | 13 (65%) | NS | 7/19 (36%) | 0.0084 |
| LV.SF.LTR, MOI = 100 | 29 | 11 (38%) | 18 (62%) | NS | 13/29 (45%) | 0.0007 |
| LV.SF.LTR all | 49 | 18 (37%) | 31 (63%) | 0.038 | 20/48 (42%) | 0.0007 |
| RV.SF.LTR | 10 | 6 (60%) | 4 (40%) | NS | 4/9 (44%) | 0.0084 |
| SIN.LV.SF | 30 | 22 (73%) | 8 (27%) | NS | 2/29 (7%) | NS |
| SIN.RV.PGK | 27 | 23 (85%) | 4 (15%) | NS | 3/25 (12%) | NS |
| Mock | 20 | 13 (65%) | 7 (35%) | – | 0/17 (0%) | – |

[A]Tumor phenotype incidence in each group of transplanted mice was compared with that of mice transplanted with mock-treated cells (mock) by Fisher's exact test. $P < 0.05$ was considered significant. [B]Incidence of myeloid tumor infiltrate in the gut compared with that in the mock group. *n*, total number of mice analyzed.

tors expressed GFP and were pseudotyped by vesicular stomatitis virus G protein (VSV.G).

For each vector, *Cdkn2a*$^{-/-}$ BM–derived lineage marker–negative (*lin*$^-$) cells were transduced or not (mock) using the same culture conditions, a vector dose of $1 \times 10^8$ HeLa transducing units (TU)/ml (MOI = 100), and transplanted after a total culture time of 96 hours. Upon transduction, GFP$^+$ cells ranged from 84% to 95% by FACS for all vectors (Figure 1B). Average vector copy number per cell (VCN$^{\text{in vitro}}$) was measured by quantitative PCR (Q-PCR) in a sample of cells cultured in vitro for 2 weeks after transduction to exclude nonintegrated vector DNA. VCN$^{\text{in vitro}}$ was 11.3 ± 5.1 ($n = 3$) for LV.SF.LTR; 3.4 ± 0.2 with RV.SF.LTR; 15.5 ± 4.4 for SIN.LV.SF ($n = 3$); and 13 ± 4 for SIN.RV.PGK ($n = 4$). Because RV.SF.LTR transduction yielded a much lower VCN$^{\text{in vitro}}$ than LV.SF.LTR, we added another condition using LV.SF.LTR at a lower MOI (MOI = 10; $1 \times 10^7$ TU/ml) to obtain more comparable transduction levels. This resulted in VCN$^{\text{in vitro}}$ 1.6 ± 0.1 and 22% or more GFP$^+$ cells (note that the percentage of GFP$^+$ cells by FACS underestimates the transduction frequency of this vector because of low MFI).

Lethally irradiated wild-type FVB mice were transplanted with $7.5 \times 10^5$ vector-treated cells (LV.SF.LTR, $n = 29$; SIN.LV.SF, $n = 39$; RV.SF.LTR, $n = 12$; SIN.RV.PGK, $n = 34$; LV.SF.LTR, MOI = 10, $n = 22$; mock-transduced, $n = 32$, for a total of 168 mice in 7 different experiments). At 6 to 8 weeks after transplant, all mice showed normal frequency of myeloid (CD11b$^+$), B (CD19$^+$), and T (CD3$^+$) lymphoid cells in the blood, and the percentage of GFP$^+$ cells was consistent with the levels of transduction observed in vitro among all lineages (range 60%–90% GFP$^+$ for all groups or 20% for LV.SF.LTR, MOI = 10) (Figure 1C and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI37630DS1).

As expected, from 60 to 400 days after transplant, all mice in all experimental groups developed hematopoietic malignancies (7). The diagnosis was based on a blinded histopathology examination of BM, spleen, thymus, liver, kidney, lung, brain, gut, and lymph nodes and by FACS analysis of BM, blood, spleen, and thymus used to evaluate the proportion of myeloid and lymphoid cells and the relative frequency of GFP$^+$ cells (Supplemental Table 1 and Supplemental Figure 2). There was an increased frequency of myeloid tumors in the LV.SF.LTR and RV.SF.LTR groups as compared with the mock and other groups. Statistical significance, however, was only reached by LV.SF.LTR when both the MOI = 100 and MOI = 10 groups were merged to increase the sample size (31/49 vs. 7/20 for the mock; $P = 0.038$, Fisher's exact test) (Table 1). Moreover, the guts from the LV.SF.LTR and RV.SF.LTR treatment groups had a significantly increased occurrence of malignant myeloid infiltration (42% and 44%, respectively, vs. 0%–12% range of all the other groups; $P = 7 \times 10^{-4}$ and $8.4 \times 10^{-3}$ vs. mock, respectively, Fisher's exact test) (Table 1 and Figure 1D). Skewing of tumor phenotype and affected tissues may indicate an effect of vector treatment on the spontaneous oncogenesis of the mouse model.

Tumor-infiltrated tissues were systematically analyzed by Q-PCR to measure the relative amounts of donor-derived and recipient cells and the VCN in donor cells. *Cdnk2a*$^{-/-}$ cells constituted more than 70% of the affected tissue, indicating the donor origin of tumors. Donor cells in the tumor-infiltrated tissues, obtained from each group, had an average VCN of 9.9 ± 6.2 for LV.SF.LTR, MOI = 100, ($n = 29$) and 1.2 ± 1.2, MOI = 10, ($n = 20$); 3.0 ± 1.4 for RV.SF.LTR ($n = 10$); 11.4 ± 7.4 for SIN.LV.SF ($n = 32$); and 8.6 ± 4.1 for SIN.RV.PGK ($n = 20$) (Figure 1E and Supplemental Table 1).

*Assessing the oncogenic risk of vector treatment*. The mock-transduced control group had a median survival time corresponding to a donor cell age of 257 days, consistent with the median survival of *Cdkn2a*$^{-/-}$ mice (32, 33) and virtually identical to that in our previously published results (7), showing that cell manipulation and transplant procedures per se do not accelerate tumor onset and that our assay is highly reproducible. Survival in each experimental group was analyzed by Kaplan-Meier curves (Figure 2, A and B). Mice transplanted with HSPCs transduced at the higher dose of LV.SF.LTR (LV.SF.LTR, MOI = 100) died significantly earlier than the mock-transduced controls (median survival, 187 days vs. mock 248 days; $P < 0.0001$, log-rank Mantel-Cox test). Mice treated with the lower vector dose (LV.SF.LTR, MOI = 10) displayed a median survival of 211 days (Figure 2A). The median survival time was 194.5 days for RV.SF.LTR mice (Figure 2A), 227.5 days for SIN.RV.PGK mice, and 238 days for SIN.LV.SF mice (Figure 2B).

**Table 2**
log-logistic accelerated failure time model estimates the impact on survival of vector treatment in each group

| Vector | *n* | log-logistic parameter ± SE[A] | Z | *P* value vs. mock[B] |
|---|---|---|---|---|
| LV.SF.LTR, MOI = 100 | 29 | –0.292 ± 0.068 | –4.288 | $1.81 \times 10^{-5}$ |
| LV.SF.LTR, MOI = 10 | 22 | –0.113 ± 0.077 | –1.455 | $1.46 \times 10^{-1}$ |
| RV.SF.LTR | 12 | –0.202 ± 0.092 | –2.191 | $2.84 \times 10^{-2}$ |
| SIN.LV.SF | 39 | –0.040 ± 0.066 | –0.598 | $5.5 \times 10^{-1}$ |
| SIN.RV.PGK | 34 | –0.059 ± 0.070 | –0.831 | $4.06 \times 10^{-1}$ |

[A]log-logistic parameters ( ± SE) evaluate the impact on survival of vector treatment in each group. [B]Vectors with SF.LTR have a negative and significant impact on the survival of the LV.SF.LTR, MOI = 100, and RV.SF.LTR groups. $P < 0.05$ was considered significant. Z, Z score. *n*, total number of mice analyzed.

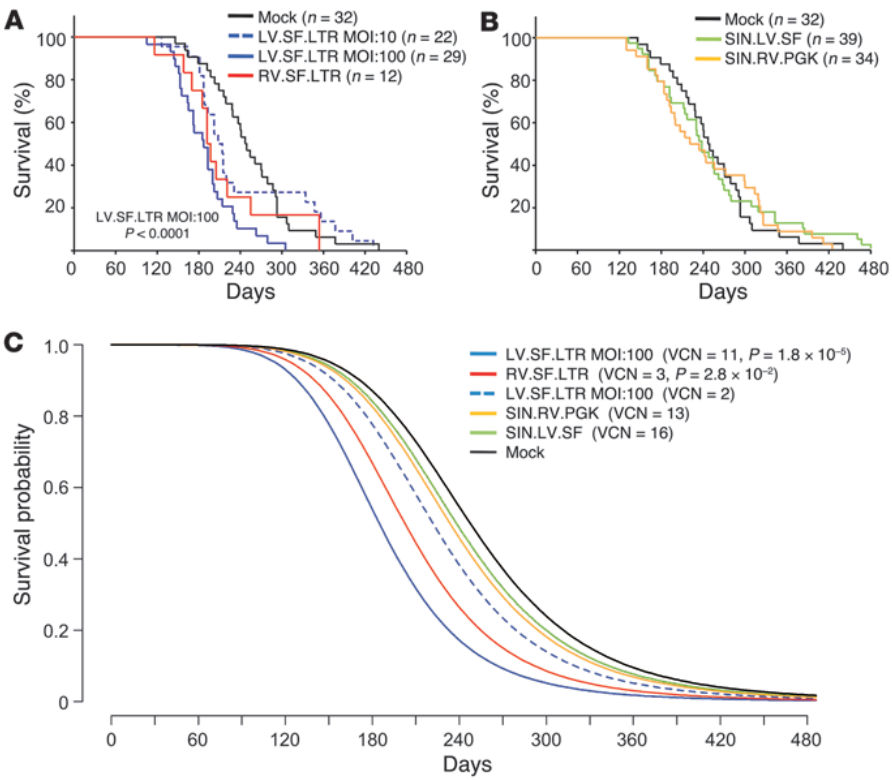


**Figure 2**
Survival curves related to vector treatment. (**A** and **B**) Kaplan-Meier survival curves of mice transplanted with cells transduced with SF.LTR vectors (**A**) and SIN.LTR vectors (**B**). For each panel, the survival curve of the mock group is shown (*n* = number of transplanted mice). Survival of the LV.SF.LTR, MOI = 100, group was significantly shorter than that of the mock group ($P < 0.0001$; Mantel-Cox log-rank test). (**C**) Survival probability over time was calculated for each group using the estimated log-logistic parameter (shown in Table 2) and compared with that of the mock group. The survival probability of the LV.SF.LTR, MOI = 100, and the RV.SF.LTR groups was significantly lower than that of the mock group (*P* values are indicated). A sample of the transduced cells was kept in vitro for 2 weeks after transduction to measure the average VCN (shown for reference).

In order to perform a risk assessment analysis that accounted for the accelerated time of death revealed in the descriptive statistics, we used an accelerated failure time model to estimate the risk of death and the survival probability for each treatment. Within this class of models, the log-logistic distribution provided the best fit to our experimental data (Supplemental Statistical Methods). The impact of the treatment (represented by the log-logistic parameter) and the survival probability during time were estimated for each treatment group and compared with that of the mock (Table 2 and Figure 2C). For the LV.SF.LTR, MOI = 100, as well as for the RV.SF.LTR groups, the survival probability was significantly lower than that of the mock ($P = 1.8 \times 10^{-5}$ and $P = 2.8 \times 10^{-2}$, respectively). On the other hand, treatment with LV.SF.LTR, MOI = 10, did not significantly reduce the survival probability as compared with that of the mock, indicating that the genotoxicity of LV.SF.LTR at this dose was below the detection limit of our in vivo assay. Notably, mice in the LV.SF.LTR, MOI = 10, and RV.SF.LTR groups were transplanted with cells carrying similar $VCN^{in\ vitro}$ (2 and 3, respectively), indicating a higher genotoxicity of RV.SF.LTR. The same analysis performed on the SIN.RV.PGK and SIN.LV.SF groups showed no significant impact of these vectors on the survival probability in our assay, even if both groups had high $VCN^{in\ vitro}$.

Because transduction with each vector resulted in a different integration load even when using the same MOI (MOI = 100), we adopted the VCN in tumor ($VCN^{tum}$) as a measure of dosage to perform VCN-matched comparisons between different vector treatment groups. Mice were stratified in groups having a $VCN^{tum}$ ranging from 1 to 6 ($VCN^{tum1-6}$) or above 6 ($VCN^{tum>6}$). The stratification criterion adopted allowed us to compare a relevant number of mice with a similar vector load (Table 3). log-rank Mantel-Cox test on the Kaplan-Meier curves showed that the mice in the LV.SF.LTR-$VCN^{tum>6}$ group died significantly earlier than those in the LV.SF.LTR-$VCN^{tum1-6}$ group ($P = 1 \times 10^{-3}$) and the mock group ($P < 1 \times 10^{-4}$) (Figure 3A). The survival of the RV.SF.LTR-$VCN^{tum1-6}$ group was also significantly reduced with respect to the mock group ($P = 2 \times 10^{-4}$) and the VCN-matched LV.SF.LTR-$VCN^{tum1-6}$ group ($P = 2 \times 10^{-2}$). Indeed, the LV.SF.LTR-$VCN^{tum1-6}$ group, which comprised 8 mice from the MOI = 100 and 13 from the MOI = 10 transduction groups, had a survival not significantly different from that of the mock group. The SIN.RV.PGK and SIN.LV.SF groups did not show any significant acceleration with respect to the mock group when stratified according to $VCN^{tum>6}$ (Figure 3B). Similar results were obtained by applying the log-logistic model to determine the survival probability of each stratified vector treatment group compared with the mock group (Supplemental Figure 3, A–C). These results showed that vector genotoxicity is dependent on the presence of active LTRs and on the VCN of

**Table 3**
Vector treatment groups stratified by mice that developed tumors with VCNs ranging from 1 to 6 and VCNs greater than 6

| Vector | Stratification group | *n* | Average $VCN^{tum}$ | Median survival (d) |
|---|---|---|---|---|
| Mock | NA | 32 | 0 | 248 |
| LV.SF.LTR | 1 to 6 | 22 | 2.8 | 214.5 |
| | >6 | 21 | 12 | 172 |
| RV.SF.LTR | 1 to 6 | 10 | 3.1 | 192 |
| SIN.LV.SF | 1 to 6 | 9 | 4.3 | 254 |
| | >6 | 23 | 14.2 | 238 |
| SIN.RV.PGK | 1 to 6 | 6 | 3.7 | 225.5 |
| | >6 | 20 | 9.5 | 238 |

*n*, number of mice for each stratification group.

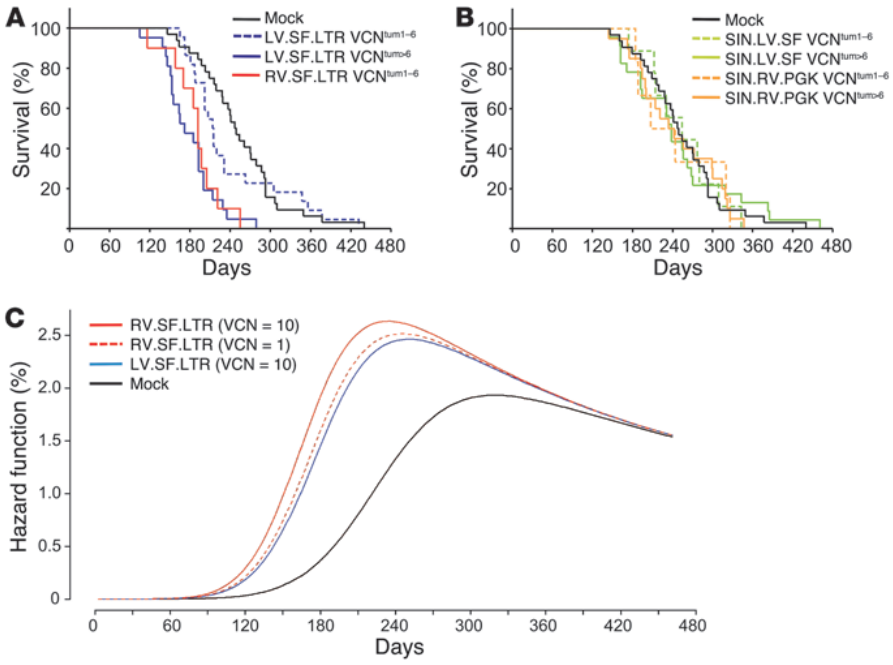


**Figure 3**
Survival curves and risk assessment related to vector treatment and dose. (**A** and **B**) Kaplan-Meier survival curves of mice treated with the SF.LTR (**A**) and SIN LTR vectors (**B**) stratified by the VCN$^{tum}$. (**C**) Percentage hazard of death over time of mice depending on the vector used and a fixed VCN$^{tum}$ of 1 (solid lines) or 10 (dashed lines).

in vitro–cultured cells or tumors. However, the negative impact on survival varies according to the type of vector and VCN.

In order to quantitatively assess the relative impact of VCN$^{tum}$ for each vector, we used the VCN$^{tum}$ of each mouse as a covariate in the accelerated failure time model. We observed that VCN$^{tum}$ acts linearly on the log-logistic hazard (Supplemental Statistical Methods). VCN$^{tum}$ affected negatively and significantly the risk of death only in interaction with the LV.SF.LTR, MOI = 100 (log-logistic parameter = –0.019; $P = 6.51 \times 10^{-5}$) and RV.SF.LTR (log-logistic parameter = –0.065; $P = 6.58 \times 10^{-3}$) but not with the LV.SF.LTR, MOI = 10, or any SIN.LTR vector (Table 4). The lack of a statistically significant impact of VCN$^{tum}$ for LV.SF.LTR, MOI = 10, suggests that genotoxicity was too low at this dosage to be measured in a reliable manner (Supplemental Figure 3D). This approach allowed estimation of the relative risk associated with different vectors at set dose levels. We thus plotted the risk of death at a fixed VCN$^{tum}$ of 1 or 10 for the genotoxic vectors (Figure 3C). RV.SF.LTR showed the highest risk, whereas LV.SF.LTR required a 10-fold higher integration load (VCN$^{tum}$) to reach the same risk of γRV.

*Vector integration analyses*. To gain functional evidence that LV.SF.LTR-driven oncogenesis in our *Cdkn2a*$^{-/-}$ model was mediated by insertional mutagenesis and to gain more insight into the low genotoxicity profile of SIN.RV.PGK and SIN.LV.SF, we compared genes targeted by integration of these vectors in tumors and in the cells used for transplant after 2 weeks of culture (in vitro). DNA from tumor-infiltrated BM (36 LV.SF.LTR, 27 SIN.LV.SF, and 23 SIN.RV.PGK mice) was subjected to a low-sensitivity linear amplification–mediated PCR (LAM-PCR) protocol aimed at identifying provirus-genomic junctions from predominant clone(s) in mixed populations (4, 7). Standard LAM-PCR protocol was used for the in vitro–cultured cells (34). We univocally mapped a total of 529 vector integration sites on the mouse genome (UCSC Mouse Genome Browser, February 2006 release) divided into 6 data sets: LV.SF.LTR, 100 sites from tumors and 70 in vitro; SIN.LV.SF, 80 from tumors and 90 in vitro; SIN.RV.PGK, 54 from tumors and 135 in vitro. The nearest gene (known to Entrez Gene or Ensembl) was then identified by bioinformatics analysis (Supplemental Table 2). Redundant integrations were excluded from calculations (total nonredundant integrations in tumors: LV.SF.LTR = 93 and SIN.LV.SF=78).

In the LV.SF.LTR and SIN.LV.SF data sets, the distribution of vector integrations displayed a pronounced tendency to integrate within genes (70%) without preference for transcription start sites (TSS), a pattern similar to that previously reported for other LVs (14, 16, 17, 23) (Supplemental Figure 4). On the other hand, 40% of SIN.RV.PGK integrations were located within genes, and 32% of the integrations clustered within ± 5 kb from the TSS.

Each data set was searched for matches to retroviral or Sleeping Beauty (SB) transposon common integration site (CIS) genes contained in the Retrovirus Tagged Cancer Gene Database (RTCGD) (35) and the frequency compared with the expected random frequency (Figure 4A, Supplemental Table 3A, and Supplemental Table 4). The SIN.RV.PGK had a pronounced tendency to target retroviral CIS in vitro and in tumors ($P < 0.0001$ vs. expected random; $\chi^2$ test) but not SB CIS. Several of the SIN.RV.PGK integrations targeting CIS genes mapped in the same narrow region previously targeted by retroviruses in the RTCGD (Supplemental Figure 5). On the other hand, the frequency of SIN.LV.SF integration near retroviral or SB CIS genes in vitro or in tumors was not significantly different from the expected random frequency. In contrast to the pattern observed for both SIN.LVs, LV.SF.LTR targeted both retroviral and SB CIS genes in

**Table 4**
log-logistic accelerated failure time model estimates the impact on survival of VCN per cell in tumors and vector treatment in each group

| Vector and VCN$^{tum}$ combined risk | log-logistic parameter ± SE[A] | Z | *P* value vs. mock[B] |
|---|---|---|---|
| LV.SF.LTR, MOI = 10 | 0.054 ± 0.036 | 1.52 | $1.28 \times 10^{-1}$ |
| LV.SF.LTR, MOI = 100 | –0.019 ± 0.005 | –3.994 | $6.51 \times 10^{-5}$ |
| RV.SF.LTR | –0.065 ± 0.024 | –2.717 | $6.58 \times 10^{-3}$ |
| SIN.LV.SF | 0.002 ± 0.003 | 0.591 | $5.54 \times 10^{-1}$ |
| SIN.RV.PGK | 0.004 ± 0.006 | 0.606 | $5.44 \times 10^{-1}$ |

[A]log-logistic parameters ( ± SE) evaluate the impact on survival of vector treatment and vector copy number on tumors in each group. [B]VCN$^{tum}$ of LV.SF.LTR, MOI = 100, and of RV.SF.LTR has a negative impact on survival as shown by the log-logistic parameter and Z score, whereas VCN$^{tum}$ of all other vectors does not. $P < 0.05$ was considered significant.

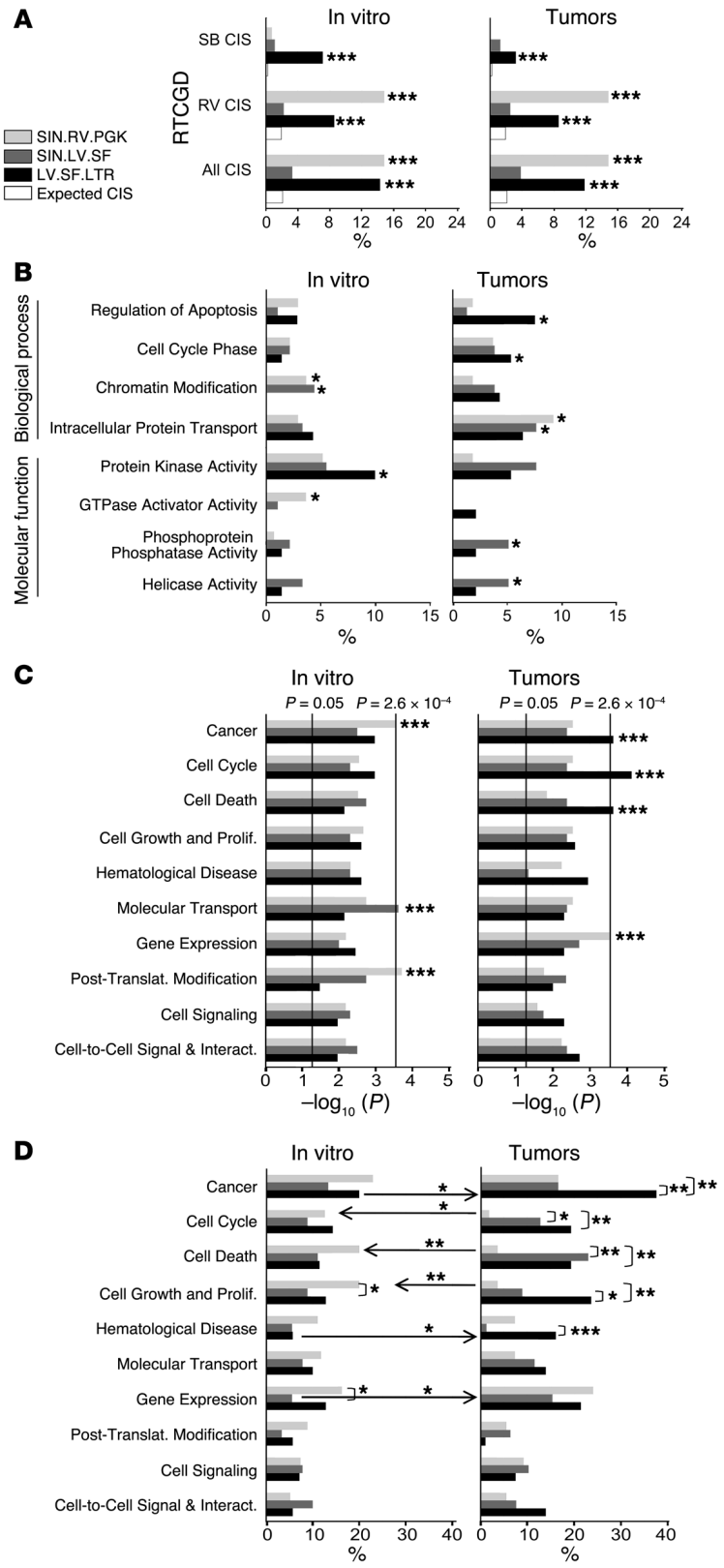


**Figure 4**
Vector integration site analysis in cells before transplant and in tumors. (**A**) Percentage of retroviral, SB transposon, and total (all) RTCGD CIS genes targeted in vitro and in tumors by each vector, as indicated. The expected random frequency was calculated as fraction of all mouse genes (25,613 genes). Significant overrepresentation versus the expected frequency ($P < 0.05$; $\chi^2$ test) is indicated by asterisks. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. (**B**) Percentage of vector integrations (as in **A**) targeting the indicated GO classes in vitro and in tumors. Significantly overrepresented classes are indicated by asterisks. Fisher's exact test. (**C**) Significance of overrepresented functional pathways of the IPA software is shown as $-\log_{10} P$ value. The significance threshold of $P < 0.05$ is indicated. Multiple comparison error correction by Bonferroni's method decreases the significance level to $P < 0.00026$ (63 gene classes for 3 vectors = 189 gene comparisons, $\alpha$ level of 0.05). Prolif., Proliferation; Post-Translat., Post-Translational; Interact., Interaction. (**D**) Percentages of genes belonging to the indicated IPA functional pathways were compared for each data set. Statistically significant differences ($P < 0.05$, Fisher's exact test) between the vectors in the same condition (bracket) or between the in vitro and tumor data set for the same vector (arrow) are indicated. Arrows from left to right indicate a significant enrichment from in vitro to tumors of the given gene class.

vitro and in tumors at significantly high frequency ($P < 0.0001$ vs. expected random; $\chi^2$ test).

The new data sets were then evaluated with DAVID-EASE (36) and Ingenuity Pathways Analysis (IPA) software for overrepresentation of Gene Ontology (GO) classes and signaling or disease pathways (Figure 4, B and C). Analysis was performed at high stringency and the results limited to the significantly overrepresented classes with an increase of 3-fold or greater with respect to the expected random frequency (DAVID-EASE; Supplemental Tables 3 and 4) or with 3 or more genes in at least 1 data set (IPA; Supplemental Table 5) and validated by Bonferroni's correction for multiple comparison error.

For LV.SF.LTR, the GO class Phosphorylation and Kinase Activity was overrepresented in vitro, while the GO classes Regulation of Apoptosis, Mitotic Cell Cycle, and B Cell Differentiation and the IPA classes Cancer, Cell Cycle, and Cell Death were all strongly overrepresented in tumors. As shown by *P* value ranking, the latter were the most significant overrepresentations found among all data sets, consistently with the observed oncogenic effect of LV.SF.LTR.

For SIN.LV.SF, the only overrepresented gene classes were the GO Chromatin Modification and the IPA Molecular Transport in vitro and the GO Helicases, Protein Phosphatase, and Intracellular Protein Transport in tumors.

For SIN.RV.PGK, the IPA Cancer and Post-Translational Modification classes were strongly overrepresented in vitro together with the GO classes Chromatin Modification and GTPase Activator, whereas the only overrepresented gene classes in tumors were the GO Protein Transport and Localization and the IPA Gene Expression classes.

We then determined the targeting frequency for the overrepresented IPA classes of each data set and performed 2-tailed Fisher's exact test for comparing vectors and conditions (Figure 4D). In tumors, LV.SF.LTR integrations at Cancer and Hematological Disease genes were significantly enriched from the in vitro data set and were more frequent than observed for the SIN.LV.SF and SIN.RV.PGK. Interestingly, SIN.RV.PGK integrations at Cell Cycle, Cell Death, and Cell Growth

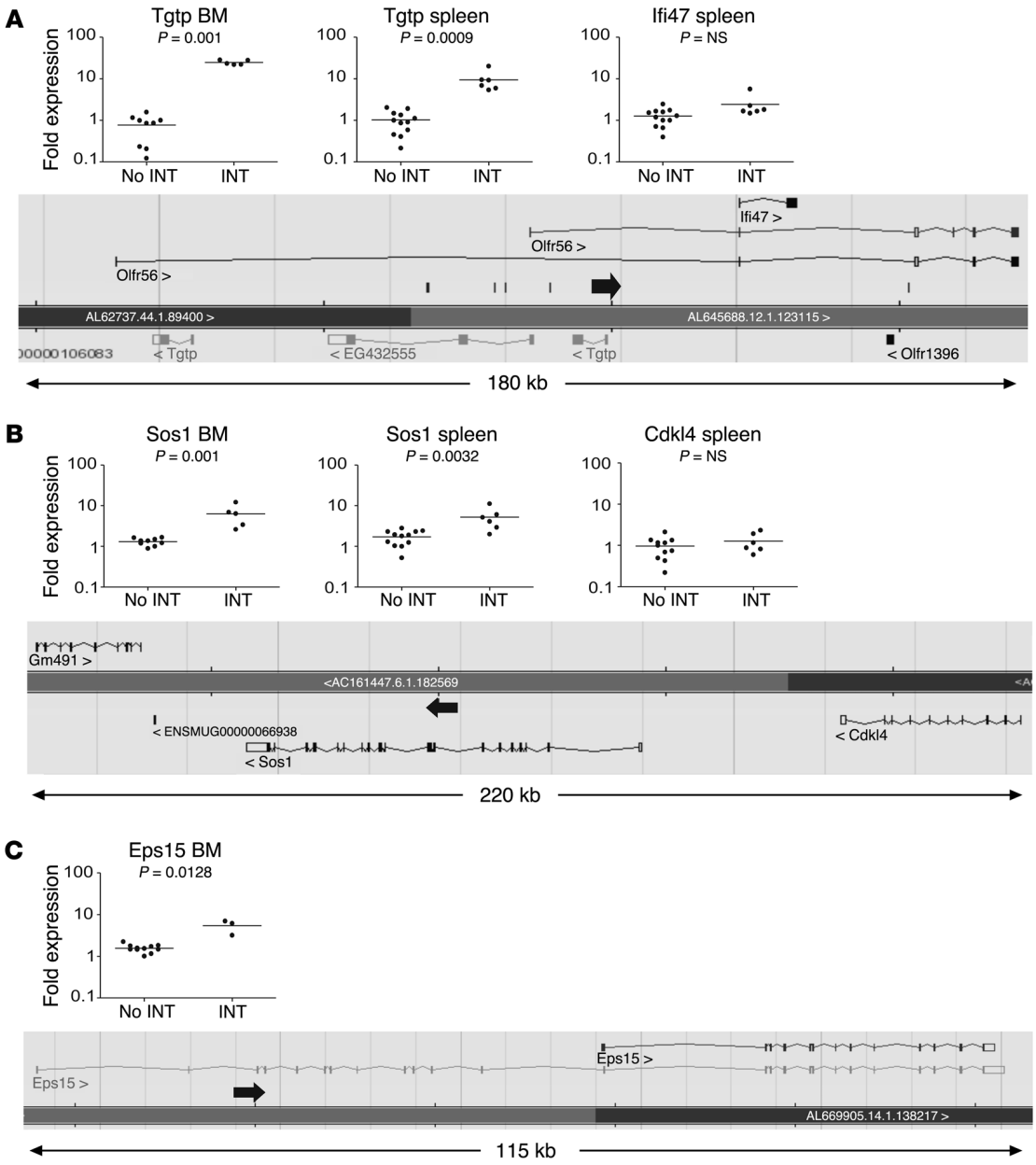


**Figure 5**

Gene expression analysis at LV.SF.LTR integration sites in tumors. (**A**–**C**) Expression of the indicated genes was measured by Q–RT-PCR on tumor-infiltrated BM or spleen cDNA (see also Supplemental Table 6). Expression data for primary and serially transplanted tumors with an integrated vector near the tested gene (INT) and phenotype-matched tumors with integrated vector in different sites or without integrations (No INT) are plotted. Each point is the fold change relative to matched-type tumor-infiltrated BM or spleen from the mock group (control level = 1); the horizontal bar represents the average. *P* value of the Mann-Whitney test comparison between the samples is indicated. $P < 0.05$ is considered significant. Genomic region targeted by the vector (vector position and orientation are represented by arrows) is shown below each set of expression data. Genes above the thick horizontal bar (chromosome) are transcribed from left to right; those below the chromosome are transcribed in the opposite direction. (**A**) *Tgtp*, which encodes for an interferon-inducible T cell–specific GTPase and whose TSS maps 530 bp from the vector integration, was overexpressed in both tumor-infiltrated BM and spleen of 2 primary and 4 secondary transplanted mice bearing the same integration; the expression of other genes surrounding the integration was not altered (see details in Supplemental Table 6). (**B**) Another integration from the same groups of mice mapped within the *Sos1* (37) oncogene, leading to its significant overexpression. (**C**) Vector integration occurred within the *Eps15* (38) oncogene, leading to its overexpression in tumors of 1 primary and 2 secondary transplanted mice.

and Proliferation genes were found at significantly reduced frequency in tumors compared with the in vitro data set.

*Mechanism of insertional mutagenesis by LV.SF.LTR.* In order to characterize the oncogenic mechanism of LV.SF.LTR, we measured the transcription level of genes near the vector integration site in early occurring tumors. When possible, tumors selected for analysis were transplanted into secondary mice to obtain biological replicates. For each gene near the integration site, we compared the average expres-

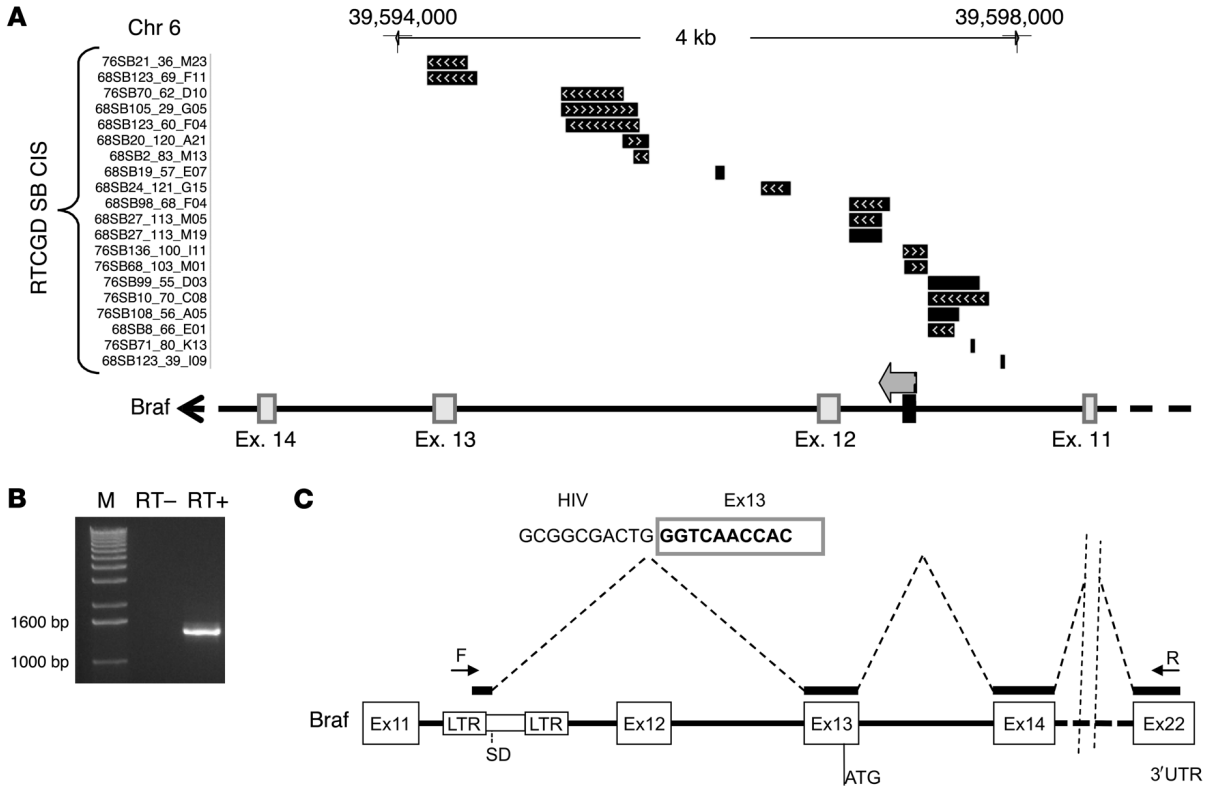


**Figure 6**

Oncogenic LV/*Braf* chimeric transcripts in an LV.SF.LTR tumor. (**A**) Genomic position of an LV.SF.LTR integration in a myeloid tumor targeting intron 11 of *Braf*. Chromosome (Chr) number and coordinates are indicated on top. The genomic interval covering exons 11 to 14 (gray boxes) is depicted. The position of the LV.SF.LTR integration (black box; LTR direction is indicated by the gray arrow) clusters with 20 SB integrations from sarcomas (39) in a narrow 4-kb region within introns 11 and 12. (**B**) RT-PCR using primers complementary to LV LTR and exon 22 of *Braf* on cDNA from the tumor described in **A** amplified a 1500-bp product. RT+, tumor cDNA; RT–, tumor RNA processed without reverse transcriptase; M molecular size markers. (**C**) The sequence of the RT-PCR product in **B** aligns to LV and to *Braf* exons. Black bars, amplified cDNA sequence; dashed lines, splicing events; F and R arrows, primers used for cDNA amplification; 3′UTR, 3′ untranslated region of *Braf*; SD, LV 5′ splice donor site. The cDNA sequence was LV specific up to the splice donor site (HIV) fused to the correct splice junction of exon 13 of *Braf* (boxed); exon 12 appears to be skipped. The first putative starting ATG codon in exon 13 is in the correct frame to produce a truncated Braf protein (indicated).

sion level in tumors carrying an integrated vector near or within that gene to the expression levels in tumors of identical phenotype but with different or no integration (Figure 5 and Supplemental Table 6). Q–RT-PCR was performed on cDNA from tumor-infiltrated BM and/or spleen tissue of 12 mice to test the expression of 18 genes surrounding 11 integration sites contained in 6 different tumors.

Two primary lymphoid tumors and 4 secondary transplants (2 from each primary tumor; tumors originated from different mice transplanted with the same in vitro–transduced cell populations) shared 4 integration sites. Among 9 genes tested that surrounded these integrations, *Tgtp* (Figure 5A) and *Sos1* (Figure 5B) were strongly and significantly overexpressed with respect to the controls (24.7 ± 3-fold increase and 6.3 ± 3.8-fold increase, respectively; $P < 0.001$, $n = 5$ vs. 9). These findings were confirmed in the spleen (9.5 ± 5.4-fold increase, $P < 0.001$; and 5.2 ± 3.3-fold increase, $P < 0.01$, respectively; $n = 6$ vs. 12).

In another lymphoid tumor and its 2 secondary transplants, *Eps15* showed a 5.4 ± 2-fold increase compared with the controls ($P = 0.013$, $n = 3$ vs. 11) among 5 genes near the vector integration (Figure 5C). All other genes tested were not significantly different from the controls. Of note, *SOS1* and *EPS15* have been implicated as oncogenes in human cancer development (37, 38).

In a myeloid tumor, 1 LV.SF.LTR integration mapped within intron 11 of *Braf*, a genomic region targeted several times by transposon integrations in sarcomas of *Arf*$^{-/-}$ mice (39) (Figure 6A). In this tumor, we detected a chimeric LV-*Braf* transcript that contained LV LTR and leader sequence up to the splice donor motif fused to the start of exon 13 and the remaining coding sequence of *Braf* (Figure 6B). This transcript must originate from the LV 5′ LTR by splicing out the genomic sequence spanning from the LV splice donor to the acceptor site of exon 13 (Figure 6C). The putative protein encoded by this transcript is a truncated Braf molecule with constitutive kinase activity similar to that previously reported upon transposon integration within the same region and that has been directly implicated in cell transformation (39).

In 1 myeloid and 1 lymphoid tumor, the *Nsd1* gene was targeted by 2 independent integrations 2,373-bp apart and in opposite orientation from each other (introns 5 and 6). The levels of expression of *Nsd1* appeared to be reduced to about 40% in both tumors bearing the integration ($n = 2$ vs. 5). Interestingly, *NSD1* haploinsufficiency is the major cause of Sotos syndrome and is associated with malignant tumor formation (40).

Overall, in each of the 6 tumors tested, we found at least 1 integration that resulted in either oncogene overexpression, generation

of aberrant transcripts encoding a truncated constitutively active oncogenic protein, or putative haploinsufficiency of a tumor suppressor gene. These genotoxic events recapitulate those previously described for γ-retroviruses or transposon-driven oncogenesis.

## Discussion

Here we show that LV with chimeric γRV LTR results in strong dose-dependent acceleration of tumor onset in the *Cdkn2a*$^{-/-}$ HSPC transplantation model, as observed for the γRV counterpart. On the other hand, both LV and γRV with SIN LTR appear to be neutral in our model. Although we cannot exclude that the SIN vectors tested could still display some degree of genotoxicity, their reduced oncogenic potential when compared with the vectors bearing active LTRs clearly indicates that transcriptionally active LTRs are major determinants of genotoxicity in retroviral vectors.

Beside active LTRs and vector dosage, we demonstrate that additional factors modulate vector genotoxicity to an unexpected extent. By comparing the survival of mice carrying matched copy numbers of LV or γRV with active LTRs either in the transplanted cells or in the tumors, we found that γRV was significantly more genotoxic than LV. By modeling the impact of vector and dosage on survival, we estimated that an approximately 10-fold higher integration load of LV with active LTRs is required to approach the same risk of a matched-design γRV. This does not mean that a single integration of our genotoxic LV within a cell might not trigger oncogenesis but rather that the relative oncogenic risk associated with an integration of LV and γRV, when bearing matched active LTRs, differs. This difference likely reflects the γRV integration bias for promoters (7, 14, 17, 23, 25–29) and selected gene classes involved in growth control and cancer (25–28), which may increase the probability of oncogene activation and, consequently, cancer development. Indeed, our vector integration analysis with respect to genomic features and gene classes supports a role of ISS in genotoxicity.

In the first detailed integration analysis of a SIN γRV, we observed a strong tendency to integrate not only near promoters but also near RTCGD CIS genes, which are hotspots of γ-retroviral integration retrieved from virus-induced tumors (35, 41) and are enriched in cancer genes. A bias for integration at Cancer genes was independently confirmed by IPA analysis, which showed strong overrepresentation of this gene class in the SIN.RV.PGK in vitro data set. Notably, this bias was observed despite the fact that the SIN γRV did not show genotoxicity in vivo and integrations at Cancer, Cell Cycle, and Cell Death genes appeared even counterselected in tumors. These findings strongly support the notion that γRVs have an intrinsic bias for integration at gene subsets enriching for cancer genes, which can be revealed even when genetic selection of cells harboring integration at oncogenes is prevented by the lack of strong transcriptional enhancers in the vector. On the contrary, SIN.LV.SF integrates near RTCGD CIS or Cancer genes at a frequency not significantly different from the expected random frequency.

LV.SF.LTR, however, showed a high frequency of integrations at RTCGD CIS and Cell Cycle, Apoptosis, and Cancer genes both in vitro and in tumors. Integrations at Cancer and Hematological Disease genes were further significantly enriched in tumors. The finding that only the genotoxic LV.SF.LTR and not the LV with SIN LTRs enriched for these types of integrations suggests that genetic selection of clones with integrations altering the expression of survival/proliferation genes occurred in vitro and in vivo. Genetic selection in vitro could occur because the cells were grown for 2 weeks before analysis (21, 42). Moreover, LV.SF.LTR targeted both retroviral and SB CIS genes, whereas the γRV tested targeted only retroviral CIS, further supporting the notion that the spectrum of targetable genes differs between the 2 vector types and that genetic selection rather than a genomic integration bias acquired by the LTR modification was the major driving force determining the LV.SF.LTR pattern of integration. Of note, because γRVs are unable to infect quiescent cells and are biased for integration at certain gene classes, the combination of effective insertional activation, typical of γRV, with the broad tropism and ISS of our LV.SF.LTR provides a mutagenic tool that may open new avenues to oncogene discovery in hematopoietic and solid tumors. We cannot exclude that the SF sequences introduced into the HIV LTR may influence the LV integration pattern by recruiting transcription factors that tether the preintegration complex to specific sites. This appears unlikely, however, because SIN.LV.SF, which bears the SF sequence in internal position, does not show an enrichment of integrations at Cancer genes as does LV.SF.LTR.

The characterization of the impact of LV.SF.LTR integration on the expression of targeted genes in tumors indicates that the mechanism of oncogenesis by LV.SF.LTR recapitulates the essential features discovered for γ-retroviruses, retrotransposons, and γ-RVs (3, 4, 7, 20, 39, 43–45). The lower risk of insertional mutagenesis by LV with active LTRs as compared with design-matched γRV can thus be explained by the lower frequency by which LVs target promoters of RTCGD CIS genes, many of which are involved in cancer.

The negative correlation between survival and $VCN^{rum}$ observed in mice treated with the same LV.SF.LTR dose suggests the occurrence of synergistic interaction between multiple integrations within the same cell in driving oncogenesis. This finding is in agreement with previous studies describing the cooccurrence of cooperating oncogenes in insertional mutagenesis screening performed in wild-type and tumor-prone mice (46–48) and supports the notion that both the total number of integrations administered and the average VCN per transplanted cell are relevant factors to be considered when estimating the oncogenic risk of a genotoxic vector (1, 2).

Accelerated tumor onset with an increased incidence of T cell lymphomas is a characteristic feature of MLV-mediated oncogenesis in *Cdkn2a*$^{-/-}$ mice (20, 32). In our study, the genotoxic LV.SF.LTR and RV.SF.LTR vectors induced an increased incidence of myeloid tumors that widely infiltrated peripheral organs. The reason(s) for the different phenotypic skewing induced by MLV and our vectors may be the different types of enhancers contained within the LTR and/or the different cell types targeted during chronic MLV infection in vivo as compared with a single ex vivo transduction of hematopoietic progenitors (49).

Remarkably, SIN.LV with internal PGK or SF promoter (this and our previous study; ref. 7) and SIN.RV.PGK did not accelerate tumor onset in the *Cdkn2a*$^{-/-}$ model, validating the improved safety of this configuration for both vector platforms. The improved safety of the SIN γRV design combined with a moderate internal cellular promoter (EF1α) was also recently shown by Zychlinski et al. (42) in an in vitro immortalization model.

The lack of genotoxicity of the SIN.LV.SF in our study is surprising given that (a) LV containing the same sequence duplicated within the LTR was genotoxic in our model; (b) SIN.RVs with an internal SF promoter were genotoxic both in vitro (21, 42) and upon transduction and serial transplantation of wild-type murine HSC (22); (c) a SIN LV with a strong LTR-derived murine stem

cell virus promoter in internal position was genotoxic in vitro (50); and (d) enhancers are well documented as interacting with distant promoters (20). We should consider that the background oncogenesis of *Cdkn2a*$^{-/-}$ HSPCs may hamper the detection of low-frequency or late-onset tumors induced by residual genotoxicity of the SIN.LV.SF vector and thus avoid the potential interpretation that LV can be safe even when containing strong enhancer sequences. Although our model did not reveal a difference in genotoxicity between SIN LV and SIN γRV, we show that the tendency of γRV to target cellular promoters and integrate at retroviral CIS and cancer genes remains unaffected by modifying the LTR for self inactivation. Therefore, we may speculate that SIN γRV with strong internal promoters may still pose a higher risk of genotoxicity than matched-design LV.

Nonetheless, the reduced oncogenic potential of SIN.LV.SF with respect to LV.SF.LTR is compelling and may help to unravel features of vector design contributing to genotoxicity. A possible explanation is that engagement in transgene transcription may reduce the probability and strength of interaction of the SF enhancer with neighboring genes (21, 42), while the presence of duplicated enhancer sequences increases the probability of transactivation. Another contributing factor is suggested by our finding of oncogenic LV/Braf chimeric transcripts in an LV.SF.LTR tumor, which suggests that the capacity to generate fusion transcripts by splicing capture may play an important role in vector genotoxicity. The mechanism of splicing mediated activation of oncogene by LVs with active LTRs has been recently described in a cell-culture assay for insertional mutagenesis (51). The placement of enhancer/promoter sequences in the LTR upstream of a strong splice donor site increases the probability of chimeric transcript formation as compared with other configurations such as the SIN.LV.SF vector, which lacks splice donor sites downstream of the internal SF promoter. We may thus speculate that safe vector design should avoid splice donor sites downstream of strong promoters in particular if a weak polyadenylation site is present in the vector. The inclusion of strong polyadenylation sites in gene therapy vectors to reduce the probability of readthrough into cellular genes has been proposed (52). However, the impact of this manipulation on vector genotoxicity remains to be directly tested, as it could also increase the frequency of premature transcription termination of genes targeted by vector integrations, potentially causing haploinsufficiency of tumor suppressor genes.

Other gene transfer vectors, such as those derived from spumaviruses (53), avian sarcoma leukosis virus (54), and transposons (55), display integration patterns potentially safer than those of γRVs and, for some vectors, even LVs. Further development of these platforms may establish their clinical potential.

Overall, considering our present findings together with those recently reported in other studies (21, 18, 42, 50, 51, 56), the safest design for retroviral vectors to alleviate the risk of insertional mutagenesis combines a SIN LTR and a moderately active internal promoter and should be the preferred choice at least until new technologies for site-specific gene editing or addition are fully validated for clinical applications (57, 58). The experiments described in this study were performed using a neutral transgene to investigate the oncogenic potential of several intrinsic vector features. However, both the biological activity of the transgene and the influence of the disease on target cell biology may significantly affect the oncogenic risk of vector treatment.

## Methods

*Mice*. Wild-type FVB/N.129 mice were obtained from Charles River. FVB.129-*Cdkn2a*$^{tm1Rdp}$ mice were obtained from the National Cancer Institute Mouse Models of Human Cancers Consortium. All mice were bred and kept in a dedicated pathogen-free animal facility and were killed when they showed signs of severe sickness. All procedures were performed according to protocols approved by the Animal Care and Use Committee of the San Raffaele Institute (IACUC 225 and 320) and communicated to the Ministry of Health and local authorities according to Italian law.

*Vector production*. SFFV LTR enhancer/promoter sequence contained in a 415-bp EcoRV/BamHI DNA fragment from pHRSIN-CSGW17 (59) was blunt-cloned into the 7300-bp EcoRI/BamHI DNA fragment from pCCL-SIN.cPPT.hPGK.eGFP.wPRE (7) to replace the hPGK promoter and generate SIN.LV.SF. To generate the LV.SF.LTR construct, the 7300-bp EcoRI/BamHI DNA fragment from pCCLSIN.cPPT.hPGK.eGFP.wPRE was religated to generate pCCLSIN.cPPT.eGFP.wPRE (without the hPGK promoter). The same 415-bp EcoRV/BamHI DNA fragment containing the SFFV LTR enhancer/promoter sequence was blunt cloned into the BbsI site in the U3 LTR region of pCCLSIN.cPPT.eGFP.wPRE, thus generating the LV.SF.LTR.

Concentrated LV stocks pseudotyped with the VSV.G envelope were produced by transient cotransfection of 4 plasmids in 293T cells and titered on HeLa cells as described (12). γRV stocks were similarly produced and titered using prkat43.3 PGK.GFP (13), pSRS.SF91 (31) (kindly provided by C. Baum, Department of Experimental Hematology, Hannover Medical School, Hannover, Germany) as transfer plasmids, pCMV-gagpol (MLV) (60), and VSV.G envelope encoding pMD2.VSV.G plasmid.

*Isolation and transduction of hematopoietic progenitors*. Six-week-old *Cdkn2a*$^{-/-}$ mice were killed by $CO_2$ inhalation, and BM was harvested by flushing femurs and tibiae with PBS-2% FBS (FBS; Invitrogen). *Lin*$^-$ cells were purified by lineage-marker–negative selection using the Enrichment of Murine Hematopoietic Progenitors Kit (StemCell Technologies), plated at a density of $1 \times 10^6$ cells/ml, and cultured in StemSpan SFEM expansion medium (StemCell Technologies) with a cytokine cocktail composed of 100 ng/ml SCF, 100 ng/ml thrombopoietin, 100 ng/ml Flt3 ligand, and 20 ng/ml IL-3 (PeproTech).

After 24 hours prestimulation, *Lin*$^-$ cells were split and subjected to mock, LV, or γRV transduction ($10^8$ TU/ml or $10^7$ TU/ml). After 12 hours, cells were washed and resuspended in the original medium and reinfected at 48 hours. Overall, cells were kept in culture for 96 hours before transplant. A sample of cells was kept for 14 days in culture to assess GFP expression by FACS analysis and genomic DNA extraction procedures.

*Transplantation procedures, FACS, and histopathology*. All transplantation procedures were performed as previously described (7). In brief, 6-week-old wild-type female FVB mice were lethally irradiated and injected in the tail vein with $7.5 \times 10^5$ cells/mouse. For secondary transplant, sublethally irradiated 6-week-old FVB female mice (5.75 Gy) were injected in the tail vein with marrow cells collected from tumor-infiltrated primary recipients ($2 \times 10^5$ cells/mouse).

FACS analysis was performed using lineage-specific antibodies (BD Biosciences — Pharmingen) against CD11b, CD19, CD3, and the IgG isotype control on cells obtained from blood collected 6 weeks after transplant or from blood, BM, spleen, thymus, and lymph nodes of diseased mice and analyzed with FCS Express 3 software (De Novo Software) as previously described (7). Dead cells were excluded by 7-aminoactinomycin D staining (5 μg/ml); blood erythrocytes were lysed with 7% ammonium chloride (StemCell Technologies).

For histological analysis, H&E staining was performed on 4-μm-thick sections of formalin-fixed, paraffin-embedded tissues. Specimens were evaluated in blinded fashion independently by 2 investigators (F. Sanvito and M. Ponzoni) as previously described. The semiquantitative scoring system was as follows: 0, no pathological infiltration; 1, mild infiltrates; 2, moderate infiltrates;

and 3, heavy infiltrates. For CD3, antigen immunolocalization was performed using rat anti-human CD3 (AbD Serotec) on formalin-fixed, paraffin-embedded 4-μm–thick sections after antigen retrieval with microwave using Tris-EDTA, pH 9. The immunoreaction was revealed by biotinilated-conjugated anti-rat antibody (Vector Laboratories) and HRP-conjugated streptavidin and using 3,3′-diaminobenzidine (DAB) as chromogen (BioGenex).

*VCN analysis*. Genomic DNA was extracted from cells (cultured in vitro for 14 days after transduction) was purified using the QIAGEN blood and cell culture DNA kit (QIAGEN). Genomic DNA from total BM or spleen or thymus of mice was purified using the QIAGEN tissue DNA kit (QIAGEN). Q-PCR analysis was performed as described (7), with probes complementary to mouse genomic β-actin (common to wild-type and *Cdkn2a*$^{-/-}$ mouse genomic DNA), GFP (common to all vectors tested), and neomycin resistance sequences (specific for *Cdkn2a*$^{-/-}$ cells) (See Supplemental Methods for information on oligonucleotides and probes used). Engraftment level was determined as the ratio between neomycin resistance and β-actin quantifications, using the DNA of a *Cdkn2a*$^{-/-}$ mouse as the standard curve. VCN was determined as the ratio between the relative amounts of GFP (common to all vector types) and the calculated engraftment levels. A GFP standard curve was made using DNA dilutions from a homozygous GFP transgenic mouse (61). Reactions were carried out according to the manufacturer's instructions and analyzed using the ABI Prism 7900 HT Sequence Detection System (PE; Applied Biosystems).

*LAM-PCR and genomic integration site analysis*. We used 0.5–5 ng of tumor DNA and 10–100 ng of transduced pretransplant DNA, respectively, as template for LAM-PCR (62). LAM-PCR was initiated with a 25-cycle linear PCR and restriction digest using Tsp509I or HpyCH4IV and ligation of a restriction site–complementary linker cassette. The first exponential biotinylated PCR product was captured via magnetic beads and reamplified by a nested second PCR. LAM-PCR primers for LV were previously described (4, 34, 63, 64). For the γRV LAM-PCR, 2 5′-biotinilated oligonucleotides complementary to the GFP sequence (LAMGFP1, 5′-TGGAGTTCGTGACCGCCGCC-3′ and LAMGFP2, 5′-GGGATCACTCTCGGCATGGAC-3′) were used for the linear amplification step. The 2 sequential exponential amplification steps were performed with nested oligonucleotides complementary to the γRV LTR sequence (LAMRV1, 5′-GACTTGTGGTCTCGCTGTTCCTTGG-3′ and LAMRV2, 5′-GGTCTCCTCTGAGTGATTGACTACC-3′), each coupled with the previously described oligonucleotides complementary to the linker cassette (34). LAM-PCR amplicons were separated on Spreadex gels (Elchrom Scientific) and the excised bands cloned into the TOPO TA vector (Invitrogen) or PCR-purified (QIAGEN), shotgun cloned, and sequenced (GATC Biotech). Sequences were aligned to the mouse genome (*Mus musculus* genome assembly February 2006, UCSC *Mus musculus* genome version 8) using the UCSC BLAT genome browser (http://genome.ucsc.edu). Identification of the nearest gene was performed with dedicated PERL scripts. Visualization of the RTCGD CIS integrations as a feature on the UCSC BLAT output was achieved by connecting to UCSC through the RTCGD web interface (http://rtcgd.abcc.ncifcrf.gov); map position of each retroviral and SB CIS integrations were automatically loaded as custom tracks on the UCSC BLAT search engine. Analysis of overrepresentation of gene classes in integration data sets was performed with DAVID-EASE software 2.0 (http://david.abcc.ncifcrf.gov/home.jsp). For each of the 4 data sets, we restricted further analysis to the GO categories significantly overrepresented, identified by at least 3 genes in the Biological Process or Molecular Function systems and with a fold increase of 3 or more. The murine gene IDs in our data sets were converted to the respective human putative orthologs by searching the Ensembl database (http://www.ensembl.org/biomart/martview). Functional networks analyses were performed with IPA 6 software (Ingenuity Systems). Only significantly overrepresented classes identified by at least 3 genes were considered.

*RNA isolation and gene expression analysis*. Total RNA from tumor-infiltrating BM cells or spleen was isolated with the RNeasy Mini Kit (QIAGEN). cDNA preparation was performed using Mo-MLV reverse transcriptase and random hexamer primers (SuperScript III First-Strand Synthesis System for RT-PCR; Invitrogen). cDNA was used as template for TaqMan Gene Expression Assays specific to each gene (Applied Biosystems) (see Supplemental Methods for gene expression assay ID and probe sequences used). Amplification reactions were performed on a 7900HT Real-Time PCR Thermal Cycler (Applied Biosystems). The relative expression level of each gene was calculated by the $^{\Delta\Delta}$Ct method (65), normalized to *Hprt* expression (housekeeping gene control), and represented as fold change relative to mock-transduced samples (calibrator). Real-time PCR Miner software (http://www.miner.ewindup.info) (66) was used to calculate the mean PCR amplification efficiency for each gene. The qBase software program (http://www.biogazelle.com) was used to measure the relative expression level for each gene (67). Amplification of the chimerical transcript between LV.SF.LTR and *Braf* was obtained using oligonucleotides complementary to the lentiviral LTR (LTR 5′-CCTGCCACCACGACTAGAATGC-3′) and exon 22 of the *Braf* gene (BrafAS 5′-GGACTGGCTACTTGAAGGCT-3′).

*Statistics*. Statistical analysis was performed with the R-statistical (version 2.1.1; http:/www.r-project.org) or GraphPad Prism 4.0 software (GraphPad Software). For survival analysis, a parametric approach was used because smooth continuous estimates of the survivor function were necessary for predictive purposes as described (7). The formula describing the log-logistic distribution is as follows: $(\rho^k \times kt^{k-1})/[1 + (t\rho)^k]$, where $t$ is the time; ρ is risk of failure, and $k$ is the log-logistic parameter that determines the effect of the treatment. The log-logistic distribution provided the best fit to our experimental data (Supplemental Statistical Methods). Time of death and VCN$^{tum}$ obtained from LV.SF.LTR, MOI = 10; LV.SF.LTR, MOI = 100; RV.SF.LTR; SIN.LV.SF; SIN.LV.PGK (7); and SIN.RV.PGK groups were used as variables in the log-logistic accelerated failure time model. The exclusion or inclusion of the SIN.LV.PGK survival data from our previous work (7) did not significantly change the results of our statistical analysis. For all statistical comparisons, $P < 0.05$ was considered significant.

## Acknowledgments

We are grateful to A. Guffanti and A. Thrasher for helpful discussion and suggestions. This work was supported by grants from Fondazione Telethon (TIGET grant), the European Union (projects LSHB-CT-2004-005242, CONSERT and LSHC-CT-2005-518198), the NIH (2 P01 HL053750-11 CFDA no. 93.839), and Associazione Italiana per la Ricerca sul Cancro (no. 1192) to L. Naldini.

Received for publication October 1, 2008, and accepted in revised form January 14, 2009.

Address correspondence to: Eugenio Montini or Luigi Naldini, San Raffaele-Telethon Institute for Gene Therapy, via Olgettina 58, 20132, Milan, Italy. Phone: 39-02-26433869; Fax: 39-02-26434621, E-mail: montini.eugenio@hsr.it (E. Montini). Phone: 39-02-26434681; Fax: 39-02-26434621, E-mail: naldini.luigi@hsr.it (L. Naldini).

1. von Kalle, C., et al. 2004. Stem cell clonality and genotoxicity in hematopoietic cells: gene activation side effects should be avoidable. *Semin. Hematol.* **41**:303–318.
2. Baum, C., et al. 2003. Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood.* **101**:2099–2114.
3. Hacein-Bey-Abina, S., et al. 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **348**:255–256.
4. Ott, M.G., et al. 2006. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* **12**:401–409.

5. Will, E., et al. 2007. Importance of murine study design for testing toxicity of retroviral vectors in support of phase I trials. *Mol. Ther.* **15**:782–791.
6. Williams, D.A. 2006. Gene therapy advances but struggles to interpret safety data in small animal models. *Mol. Ther.* **13**:1027–1028.
7. Montini, E., et al. 2006. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat. Biotechnol.* **24**:687–696.
8. Sherr, C.J. 2001. The INK4a/ARF network in tumour suppression. *Nat. Rev. Mol. Cell Biol.* **2**:731–737.
9. Sherr, C.J. 2004. Principles of tumor suppression. *Cell.* **116**:235–246.
10. Hacein-Bey-Abina, S., et al. 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**:3132–3142.
11. Howe, S.J., et al. 2008. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J. Clin. Invest.* **118**:3143–3150.
12. Follenzi, A., Ailles, L.E., Bakovic, S., Geuna, M., and Naldini, L. 2000. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat. Genet.* **25**:217–222.
13. Roberts, M.R., et al. 1998. Antigen-specific cytolysis by neutrophils and NK cells expressing chimeric immune receptors bearing zeta or gamma signaling domains. *J. Immunol.* **161**:375–384.
14. Hematti, P., et al. 2004. Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol.* **2**:e423.
15. Lewinski, M.K., et al. 2006. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* **2**:e60.
16. Schroder, A.R., et al. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* **110**:521–529.
17. Wu, X., Li, Y., Crise, B., and Burgess, S.M. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science.* **300**:1749–1751.
18. Zufferey, R., et al. 1998. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**:9873–9880.
19. Lenz, J., et al. 1984. Determination of the leukaemogenicity of a murine retrovirus by sequences within the long terminal repeat. *Nature.* **308**:467–470.
20. Jonkers, J., and Berns, A. 1996. Retroviral insertional mutagenesis as a strategy to identify cancer genes. *Biochim. Biophys. Acta.* **1287**:29–57.
21. Modlich, U., et al. 2006. Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood.* **108**:2545–2553.
22. Modlich, U., et al. 2008. Leukemia induction after a single retroviral vector insertion in Evi1 or Prdm16. *Leukemia.* **22**:1519–1528.
23. Bushman, F., et al. 2005. Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3**:848–858.
24. De Palma, M., et al. 2005. Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. *Blood.* **105**:2307–2315.
25. Cattoglio, C., et al. 2007. Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood.* **110**:1770–1778.
26. Aiuti, A., et al. 2007. Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. *J. Clin. Invest.* **117**:2233–2240.
27. Deichmann, A., et al. 2007. Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J. Clin. Invest.* **117**:2225–2232.
28. Schwarzwaelder, K., et al. 2007. Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *J. Clin. Invest.* **117**:2241–2249.
29. Bushman, F.D. 2007. Retroviral integration and human gene therapy. *J. Clin. Invest.* **117**:2083–2086.
30. Baum, C., et al. 1997. The potent enhancer activity of the polycythemic strain of spleen focus-forming virus in hematopoietic cells is governed by a binding site for Sp1 in the upstream control region and by a unique enhancer core motif, creating an exclusive target for PEBP/CBF. *J. Virol.* **71**:6323–6331.
31. Schambach, A., et al. 2000. Context dependence of different modules for posttranscriptional enhancement of gene expression from retroviral vectors. *Mol. Ther.* **2**:435–445.
32. Lund, A.H., et al. 2002. Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. *Nat. Genet.* **32**:160–165.
33. Serrano, M., et al. 1996. Role of the INK4a locus in tumor suppression and cell mortality. *Cell.* **85**:27–37.
34. Schmidt, M., et al. 2007. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat. Methods.* **4**:1051–1057.
35. Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A., and Copeland, N.G. 2004. RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res.* **32**:D523–D527.
36. Hosack, D.A., Dennis, G., Jr., Sherman, B.T., Lane, H.C., and Lempicki, R.A. 2003. Identifying biological themes within lists of genes with EASE. *Genome Biol.* **4**:R70.
37. Kratz, C.P., et al. 2007. Mutation analysis of Son of Sevenless in juvenile myelomonocytic leukemia. *Leukemia.* **21**:1108–1109.
38. Rogaia, D., et al. 1997. The localization of the HRX/ALL1 protein to specific nuclear subdomains is altered by fusion with its eps15 translocation partner. *Cancer Res.* **57**:799–802.
39. Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D.A. 2005. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature.* **436**:272–276.
40. Martinez-Glez, V., and Lapunzina, P. 2007. Sotos syndrome is associated with leukemia/lymphoma. *Am J. Med. Genet. A.* **143A**:1244–1245.
41. Wu, X., Luke, B.T., and Burgess, S.M. 2006. Redefining the common insertion site. *Virology.* **344**:292–295.
42. Zychlinski, D., et al. 2008. Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Mol. Ther.* **16**:718–725.
43. Kustikova, O.S., et al. 2007. Retroviral vector insertion sites associated with dominant hematopoietic clones mark "stemness" pathways. *Blood.* **109**:1897–1907.
44. Hacein-Bey-Abina, S., et al. 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science.* **302**:415–419.
45. Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. 2005. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature.* **436**:221–226.
46. Modlich, U., et al. 2005. Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. *Blood.* **105**:4235–4246.
47. de Ridder, J., et al. 2007. Co-occurrence analysis of insertional mutagenesis data reveals cooperating oncogenes. *Bioinformatics.* **23**:i133–i141.
48. Uren, A.G., et al. 2008. Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell.* **133**:727–741.
49. Celander, D., and Haseltine, W.A. 1984. Tissue-specific transcription preference as a determinant of cell tropism and leukaemogenic potential of murine retroviruses. *Nature.* **312**:159–162.
50. Ramezani, A., Hawley, T.S., and Hawley, R.G. 2008. Combinatorial incorporation of enhancer blocking components of the chicken {beta}-globin 5'HS4 and human T-cell receptor {alpha}/{delta} BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. *Stem Cells.* **26**:3257–3266.
51. Bokhoven, M., et al. 2009. Insertional gene activation by lentiviral and gammaretroviral vectors. *J. Virol.* **83**:283–294.
52. Schambach, A., Galla, M., Maetzig, T., Loew, R., and Baum, C. 2007. Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors. *Mol. Ther.* **15**:1167–1173.
53. Bauer, T.R., Jr., et al. 2008. Successful treatment of canine leukocyte adhesion deficiency by foamy virus vectors. *Nat. Med.* **14**:93–97.
54. Hu, J., et al. 2008. Reduced genotoxicity of avian sarcoma leukosis virus vectors in rhesus long-term repopulating cells compared to standard murine retrovirus vectors. *Mol. Ther.* **16**:1617–1623.
55. Izsvak, Z., and Ivics, Z. 2004. Sleeping beauty transposition: biology and applications for molecular therapy. *Mol. Ther.* **9**:147–156.
56. Ryu, B.Y., et al. 2008. An experimental system for the evaluation of retroviral vector design to diminish the risk for proto-oncogene activation. *Blood.* **111**:1866–1875.
57. Lombardo, A., et al. 2007. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**:1298–1306.
58. Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science.* **300**:764.
59. Ikeda, Y., Takeuchi, Y., Martin, F., Cosset, F.L., Mitrophanous, K., and Collins, M. 2003. Continuous high-titer HIV-1 vector production. *Nat. Biotechnol.* **21**:569–572.
60. Naldini, L., et al. 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science.* **272**:263–267.
61. Hadjantonakis, A.K., Gertsenstein, M., Ikawa, M., Okabe, M., and Nagy, A. 1998. Generating green fluorescent mice by germline transmission of green fluorescent ES cells. *Mech. Dev.* **76**:79–90.
62. Schmidt, M., et al. 2001. A model for the detection of clonality in marked hematopoietic stem cells. *Ann. N. Y. Acad. Sci.* **938**:146–155; discussion 155–146.
63. Schmidt, M., et al. 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.
64. Schmidt, M., et al. 2001. Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. *Hum. Gene Ther.* **12**:743–749.
65. Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**:e45.
66. Zhao, S., and Fernald, R.D. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J. Comput. Biol.* **12**:1047–1064.
67. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**:R19.