RESEARCH ARTICLE

# Lentiviral Vector Promoter is Decisive for Aberrant Transcript Formation

Simone J. Scholz,[1,†] Raffaele Fronza,[1,2] Cynthia C. Bartholomä,[1,§] Daniela Cesana,[3] Eugenio Montini,[3] Christof von Kalle,[1] Irene Gil-Farina,[1] and Manfred Schmidt[1,2,*]

[1]Department of Translational Oncology, German Cancer Research Center and National Center for Tumor Diseases, Heidelberg, Germany; [2]GeneWerk GmbH, Heidelberg, Germany; [3]San Raffaele Telethon Institute for Gene Therapy, Milan, Italy.
[†]Current address: Bristol-Myers Squibb GmbH & Co. KGaA, München, Germany.
[§]Current address: GeneWerk GmbH, Heidelberg, Germany.



Lentiviral vectors hold great promise for the genetic correction of various inherited diseases. However, lentiviral vector biology is still not completely understood and warrants the precise decoding of molecular mechanisms underlying integration and post-translational modification. This study investigated a series of self-inactivating (SIN) and full long terminal repeat (LTR) lentiviral vectors that contained different types of promoters with or without a transgene to gain deeper insights in lentiviral target site selection and potential perturbation of cellular gene expression. Using an optimized nonrestrictive linear amplification-mediated polymerase chain reaction (nrLAM-PCR) protocol, vector structure-dependent integration site profiles were observed upon transduction of mouse lin$^-$ hematopoietic progenitors *in vitro*. Initial target site selection mainly depended on the presence of the promoter while being independent of its nature. Despite the increased propensity for read-through transcription of SIN lentiviral vectors, the incidence of viral–cellular fusion transcript formation involving the canonical viral splice donor or cryptic splice sites was reduced in both unselected primary lin$^-$ cells and transformed 32D cells. Moreover, the strength of the internal promoter in vectors with SIN LTRs is decisive for *in vitro* selection and for the abundance of chimeric transcripts, which are decreased by moderately active promoters. These results will help to better understand vector biology and to optimize therapeutic vectors for future gene therapy applications.

**Keywords:** lentiviral vectors, integration site, fusion transcripts, genotoxicity

## INTRODUCTION

RETROVIRAL VECTORS ARE EFFICIENT TOOLS for the long-term correction of genetic disorders due to their stable integration into the host genome.[1–4] At the same time, integration is a critical factor for their genotoxic potential, which is a major hurdle on the way to retroviral gene therapy becoming a standard treatment. The site of integration depends on numerous factors, for example cell type, chromatin structure, and various vector elements.[5–8] A major determinant herein is the genus of the parental virus with its respective integrase.[9] Gammaretrovirus-derived vectors favor transcription start sites and active regulatory elements, whereas lentiviral vectors prefer gene-coding regions of actively transcribed genes.[10–13] The integration pattern of gammaretroviral vectors in combination with strong enhancer/promoter elements within the long terminal repeat (LTR) has been shown to confer an increased risk of insertional activation of proto-oncogenes triggering severe side effects in several clinical trials.[14–17] In this respect, the lentiviral integration pattern is supposed to be superior, while self-inactivating (SIN) LTRs and cell type-specific and moderate internal promoters further reduce the risk of insertional mutagenesis.[18–20] However, the influence of these functionally active vector internal elements on target site selection is still unclear and remains to be elucidated.

*Correspondence: Dr. Manfred Schmidt, Department of Translational Oncology, National Cancer Institute and German Cancer Research Center, Im Neuenheimer Feld 581, Heidelberg, 69120, Germany. E-mail: manfred.schmidt@nct-heidelberg.de



DOI: 10.1089/hum.2017.162

Recently, clinical and preclinical studies revealed that lentiviral integration can be associated with side effects other than enhancer-mediated gene activation, namely the formation of aberrant transcripts. SIN lentiviral vectors display an increased frequency of read-through transcription compared to their wild-type (WT) LTR counterpart.[21,22] Subsequent aberrant splicing can result in viral–cellular fusion transcripts, which may deregulate cellular gene expression and potentially promote unwanted effects. In a clinical gene therapy trial for $\beta$-thalassemia, lentiviral vector-induced aberrant splicing stabilized a truncated *HMGA2* transcript, which caused clonal dominance.[23] This was associated with transfusion independence but no adverse effects. On the contrary, a truncated transcript generated by the fusion of a lentiviral vector with a cellular proto-oncogene contributed to tumor formation in a mouse model.[19] Chimeric transcript formation might also lead to gene inactivation due to downregulation of the full length messenger, as has been recently observed in a murine study. Intronic integration and subsequent aberrant splicing of a lentiviral gene marking vector induced haploinsufficiency of *Ebf1*, a known tumor suppressor gene, thereby causing leukemia.[24] These results emphasize the need for a deeper understanding of the fundamental interplay between viral vectors and the host-cell transcription machinery and its consequences.[25]

Recent reports showed that chimeric transcript formation is a very common event that likely depends on the cell type and LTR configuration.[26,27] Thus, the exact functions and interactions of viral and cellular elements during integration and transcription are not yet resolved. Vector elements such as the LTR (full or SIN), enhancer/promoter, and transgene are the most commonly changed features during vector design.[28] This work examined the influence of these functional vector elements on lentiviral integration and cellular gene expression by high-resolution genome- and transcriptome-wide analysis. The results highlight the importance of the internal promoter in SIN lentiviral vectors during target site selection and in the context of chimeric transcript formation. The quantification of these transcripts will help to improve lentiviral vector design and eventually the safety of gene therapeutic applications.

## MATERIALS AND METHODS

### Lentiviral vector production and titration

VSV-G pseudotyped lentiviral vectors were produced by transient four-plasmid co-transfection of 293T cells using polyethylenimine, harvested 2 days after transfection and concentrated by ultracentrifugation. The vector titer was determined on HeLa cells by flow cytometry 72 h after infection with serial dilutions of the viral suspension.

### Isolation and transduction of hematopoietic progenitor cells

Mouse bone marrow was isolated from the femur and tibia by flushing with phosphate-buffered saline with 2% fetal bovine serum (FBS; Invitrogen) and lin$^-$ cells were purified by lineage-marker negative selection using the Enrichment of Murine Hematopoietic Progenitors kit (STEMCELL Technologies, Inc.). Cells were plated at a density of $1\times10^6$ cells/mL and cultured for 24 h in StemSpan™ serum-free expansion medium (STEMCELL Technologies, Inc.) supplemented with cytokines (100 ng/mL of stem cell factor [SCF], 100 ng/mL of Flt3-ligand, 100 ng/mL of thrombopoietin [TPO], and 20 ng/mL of interleukin [IL]-3; PeproTech). Lin$^-$ cells were then subjected to mock or lentiviral vector transduction and, 18 h post-infection, cells were washed and maintained in RPMI supplemented with 10% FBS, 100 IU/mL of penicillin, 100 mg/mL of streptomycin, 20 ng/mL of SCF, 20 ng/mL of Flt3-l, 20 ng/mL of TPO, and 4 ng/mL of IL-3 for up to 60 days.

### DNA and RNA isolation, cDNA synthesis, and vector copy number determination

Genomic DNA was isolated using the High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's instructions. Total mRNA was extracted with the RNeasy MiniKit (Qiagen), including the optional on column DNA digestion step. Following a rigorous DNase treatment using the Ambion™ TURBO DNA-free™ kit (Life Technologies), cDNA was synthesized with the Verso™ cDNA Kit (Thermo Scientific). A control reaction lacking the reverse transcriptase (–RT) was performed for subsequent nonrestrictive linear amplification-mediated polymerase chain reaction (nrLAM-PCR) analyses. To further exclude DNA contamination, a reverse transcription (RT) PCR was performed with *Hprt* specific primers annealing in different exons (data not shown). Vector copy numbers (VCNs) were determined by quantitative PCR, with primers annealing within *Hprt* and within the vector: 5′-GGGCATCTACTTACAGTCTGG/5′-GAAGCTGT CCTTATTCTGAATGC and 5′-GAGCTCTCTGG CTAACTAGG/5′-GCTAGAGATTTTCCACACTG, respectively.

### *In vitro* genotoxicity assay

32D cells (ATCC) were cultured in RPMI supplemented with 10% fetal calf serum, 1% penicillin/

streptomycin, and 5% IL-3. Transduction was performed at a density of $5\times10^5$ cells/mL in the presence of polybrene followed by an expansion period of several days. Colony assays were executed as described previously.[29] Transduction efficiency and vector presence in selected colonies were determined by flow cytometry and PCR, respectively (primers: 5′-AAACGGCCACAAGTTCA GCG and 5′-TCTTCTGCTTGTCGGCCATG).

### Integration site and chimeric transcript analysis

Integration sites (IS) were analyzed by nrLAM-PCR using an optimized protocol (see below) for increased sensitivity and accuracy. Read-through and fusion transcripts were detected via LAM- or nrLAM-PCR. (nr)LAM-PCR, 454 sequencing, and bioinformatical data analysis were performed as described before,[30–33] with the following adjustments in the nrLAM-PCR protocol. After magnetic capture of the linear PCR product, beads were washed and resuspended in 10 $\mu$L of mastermix for linker cassette ligation using CircLigase™ ssDNA Ligase (Epicentre): 1.5 $\mu$L of water, 1 $\mu$L of 10$\times$CircLigase™ reaction buffer, 1 $\mu$L of ssDNA linker (10 pmol/$\mu$L), 0.5 $\mu$L of $MnCl_2$ (50 mM), 0.5 $\mu$L of ATP (1 mM), and 5 $\mu$L of 50% PEG 8000 (preheated for 5 min to 60°C). After 1 h of incubation at 60°C, the reaction was stopped by addition of 90 $\mu$L of $H_2O$. Subsequently, DNA–bead complexes were washed and resuspended in 10 $\mu$L of $H_2O$. Exponential amplification by nested PCR and 454 sequencing were performed as previously described.[31] Primer sequences for read-through and fusion transcript analyses were as follows: 5′-bio-AGTAGTGTGTGCCCGTCTGT, 5′-bio-GTGTGACTCTGGTAACTAGAG, 5′-GATCCCT CAGACCCTTTTAGTC and 5′-bio-AGTAGTGTGTG CCCGTCTGT, 5′-bio-CAGTGGCGCCCGAACAG, 5′-GAAAGCGAAAGGGAAACCAG, respectively. A –RT control (see above) was analyzed in addition for each sample as a control for DNA contamination.

### Clonality and comparative common integration site analysis

The Shannon value $H_j$ for the sample $j$ is defined as $H_j=\Sigma_i f_{ji} \log f_{ji}$, where $f_{ji}$ is the reads fraction that contributes to the IS $i$ in the sample $j$, whereas the Simpson index $S_j$ is defined as $S_j=1/\Sigma_i {f_{ji}}^2$. The pairwise comparison of the common integration site (CIS) profile was performed to evaluate the similarity of the samples. The CIS definition was as follows: a CIS of 2nd order comprises two IS within 30 kb, a CIS of 3rd order comprises three IS within 50 kb, a CIS of 4th order comprises four IS within 100 kb, and a CIS of 5th and higher order comprises five or more IS within 200 kb.[34] Each combination of the insertion site set of sample $i$ and $j$ were analyzed in order to obtain a mixed CIS profile. Normalized entropy $N$ between 0 and 1 was then computed for each CIS >4th order in the mixed profile. The normalized entropy value for the CIS $l$ in the mixed profile of the sample $i$ and $j$ is defined as $N_{ijl}=(\Sigma_k p_{ijlk})/\log 2$, where $p_{ijlk}$ is the IS fraction belonging to the sample $k$. Then, all those CIS with $N_{ijl} > 0.5$ were defined as shared. The fraction of shared CIS with respect to the total number of CIS was used as a measure of association between the two samples. The values together form a $5\times5$ association matrix, with values ranging from 0 (no shared CIS) to 1 (all CIS are shared). The matrix was then loaded into R suite and converted to a dissimilarity table. Briefly, for each element $e_{ij}$ in the matrix, with $i=1, 2, \ldots, 5$, a dissimilarity table was constructed where the dissimilarity values $d_{i,j}=1-e_{ij}$ are used as a distance in order to perform a hierarchical clustering. The tree structure is then obtained by plotting the results returned from the R function hclust().

## RESULTS

For high-resolution IS analyses, nrLAM-PCR performance was substantially improved in terms of sensitivity and efficiency (Supplementary Fig. S1 and S2; Supplementary Data are available online at www.liebertpub.com/hum). It was hypothesized that the prolonged culture of bulk-infected cells and subsequent in-depth analysis of integration profiles would enable the dissection of initial target site selection and the detection of potential *in vitro* clonal selection, thus allowing the assessment of the consequences of vector design on a high number of transduced cells at a given time point. To define the influence of viral vector elements on target site selection, we investigated the IS profiles in lineage negative ($lin^-$) murine stem and progenitor cells transduced with a series of lentiviral vectors that differed in LTR configuration, promoter, and/or transgene (Fig. 1). Infected cells were cultured for up to 60 days, and VCNs were determined 14 days post transduction to range between 2 and 10.4 (Table 1).

First, nrLAM-PCR followed by 454 pyrosequencing was performed on $lin^-$ cells 14, 30, 45, and 60 days after lentiviral transduction. In total, the study identified 5,159 (LV.SF.LTR), 9,414 (SIN.LV.SF), 8,461 (SIN.LV.PGK), 5,033 (LV.W), and 6,719 (LV.WET) uniquely mappable IS (Table 2). Diversity analyses based on the number of IS, VCN, and transduction efficiency showed a highly polyclonal pattern for all vectors 14 days

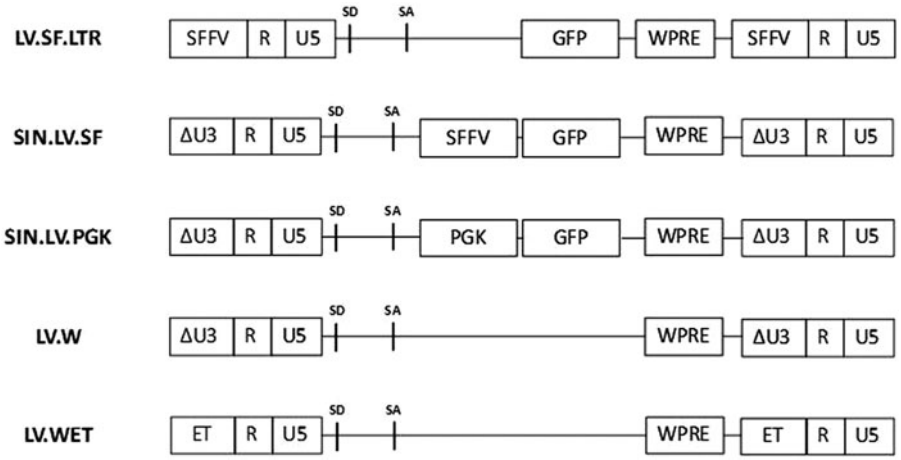


**Figure 1.** Lentiviral vector constructs employed to assess vector architecture's effect on integration site and clonal selection. SFFV, enhancer/promoter of the spleen focus forming virus; SD, splice donor; SA, splice acceptor; GFP, green fluorescent protein; WPRE, woodchuck hepatitis virus post-transcriptional regulatory element; ΔU3, deletion of U3; PGK, phosphoglycerate kinase promoter, ET, liver-specific transthyretin promoter.



post-transduction. Over time, emergence of individual cell clones and loss of diversity was observed in cells transduced with the full SFFV LTR-driven vector LV.SF.LTR (Fig. 2A). Sample heterogeneity was assessed by two diversity indexes, the Shannon index and the Simpson index.[35] This confirmed a clear reduction in both clonality (Shannon index; Fig. 2B) and homogeneity (Simpson index; Fig. 2C) in LV.SF.LTR-transduced cells 30–45 days post-transduction. In contrast, SIN lentiviral vectors and a vector with a liver-specific LTR promoter did not show dominant clones and only exhibited a slight diversity reduction. All vectors showed the typical lentiviral integration pattern, with $62.6 \pm 0.9\%$ of all integrations occurring in-gene and no obvious differences in the global chromosomal distribution.

To obtain a deeper insight into lentiviral target site selection, comparative analyses of common integration sites (CIS) were performed at an early time point after transduction (48 h). By LAM-PCR and subsequent 454 sequencing 1,200 (LV.SF.LTR), 899 (SIN.LV.SF), 1,051 (SIN.LV.PGK), 847 (LV.W), and 1,226 (LV.WET) IS were identified. Comparable data sets were obtained by extracting equal numbers of IS from each integration site pool. Comparative CIS analyses were performed pairwise, with the proportion of shared CIS being representative for the similarity of their IS profiles. Based on the ratio of shared and unique CIS, a phylogenetic tree was created. All vectors comprising a professional promoter formed one common cluster, while the only vector lacking promoter sequences did not integrate in this cluster (Fig. 3A). The distance of the cluster to the empty vector without promoter and transgene (LV.W) was relatively high compared to the distances between vectors within the cluster (height 1.0 vs. 0.01–0.2). In contrast, comparative CIS analysis 14–60 days post-transduction showed that the further development of the integration site pool is mainly dependent on the nature and strength of the promoter (Fig. 3B). Both vectors with a strong SFFV promoter in an internal position or in the LTR (SIN.LV.SF and LV.SF.LTR) constituted one cluster, while another cluster was formed by two vectors with a weak or without a promoter (SIN.LV.PGK

**Table 1.** *nrLAM and vector copy number analysis performed 14 days post transduction on lineage negative (lin⁻) murine stem and progenitor cells transduced with a series of lentiviral vectors*

| *Vector* | *Raw sequences* | *Sequences* | *IS* | *VCN* | *Clones* | *Seqs/IS* |
|---|---|---|---|---|---|---|
| LV.SF.LTR | 15,744 | 4,554 | 2,463 | 4.6 | 535 | 1.8 |
| SIN.LV.SF | 16,878 | 5,570 | 3,454 | 4.6 | 751 | 1.6 |
| SIN.LV.PGK | 11,651 | 3,721 | 2,531 | 10.4 | 243 | 1.5 |
| LV.W | 26,604 | 8,293 | 2,597 | 2.0 | 1,299 | 3.2 |
| LV.WET | 13,427 | 4,466 | 2,764 | 6.2 | 446 | 1.6 |

nrLAM, nonrestrictive linear amplification-mediated; IS, integration sites; VCN, vector copy number.

**Table 2.** *Integration site analysis performed on lineage negative (lin⁻) murine stem and progenitor cells at different time points post transduction with different lentiviral vectors*

| *Time point* | *14 days* | | *30 days* | | *45 days* | | *60 days* | | *Total* |
|---|---|---|---|---|---|---|---|---|---|
| *Vector* | *No. Seq* | *IS* | *No. Seq* | *IS* | *No. Seq* | *IS* | *No. Seq* | *IS* | *IS* |
| LV.SF.LTR | 4,554 | 2,463 | 4,802 | 1,957 | 5,203 | 772 | 4,574 | 1,040 | 5,159 |
| SIN.LV.SF | 5,570 | 3,454 | 5,531 | 2,848 | 5,619 | 2,754 | 8,474 | 2,938 | 9,414 |
| SIN.LV.PGK | 3,721 | 2,531 | 5,304 | 2,727 | 5,581 | 2,857 | 5,532 | 2,400 | 8,461 |
| LV.W | 8,293 | 2,597 | 5,938 | 2,025 | 7,407 | 1,706 | 8,466 | 1,655 | 5,033 |
| LV.WET | 4,466 | 2,764 | 4,884 | 2,690 | 4,318 | 2,232 | 5,257 | 2,114 | 6,719 |

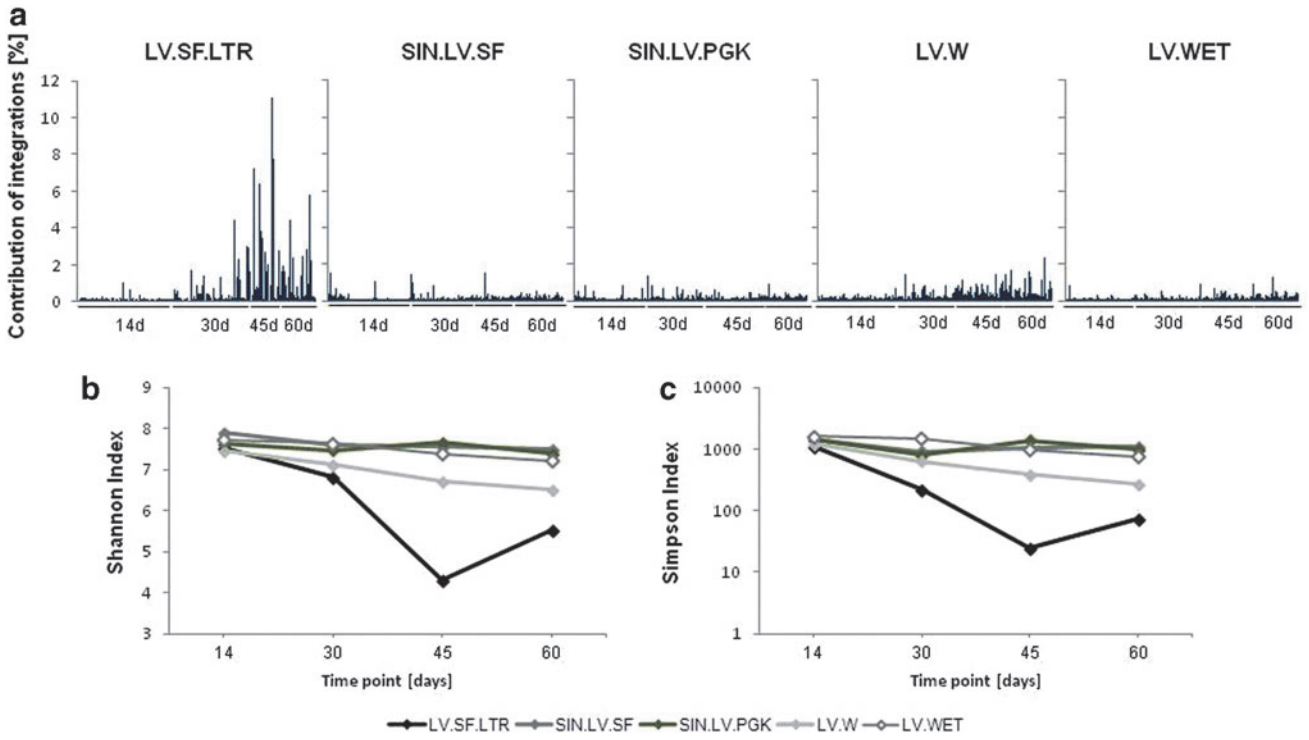


**Figure 2.** Clonality of transduced cell populations over time. **(A)** Contribution of single integration clones at various time points after transduction. Clonal contribution was assessed using 454/Roche and MiSeq/Illumina sequence counts. Clonal skewing was only observed with LV.SF.LTR-transduced cells. **(B** and **C)** Assessment of sample diversity using **(B)** Shannon and **(C)** Simpson indexes. The diversity of LV.SF.LTR-transduced cells dropped down 30–45 days post transduction. d, days; LTR, long terminal repeat.

and LV.W). The vector bearing a tissue-specific promoter (LV.WET) showed the lowest similarity in the selected IS profile. The prolonged culture of bulk cells and subsequent analyses with high-resolution nrLAM-PCR enabled vector-driven clonal selection to be measured in an easy and fast *in vitro* setting.

Next, the study aimed to understand how vector structure and elements affect the incidence of viral–cellular read-through transcription, abnormal splicing, and the characteristics of chimeric transcripts. Thus, aberrant transcript formation was investigated in a transcriptome-wide fashion using nrLAM-PCR on cDNA from murine lin$^-$ cells 14 days after transduction. Functional analyses were conducted with the vectors LV.SF.LTR, SIN .LV.SF, and SIN.LV.PGK, showing that at differ-

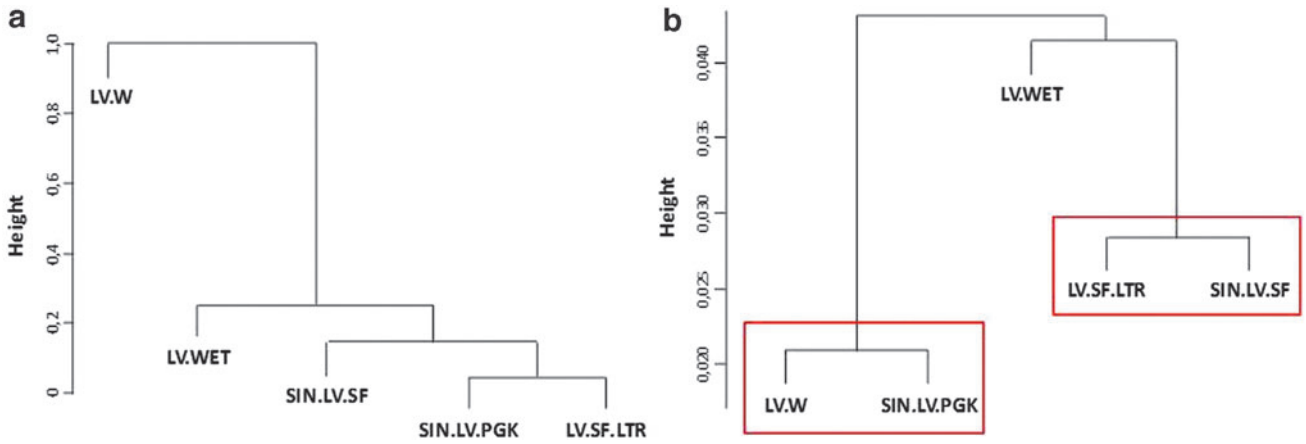


**Figure 3.** Comparative analysis of integrations site profiles of different lentiviral vector constructs at different time points post transduction. To determine the relationship of the vectors, comparative CIS analyses were performed pairwise, and dendrograms were generated based on the percentage of common CIS, with the height representing the distance between the vectors. **(A)** All promoter-containing vectors clustered together 48 h post transduction. **(B)** After 14–60 days post infection, cluster formation was observed according to promoter strength. CIS, common integration site.

ent frequencies, all three vectors may give rise to chimeric read-through (Fig. 4A) and fusion (Fig. 4B) transcripts. High-throughput sequencing revealed 351 (LV.SF.LTR), 674 (SIN.LV.SF), and 74 (SIN.LV.PGK) viral–cellular read-through transcripts (Fig. 4C).

To answer the question of whether the genomic region surrounding the insert has an influence on the frequency of read-through transcription, the occurrence of transcripts in gene coding and non-coding regions was compared with the IS distribution. All vectors showed a significantly higher read-through rate within gene coding regions, mainly in introns ($75.1 \pm 3.1\%$ vs. $62.4 \pm 0.7\%$; $p = 2.9 \times 10^{-13}$, $\chi^2$ test). Furthermore, transcripts containing exonic sequences were significantly enriched with SIN.LV.PGK compared to SIN.LV.SF and LV.SF.LTR (9.5% vs. $5.1 \pm 0.1\%$; $p = 9.9 \times 10^{-2}$–$1.5 \times 10^{-3}$, $\chi^2$ test; Fig. 5A).

Despite the high rate of read-through transcription, only rare chimeric fusion transcripts were observed: eight with LV.SF.LTR, three with SIN.LV.SF, and none with SIN.LV.PGK (Fig. 4D). Even though LTR-induced transcription is largely abrogated with the SIN-LTR design, full-length vector transcripts have been observed with both gammaretro- and lentiviral SIN vectors.[36,37] Since the lentiviral splice site is located upstream of the internal promoter, the fusion transcripts detected in the SIN.LV.SF setting might result from LTR-directed or cellular read-through transcription.

To test whether these transcripts could arise from an increased SIN LTR residual activity due to the potent SFFV promoter at the internal position, the amount of LTR-induced transcripts generated by SIN.LV.SF and SIN.LV.PGK was quantified by qRT-PCR. With this approach, it was found that full-length transcripts are 16 times more frequent with the strong SFFV promoter than with the weak PGK promoter (Fig. 5B and C). To assess the contribution of vector-induced fusion transcripts to the biological safety of lentiviral vectors, an *in vitro* genotoxicity assay was employed. Herein, the genotoxic potential of a vector is measured by the functional

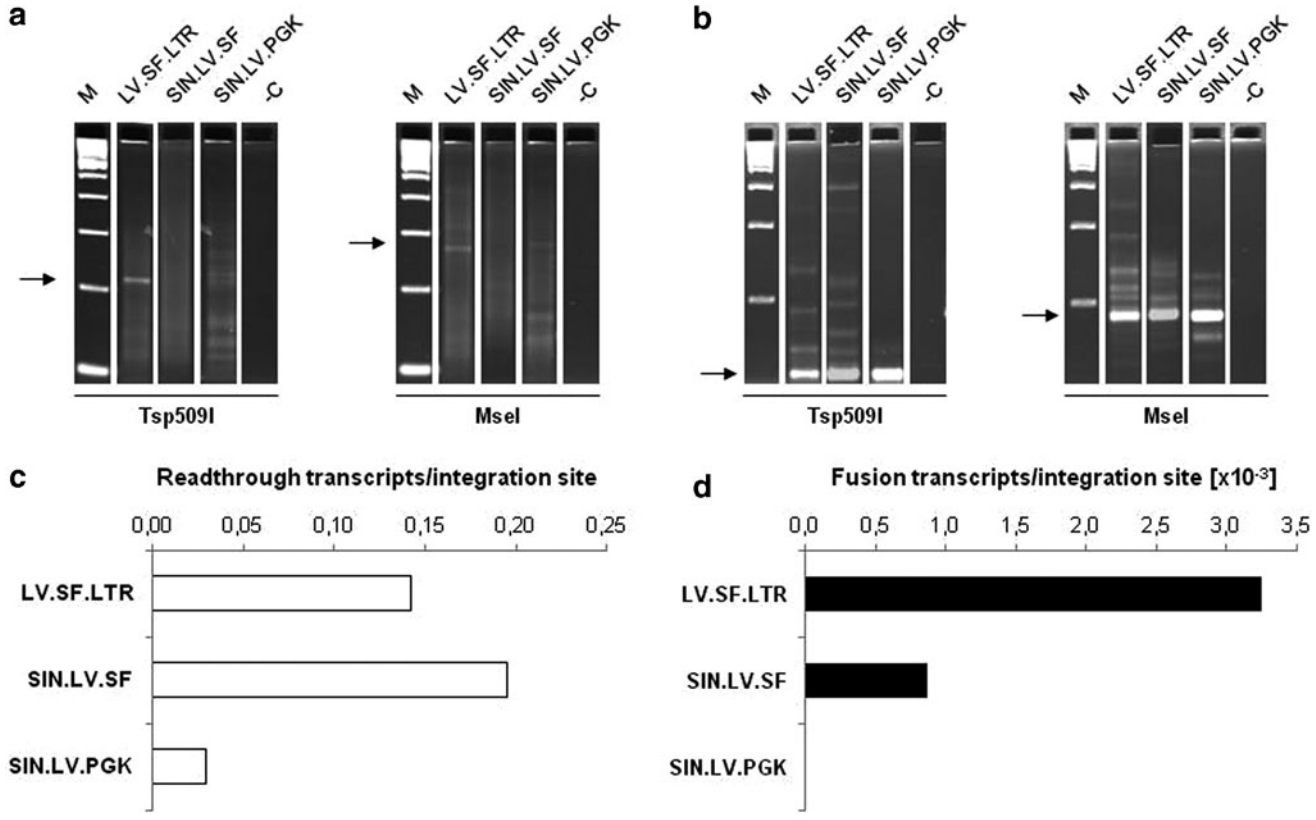


**Figure 4.** Aberrant transcript formation of lentiviral vectors. 3′ LAM-PCR analysis was performed on 150 ng of cDNA of lentiviral transduced lineage negative cells for the detection of chimeric **(A)** read-through and **(B)** fusion transcripts. *Arrows* indicate the size of the internal control fragment. Abundant read-through transcription was observed with all vectors, while only a small fraction resulted in fusion transcripts. Number of chimeric read-through **(C)** and fusion **(D)** transcripts detected per integration site. –C, RNA control (cDNA synthesis without reverse transcriptase); LAM-PCR, linear amplification-mediated PCR; cDNA, complementary DNA.

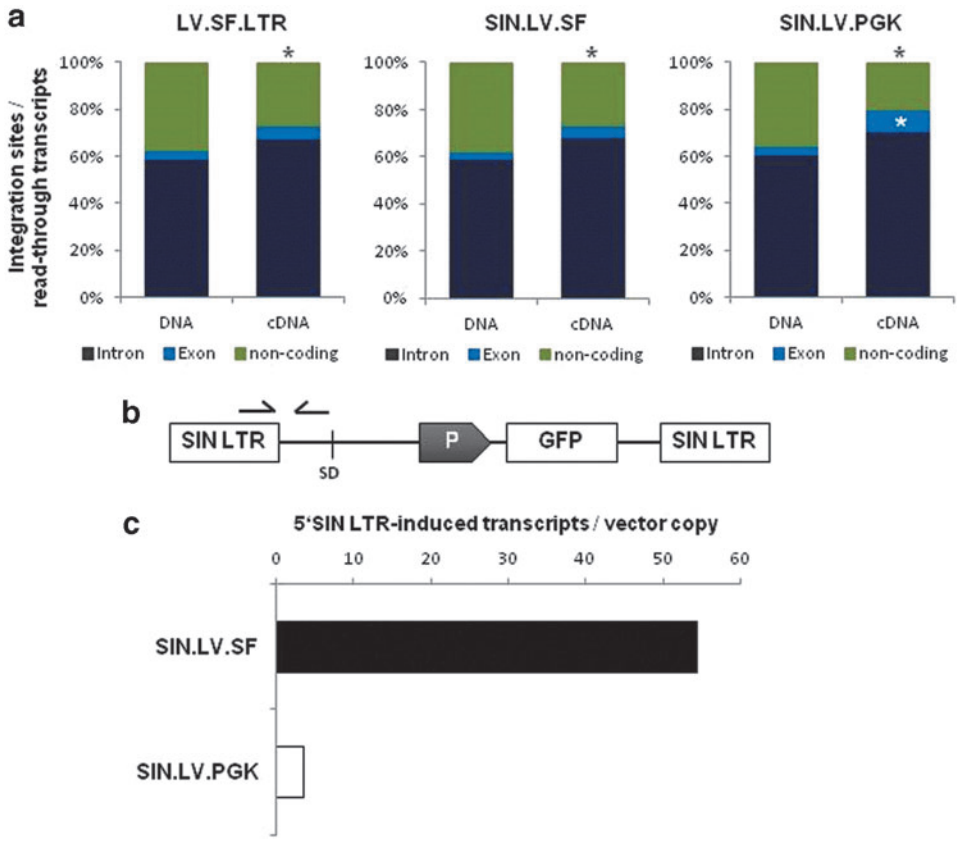


**Figure 5.** Distribution of viral–cellular read-through transcripts. **(A)** Comparison of the distribution of vector integration sites and read-through transcripts. Asterisks indicate significant *p*-values, which were determined by chi-square test. **(B** and **C)** Frequency of 5′ SIN LTR-induced transcripts. The number of transcripts was analyzed by qPCR and normalized to the number of vector copies per cell. The location of primer sets used for qPCR is depicted in **(B)**. SIN, self-inactivating; qPCR, quantitative PCR.

transformation of a factor-dependent cell line.[29] The transformation potential of the LV.SF.LTR vector was determined at $19.5 \pm 3.8 \times 10^{-6}$ per vector copy, whereas the ones of the SIN vectors were 20 and even 50 times lower ($1.0 \pm 0.5 \times 10^{-6}$ for SIN.LV.SF and $0.4 \pm 0.4 \times 10^{-6}$ for SIN.LV.PGK; Fig. 6A). For each vector, several transformed colonies were randomly picked and expanded for further functional analyses (12 LV.SF.LTR, 12 SIN.LV.SF, and all [3] SIN.LV.PGK colonies). In total, 55 LV.SF.LTR, 71 SIN.LV.SF, and 30 SIN.LV.PGK integrants were identified, with at least one being located in the 5′ region of the growth hormone receptor (*Ghr*) in all LV.SF.LTR and 83% of SIN.LV.SF colonies, but not in SIN.LV.PGK colonies. *Ghr* fusion transcripts were shown to be the major cause for the functional transformation of IL-3-dependent cells by lentiviral vectors.[38,39] The transcriptome-wide analysis of fusion transcripts using the optimized nrLAM-PCR protocol on cDNA revealed that almost 60% of LV .SF.LTR, 25% of SIN.LV.SF, and 3% of SIN.LV.PGK integrants led to the formation of fusion transcripts (Fig. 6B). Genes involved in fusion transcript formation mainly played a role in protein synthesis, cell proliferation, and signal transduction. *Ghr* fusion transcripts were detected in all LV.SF.LTR (12/12) and some SIN.LV.SF (4/12) colonies (Fig. 6C). These transcripts were generated from integrants located within a 25 kb window directly upstream of exon 2, while integrants >25 kb upstream or downstream of exon 2 did not give rise to fusion transcripts. Gene expression analysis showed an upregulation of *Ghr* in almost all colonies transduced with the two SFFV vectors. In contrast, SIN.LV.PGK colonies did not show any upregulation (Fig. 6D).

## DISCUSSION

Vector integration is the key event for the success of retroviral gene therapy and its long-lasting therapeutic effect, but also represents a potential cause for the development of adverse events due to insertional mutagenesis. Major determinants for the genotoxic profile of retroviral vectors are the site of integration and their ability to deregulate the

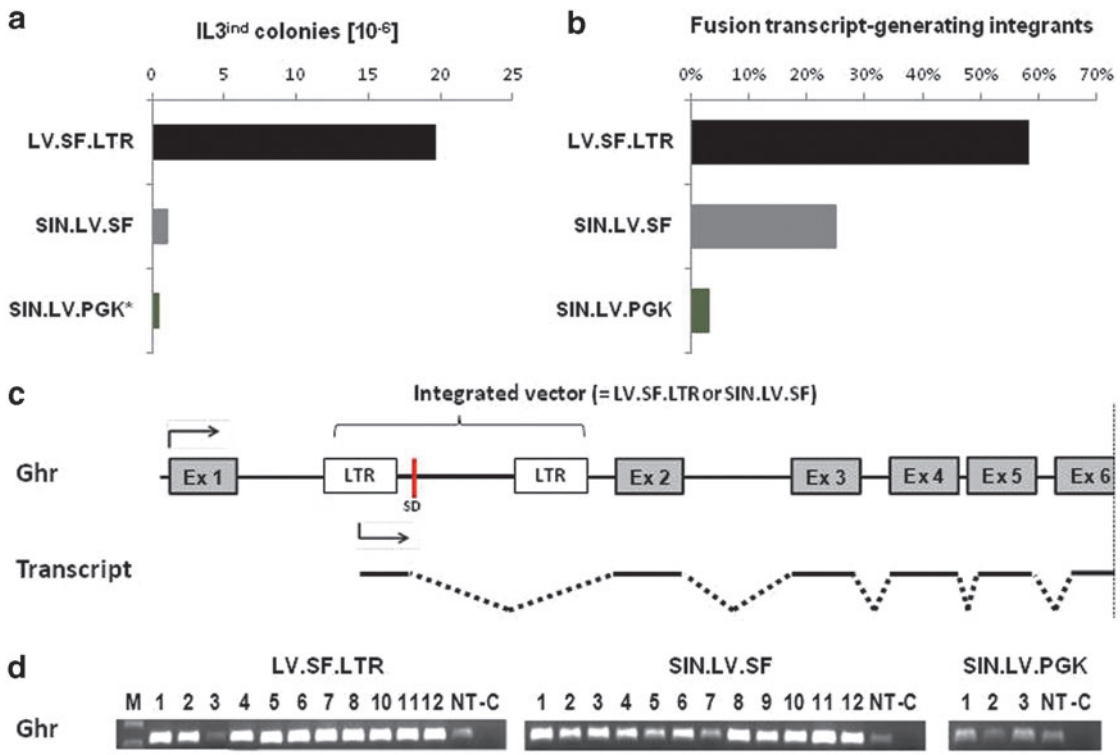


**Figure 6.** Genotoxic potential of lentiviral vectors. **(A)** Frequency of interleukin (IL)-3-independent colonies. The genotoxic potential was assessed by functional transformation of IL-3-dependent 32D cells upon transduction with the various vectors. **(B)** Proportion of integrants leading to fusion transcripts in transformed colonies. IL-3-independent colonies were expanded and analyzed for integration sites and fusion transcripts by nrLAM-PCR on DNA/cDNA. **(C)** Scheme of the vector insertions within *Ghr* and the resulting chimeric transcripts. Vector integration occurred in the first intron of *Ghr*, leading to chimeric transcripts that were initiated in the 5′ LTR of the vector. Read-through transcription and subsequent aberrant splicing resulted in the fusion of the viral splice donor (SD) to the cellular splice acceptor of exon 2 (Ex2). **(D)** Gene expression analysis of *Ghr* in transformed colonies. *Ghr* expression was measured by RT-PCR using primers that anneal to exon 2 and 4, thus measuring both normal and chimeric *Ghr* transcripts. *SIN.LV.PGK colonies could only be expanded after re-addition of IL3. M, 100 bp marker; NT, non-transduced control; –C, water control.



expression of surrounding genes. This study investigated a series of lentiviral vectors using functional integrome analysis to determine the influence of vector structure elements, such as LTR configuration or the promoter, on target site selection, clonal development, transcription of cellular genes, and the consequences for vector biosafety.

So far, standard *in vitro* IS analyses concentrate on late time points after transduction, thus hampering the determination of the vector's initial target site selection preferences and the ability to differentiate them from integration features that are rather due to clonal selection. With this aim in mind and to enable the detection of potential clonal selection processes taking place over time, it is mandatory to analyze early time points when no remarkable clonal selection has yet occurred. This study scrutinized the integration profiles of lentiviral vectors differing in LTR configuration, promoter, and presence of a transgene in hematopoietic stem and progenitor cells (HSPCs) 48 h after transduction. The results show that LTR configuration and the presence of a transgene play minor roles during target site selection. In contrast, early IS profiles are highly influenced by the presence of a promoter, independently of its nature and activity. The preferential integration of MLV-derived vectors in gene regulatory regions is mediated by the binding of transcription factors to the pre-integration complex and cellular transcription factor binding sites (TFBS), while human immunodeficiency virus type 1 (HIV-1)-derived vectors do not show such a bias, but rather an underrepresentation of TFBS around the IS. However, this negative bias is removed when the HIV-1 U3 is replaced by the U3 region of MLV.[40] Altogether, this suggests that ectopic promoter sequences have a major impact on the target site selection of lentiviral vectors, most likely due to transcription factor-mediated interactions, such as described for LEDGF, which are not present in promoter-less or WT HIV-1 LTR containing vectors.

When the early 48 h IS profiles were compared with later time points at 1,460 days, an obvious over time clonal selection of cells transduced with the complete SFFV LTR-driven lentiviral vector was

observed. In addition, high-resolution IS analysis and comparative CIS analysis also demonstrated that a SIN vector with an internal strong (SFFV) promoter may have a certain albeit low potential to induce clonal selection. In agreement with this, the SIN.LV.SF vector exhibited a higher genotoxic potential when compared to the SIN.LV.PGK in an *in vitro* genotoxicity assay.[29] Overall, the clustering of IS profiles determined by comparative CIS analysis occurred according to the promoter and its respective activity in HSPCs. This indicates the decisive role of promoter strength and/or nature even in context of SIN LTR vectors. The genotoxic potential of LV.SF.LTR was also illustrated in a murine mouse model, leading to an acceleration of tumor onset and biased integration site profiles *in vivo* and *in vitro*. This study shows that, if any, SIN.LV.SF vector's genotoxicity is substantially reduced and below detection limit, probably due to background oncogenesis of Cdkn2a$^{-/-}$ HSPCs and the too short (2 weeks) *in vitro* culture.[19] Here, it has been shown that the prolonged culture of transduced cells and meticulous comparative integration site analysis enable the detection of even subtle clonal selection processes in bulk cell populations and thus may provide a good prospect for early biosafety measurements of new therapeutic vectors.

Using nrLAM-PCR based transcriptome analysis of unselected vector-transduced bulk cells, the study shows that both full and SIN LTR vectors generate substantial amounts of read-through transcripts, though the frequency is clearly reduced with SIN LTR configurations and weak internal promoter usage. As 70–80% of the termination signals reside within the first 124 nucleotides of the HIV-1 transcriptional control region overlapping enhancer and promoter sequences,[41] the high read-through activity of a full SFFV LTR-driven vector observed in this study points out that an ectopic U3 region, albeit placed within the LTR, cannot completely compensate for the deletion of these HIV-1 U3 inherent termination signals.

Most read-through transcripts were detected within gene-coding regions. Genomic sequences and/or chromatin structure can influence the expression of the vector.[42] The open form of the chromatin and the increased transcriptional activity in the area of actively transcribed genes might lead to higher levels of viral–cellular read-through transcripts. Another possibility is the mechanism of differential regulation of the identical 5′ and 3′ LTRs. In WT HIV-1, the polyadenylation signal in the 5′ LTR is suppressed by the canonical splice donor. The introduction of an intron downstream of the 3′ LTR can interfere with this regulation, thereby weakening the poly(A) site at this position.[43] In addition, the chimeric transcripts detected in this setting can originate from the vector as well as from cellular genes. Read-through transcription from cellular genes therefore increases the amount of chimeric transcripts containing gene-coding sequences. With a vector harboring a weak promoter, the ratio of vector-induced transcripts is further decreased. Vector sequences within introns of cellular transcripts will be spliced out to a large extent by the normal cellular RNA processing, while vector sequences within exons will remain in the transcript. This may result in a higher ratio of exon-containing chimeric transcripts, as observed with SIN.LV.PGK.

The lack of transcriptional termination in SIN lentiviral vectors poses an increased risk of activating cellular oncogenes downstream of the viral integration site.[44] This study shows that the activity of the internal promoter has a significant impact on the incidence of chimeric read-through and fusion transcripts. Moreover, these transcripts partially contribute to the genotoxic potential of lentiviral vectors. This is in agreement with a study by Bokhoven *et al.*, reporting that the major mechanism of mutagenesis used by lentiviral vectors involves vector-mediated aberrant splicing.[38] The identification of cryptic splice sites within the vector construct allows the recoding of these sequences and reduction of residual aberrant splicing.[45] However, aberrant transcript formation cannot be stalled completely,[26] and the viral canonical splice donor cannot be recoded due to its important role in the regulation of the termination and polyadenylation signals within the 5′ LTR.[43] Mutations in this sequence indeed lead to a drastic decrease of virus titer and viral gene expression.[46] Thus, this study concentrated its investigation of aberrantly spliced mRNAs on the canonical splice donor and surrounding cryptic splice sites. In lin$^{-}$ and IL-3$^{-}$ dependent murine cells, fusion transcripts were yielded utilizing both canonical and cryptic splice sites.

Fusions with exon-like sequences suggest the formation of new transcripts upon integration of the lentiviral vector, potentially with growth promoting features. Primary HSPCs contained only few chimeric fusion transcripts, which were solely formed under utilization of the canonical lentiviral splice donor. In contrast, up to 100% of transformed 32D clones generated aberrant transcripts. These results coincide with a study showing that the frequency of aberrant splicing, as well as the utilization of different splice sites, strongly varies, depending on the cell line and selection conditions.[27] The accumulation of

aberrant transcripts in the cell is, to some extent, prevented by nonsense-mediated mRNA decay (NMD).[27] Indeed, despite high frequencies of read-through transcription, spliced chimeric transcripts were hardly detected.

It was found that the activity of the internal promoter in SIN vectors critically impacts the incidence of read-through and fusion transcripts. This is likely due to the higher transcription rate conferred by a strong promoter such as SFFV. With the increasing amount of transcripts that have to be processed, the probability of a mistake in this process, for example the recognition of termination and splice signals, rises. Yet, an additional factor has to account for the differential formation of viral-cellular fusion transcripts using splice sites upstream of the internal promoter. In an *in vitro* genotoxicity assay, an internal SFFV enhancer/promoter was shown to increase the expression of an intact lentiviral 5′ LTR, leading to higher transformation rates compared to a vector without an internal promoter.[39] Although the promoter sequences are deleted from the LTR in SIN vectors, a residual promoter activity exists within the leader region, resulting in full-length transcripts.[36] This study shows for the first time that this residual activity of the 5′ SIN LTR is influenced by the activity of the internal promoter, resulting in a higher amount of aberrantly spliced viral–cellular transcripts. Nevertheless, a portion of the transcripts detected in this experiment might also be the result of cellular read-through transcription, as previously suggested.[26]

The assessment of the genotoxic potential confirmed the importance of the LTR configuration and internal promoter and their impact on genotoxicity via the formation of viral–cellular transcripts. The quantitative analysis of transformed 32D colonies showed that the proportion of integrants generating fusion transcripts is clearly reduced in SIN vectors and even further in a SIN vector bearing a weak internal PGK promoter. *Ghr* fusion transcripts, which are the major cause for the functional transformation in this assay, were detected with both full and SIN LTR SFFV-containing vectors leading to its upregulation, whereas a SIN PGK vector did not result in the deregulation of gene expression.

The data demonstrate the high potential of high-resolution integration site analysis of *in vitro* cultured bulk cells. The study shows that functional vector elements contribute to the initial target site selection of lentiviral vectors as well as to the later over time clonal development. Above that, the incidence of chimeric transcripts and the genotoxic potential are strongly influenced by the vector design, in particular by LTR configuration and promoter. It is concluded that, also in the context of SIN vectors, powerful promoters should be avoided or at least be restricted to the lineages in which high transgene expression is required.

## ACKNOWLEDGMENTS

Funding was provided by the European Commission's 7th Framework Program through contracts FP7-HEALTH-FS-2009-222878-PERSIST, FP7-HEALTH-2012-INNOVATION-1-305011-Net4CGD, FP7-HEALTH-F5-2012-305421-EUROFANCOLEN, and the NC3R CRACK IT Challenge 21: InMutaGene. The authors thank the DKFZ Genomics and Proteomics Core Facility.

## AUTHOR DISCLOSURE

M.S. and C.v.K. are co-founders of GeneWerk GmbH. R.F. and M.S. are part-time employers and C.B. is a full-time employee of GeneWerk GmbH. S.J.S., D.C., E.G., and I.G.F. have no competing financial interests to disclose.



## REFERENCES

1. Morris EC, Fox T, Chakraverty R, et al. Gene therapy for Wiskott–Aldrich syndrome in a severely affected adult. Blood 2017 Jul 17 [Epub ahead of print];. DOI: 10.1182/blood-2017-04-777136.

2. Hacein-Bey Abina S, Gaspar HB, Blondeau J, et al. Outcomes following gene therapy in patients with severe Wiskott–Aldrich syndrome. JAMA 2015;313:1550–1563.

3. Ylä-Herttuala S. ADA-SCID gene therapy endorsed by European Medicines Agency for marketing authorization. Mol Ther 2016;24:1013–1014.

4. Sessa M, Lorioli L, Fumagalli F, et al. Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an *ad-hoc* analysis of a non-randomised, open-label, Phase 1/2 trial. Lancet 2016;388:476–487.

5. Biasco L, Ambrosi A, Pellin D, et al. Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. EMBO Mol Med 2011;3:89–101.

6. Pasi M, Mornico D, Volant S, et al. DNA minicircles clarify the specific role of DNA structure on retroviral integration. Nucleic Acids Res 2016;44:7830–7847.

7. Papanikolaou E, Paruzynski A, Kasampalidis I, et al. Cell cycle status of CD34(+) hemopoietic stem cells determines lentiviral integration in actively transcribed and development-related genes. Mol Ther 2015;23:683–696.

8. Demeulemeester J, De Rijck J, Gijsbers R, et al. Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. Bioessays 2015;37:1202–1214.

9. Lewinski MK, Yamashita M, Emerman M, et al. Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog 2006;2:e60.
10. Schröder ARW, Shinn P, Chen H, et al. HIV-1 Integration in the human genome favors active genes and local hotspots. Cell 2002;110:521–529.
11. Wu X, Li Y, Crise B, et al. Transcription start regions in the human genome are favored targets for MLV integration. Science 2003;300:1749–1751.
12. Cattoglio C, Pellin D, Rizzi E, et al. High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. Blood 2010;116:5507–5517.
13. Serrao E, Engelman AN. Sites of retroviral DNA integration: from basic research to clinical applications. Crit Rev Biochem Mol Biol 2016;51:26–42.
14. Hacein-Bey-Abina S, Garrigue A, Wang GP, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 2008;118:3132–3142.
15. Howe SJ, Mansour MR, Schwarzwaelder K, et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest 2008;118:3143–3150.
16. Stein S, Ott MG, Schultze-Strasser S, et al. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med 2010;16:198–204.
17. Braun CJ, Boztug K, Paruzynski A, et al. Gene therapy for Wiskott–Aldrich syndrome—long-term efficacy and genotoxicity. Sci Transl Med 2014;6:227ra33.
18. Modlich U, Bohne J, Schmidt M, et al. Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. Blood 2006;108:2545–2553.
19. Montini E, Cesana D, Schmidt M, et al. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest 2009;119:964–975.
20. Cesana D, Ranzani M, Volpin M, et al. Uncovering and dissecting the genotoxicity of self-inactivating lentiviral vectors *in vivo*. Mol Ther 2014;22:774–785.
21. Almarza D, Bussadori G, Navarro M, et al. Risk assessment in skin gene therapy: viral–cellular fusion transcripts generated by proviral transcriptional read-through in keratinocytes transduced with self-inactivating lentiviral vectors. Gene Ther 2011;18:674–681.
22. Schambach A, Galla M, Maetzig T, et al. Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors. Mol Ther 2007;15:1167–1173.
23. Cavazzana-Calvo M, Payen E, Negre O, et al. Transfusion independence and HMGA2 activation after gene therapy of human $\beta$-thalassaemia. Nature 2010;467:318–322.
24. Heckl D, Schwarzer A, Haemmerle R, et al. Lentiviral vector induced insertional haploinsufficiency of Ebf1 causes murine leukemia. Mol Ther 2012;20:1187–1195.
25. Rothe M, Modlich U, Schambach A. Biosafety challenges for use of lentiviral vectors in gene therapy. Curr Gene Ther 2013;13:453–468.
26. Cesana D, Sgualdino J, Rudilosso L, et al. Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. J Clin Invest 2012;122:1667–1676.
27. Moiani A, Paleari Y, Sartori D, et al. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. J Clin Invest 2012;122:1653–1666.
28. Cooray S, Howe SJ, Thrasher AJ. Retrovirus and lentivirus vector design and methods of cell conditioning. Methods Enzymol 2012;507:29–57.
29. Li CL, Xiong D, Stamatoyannopoulos G, et al. Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the cHS4 chromatin insulator. Mol Ther 2009;17:716–724.
30. Gabriel R, Kutschera I, Bartholomae CC, et al. Linear amplification mediated PCR—localization of genetic elements and characterization of unknown flanking DNA. J Vis Exp 2014;e51543.
31. Paruzynski A, Arens A, Gabriel R, et al. Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. Nat Protoc 2010;5:1379–1395.
32. Arens A, Appelt J-U, Bartholomae CC, et al. Bioinformatic clonality analysis of next-generation sequencing-derived viral vector integration sites. Hum Gene Ther Methods 2012;23:111–118.
33. Schmidt M, Schwarzwaelder K, Bartholomae C, et al. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods 2007;4:1051–1057.
34. Abel U, Deichmann A, Bartholomae C, et al. Real-time definition of non-randomness in the distribution of genomic events. PLoS One 2007;2:e570.
35. Morris EK, Caruso T, Buscot F, et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. Ecol Evol 2014;4:3514–3524.
36. Logan AC, Haas DL, Kafri T, et al. Integrated self-inactivating lentiviral vectors produce full-length genomic transcripts competent for encapsidation and integration. J Virol 2004;78:8421–8436.
37. Xu W, Russ JL, Eiden MV. Evaluation of residual promoter activity in $\gamma$-retroviral self-inactivating (SIN) vectors. Mol Ther 2012;20:84–90.
38. Bokhoven M, Stephen SL, Knight S, et al. Insertional gene activation by lentiviral and gammaretroviral vectors. J Virol 2009;83:283–294.
39. Knight S, Bokhoven M, Collins M, et al. Effect of the internal promoter on insertional gene activation by lentiviral vectors with an intact HIV long terminal repeat. J Virol 2010;84:4856–4859.
40. Felice B, Cattoglio C, Cittaro D, et al. Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS One 2009;4:e4571.
41. Yang Q, Lucas A, Son S, et al. Overlapping enhancer/promoter and transcriptional termination signals in the lentiviral long terminal repeat. Retrovirology 2007;4:4.
42. Emery DW. The use of chromatin insulators to improve the expression and safety of integrating gene transfer vectors. Hum Gene Ther 2011;22:761–774.
43. Ashe MP, Griffin P, James W, et al. Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. Genes Dev 1995;9:3008–3025.
44. Zaiss A-K, Son S, Chang L-J. RNA 3′ readthrough of oncoretrovirus and lentivirus: implications for vector safety and efficacy. J Virol 2002;76:7209–7219.
45. Knight S, Zhang F, Mueller-Kuller U, et al. Safer, silencing-resistant lentiviral vectors: optimization of the ubiquitous chromatin-opening element through elimination of aberrant splicing. J Virol 2012;86:9088–9095.
46. Furger A, Monks J, Proudfoot NJ. The retroviruses human immunodeficiency virus type 1 and Moloney murine leukemia virus adopt radically different strategies to regulate promoter-proximal polyadenylation. J Virol 2001;75:11735–11746.


Received for publication August 17, 2017;
accepted after revision August 17, 2017.

Published online: August 18, 2017.