genes

Article

Genome-Wide Analysis of Alpharetroviral Integration in Human Hematopoietic Stem/Progenitor Cells

Arianna Moiani [1], Julia Debora Suerth [2], Francesco Gandolfi [1], Ermanno Rizzi [3], Marco Severgnini [3], Gianluca De Bellis [3], Axel Schambach [2] and Fulvio Mavilio [1,*]

[1] Genethon, 1bis Rue de l'Internationale, 91020 Evry, France;
E-Mails: amoiani@genethon.fr (A.M.); fgandolfi@genethon.fr (F.G.)

[2] Institute of Experimental Hematology, Hannover Medical School, Carl-Neuberg-Str.1,
D-30625 Hannover, Germany; E-Mails: suerth.julia@mh-hannover.de (J.D.S.);
schambach.axel@mh-hannover.de (A.S.)

[3] Institute for Biomedical Technologies, Consiglio Nazionale delle Ricerche, Milan 20132, Italy;
E-Mails: ermanno.rizzi@itb.cnr.it (E.R.); marco.severgnini@itb.cnr.it (M.S.);
gianluca.debellis@itb.cnr.it (G.B.)

* Author to whom correspondence should be addressed; E-Mail: fmavilio@genethon.fr;
Tel.: +33-1-6947-2978.

Abstract: Gene transfer vectors derived from gamma-retroviruses or lentiviruses are currently used for the gene therapy of genetic or acquired diseases. Retroviral vectors display a non-random integration pattern in the human genome, targeting either regulatory regions (gamma-retroviruses) or the transcribed portion of expressed genes (lentiviruses), and have the potential to deregulate gene expression at the transcriptional or post-transcriptional level. A recently developed alternative vector system derives from the avian sarcoma-leukosis alpha-retrovirus (ASLV) and shows favorable safety features compared to both gamma-retroviral and lentiviral vectors in preclinical models. We performed a high-throughput analysis of the integration pattern of self-inactivating (SIN) alpha-retroviral vectors in human CD34$^+$ hematopoietic stem/progenitor cells (HSPCs) and compared it to previously reported gamma-retroviral and lentiviral vectors integration profiles obtained in the same experimental setting. Compared to gamma-retroviral and lentiviral vectors, the SIN-ASLV vector maintains a preference for open chromatin regions, but shows no bias for transcriptional regulatory elements or transcription units, as defined by genomic annotations and epigenetic markers (H3K4me1 and H3K4me3 histone

modifications). Importantly, SIN-ASLV integrations do not cluster in hot spots and target potentially dangerous genomic loci, such as the EVI2A/B, RUNX1 and LMO2 proto-oncogenes at a virtually random frequency. These characteristics predict a safer profile for ASLV-derived vectors for clinical applications.

Keywords: retroviral integration; alpha-retroviral vectors; gene therapy; human hematopoietic cells

1. Introduction

Transplantation of hematopoietic stem cells genetically modified by retroviral vectors has proven its clinical efficacy in a number of seminal clinical trials for the treatment of severe monogenic disorders [1–8]. However, some of these studies also showed the genotoxic risks associated with the insertion of foreign DNA in the human genome, which limit the clinical application of integrating vectors (reviewed in [9]). Several efforts have been made to improve the safety of retroviral vectors, leading to the design of safer constructs and the development of robust *in vitro* and *in vivo* genotoxic assays to predict the potential risk associated with their integration into the genome [10–12]. High-definition mapping of integration sites of vectors derived from the Moloney murine leukemia virus (MLV) and human immunodeficiency virus (HIV) in murine and human cells revealed non-random profiles with a strong tendency to target active regulatory regions for MLV-derived gamma-retroviral vectors [13,14] and transcribed regions for HIV-derived lentiviral vectors [15,16]. These integration patterns explain the relatively high risk to deregulate gene expression at the transcriptional or post-transcriptional level observed in pre-clinical, as well as in clinical studies (reviewed in [9]).

Small-scale surveys of integration sites of vectors derived from alpha-retroviruses, such as the avian sarcoma-leukosis virus (ASLV), in different cell types indicated a more random pattern compared to other retroviruses, with a slight preference for transcription units, but no apparent preference for promoters and transcription start sites (TSSs) [17–20]. This potentially more favorable integration profile prompted the development of a replication-deficient, self-inactivating (SIN) ASLV-derived vector capable of efficiently transducing murine and human cells [21]. This vector was able to sustain long-term transgene expression in murine and human hematopoietic progenitors at levels comparable to those obtained with SIN-MLV and SIN-HIV vectors and to correct the X-linked chronic granulomatous disease (X-CGD) phenotype in a mouse model of the disease [20,22].

We and others previously reported that MLV, SIN-MLV and SIN-HIV integrations are highly clustered in the human genome, with cell-specific patterns that correlate with the transcriptional program and the epigenetic landscape of each cell type [14–16,19,23–26]. In this study, we report a high-definition analysis of the integration patterns of SIN-MLV, SIN-ASLV and SIN-HIV vectors in human CD34[+] hematopoietic stem/progenitor cells (HSPCs), which was carried out to evaluate their comparative genotoxic potential in a clinically relevant target cell. We show that the SIN-ASLV integration profile is close to random, with no preferential targeting of TSSs or transcribed genes compared to SIN-MLV and SIN-HIV. The SIN-ASLV vector does not target CpG islands, conserved

non-coding regions (CNCs) or elements enriched in transcription factor binding sites (TFBS), is less frequently associated with epigenetically defined promoter and enhancer regions compared to SIN-MLV and is randomly associated with repetitive elements in the genome. Similarly, we observed no preference for transcribed regions compared to SIN-HIV. Heterochromatic regions are excluded by the integration pattern of all three vectors. Interestingly, the ASLV vector showed no apparent clustering in the genome and has no association with the typical integration hot spots observed for MLV- and HIV-based vectors. These results highlight a safer integration profile of alpha-retroviral vectors in human cells, supporting their development as a clinical gene transfer tool.

2. Experimental

2.1. Vectors and Cells

Human CD34[+] HSPCs were purified form umbilical cord blood, pre-stimulated for 48 h in serum-free Iscove's modified Dulbecco medium supplemented with 20% Fetal Calf Serum (FCS), 20 ng/mL human thrombopoietin, 100 ng/mL Flt-3 ligand, 20 ng/mL interleukin-6 and 100 ng/mL stem cell factor, as previously described [23]. HSPCs were transduced with the SIN-ASLV vector, pAlpha.SIN.EFS.EGFP.WPRE (noTATA), expressing GFP under the control of the elongation factor 1α promoter, pseudotyped in an amphotropic envelope by three-plasmid transfection in 293T cells, as previously described [20]. Cells were infected by 3 rounds of spinoculation (1500 rpm for 45 min) in the presence of 4 μg/mL polybrene. Transduction efficiency was evaluated by cytofluorimetric analysis of GFP expression 48 h after infection.

2.2. Amplification, Sequencing, and Analysis of Retroviral Integration Sites

Genomic DNA was extracted from a pool of 3.5×10^6 CD34[+]/GFP[+] cells enriched by fluorescence-activated cell sorting, after a brief period in culture to dilute unintegrated vectors. 3'-LTR vector-genome junctions were amplified by LM-PCR adapted to the GS-FLX Genome Sequencer (Roche/454 Life Sciences) pyrosequencing platform, as previously described [14]. Raw sequence reads were processed by an automated bioinformatic pipeline that eliminated small and redundant sequences [14] and mapped on the University of California at Santa Cruz (UCSC) hg19 release of the human genome [14]. All UCSC RefSeq genes having their TSS at ±50 kb from an integration site were annotated as targets. Genomic features were annotated when their genomic coordinates overlapped for ≥1 nucleotide with a ±1 kb interval around each integration site. We used UCSC tracks for both CpG islands and conserved TFBSs, and the previously described genomic coordinates of 82,335 mammalian conserved non-coding sequences (CNCs) [27]. Raw sequences having a single or ambiguous match in the genome (the latter mapping in multiple genomic positions with a difference in the identity <2) were blasted on the UCSC RepeatMasker database. DNase I hypersensitive sites from publicly available data [28] were annotated when overlapping for at least 1 bp with a ±1-kb interval around an integration. Repetitive elements were annotated when directly targeted by each integration site. Sequences having multiple matches were collapsed and counted as one when matching in the same genomic positions and were univocally associated with the single type of repetitive element they targeted.

For the association of the integrations with epigenetically defined chromatin states, we used publicly available ChIP-Seq data (NIH Roadmap Epigenomics Mapping Consortium database) that we re-annotated in the UCSC hg19 release of the human genome. We analyzed the distribution of integration sites around histone modifications (H3K4me1, H3K4me3, H3K36me3, H3K27me3) using the seqMINER platform [29]. Previously generated SIN-MLV, SIN-HIV integrations and random control sequences datasets [14] were also re-annotated on the UCSC hg19 genome. For all pairwise comparisons, we applied a 2-sided Fisher's exact test. The threshold for statistical significance was set at a *p*-value < 0.01.

3. Results and Discussion

3.1. SIN-ASLV Vectors Exhibit an Almost Random Integration Profile in the Genome of Human CD34[+] HSPCs

To generate a high-definition alpha-retroviral integration profile in human HSPCs, we transduced umbilical cord blood-derived CD34[+] cells with a previously described SIN-ASLV vector carrying a GFP expression cassette under the control of the intron-less, 240-bp version of the elongation factor-1α (EFS) promoter [20]. Cells were transduced at 10% to 20% efficiency and were selected for GFP expression by cell sorting 10 days after infection, to dilute unintegrated vectors. Vector-genome junctions were amplified from genomic DNA by ligation-mediated (LM)-PCR and pyrosequenced, as previously described [14]. Raw sequences (available at GenBank with the accession number SRR1282019) were processed by a previously described bioinformatic pipeline [14] and mapped on the UCSC hg19 release of the human genome, to obtain 8250 unique insertion sites. Two datasets of SIN-MLV (13,097) and SIN-HIV (31,827) vector integrations, previously generated in human umbilical cord blood-derived CD34[+] cells in comparable experimental conditions, and a set of in-silico generated normalized random sites (40,000) [14,26] were re-annotated on the hg19 genome and used for comparison. To identify differences in the integration preferences of SIN-ASLV compared to SIN-MLV and SIN-HIV in human HSPCs, we first analyzed the distribution of integration sites around RefSeq genes in the human genome: integration was annotated as TSS-proximal when occurring in an interval of ±2.5 kb from the TSS of any RefSeq gene, intragenic when occurring inside a RefSeq gene >2.5 kb from the TSS and intergenic in all other cases.

The high-definition profile of SIN-ASLV integration showed only a modest preference for TSSs (6.97% of the integration sites were annotated as TSS-proximal) compared to SIN-HIV and random sites (3.45% and 3.16%, respectively), which was significantly lower than that observed for the SIN-MLV vector (23.38%, *p* < 0.01). Similarly, SIN-ASLV showed only a slight tendency to integrate into genes (49.48% *vs.* 40.58% of random sites), significantly lower than that observed for SIN-HIV vectors (76.77%, *p* < 0.01). As a consequence, the frequency of SIN-ASLV integration outside transcription units was only slightly lower than random (43.55% *vs.* 56.26%) and significantly higher than those observed for the other two vectors (34.36% and 19.78%, respectively, *p* < 0.01) (Table 1). A plot of the relative distance of SIN-ASLV integration sites in an interval of ±50 kb from any TSS revealed a spread distribution with only a modest accumulation in the ±2.5 kb interval around TSS compared to the SIN-MLV vector. A higher definition map (100-bp intervals) showed the absence of

Figure 1. Genomic distribution of SIN-MLV, SIN-ASLV and SIN-HIV integrations in human HSPCs. The distribution of the distance of SIN-MLV (red bars), SIN-ASLV (yellow bars) and SIN-HIV (blue bars) integration sites from the TSS of targeted genes at 2500-bp (**a**) or 50-bp (**b**) resolution. The percentage of genes targeted at each position is plotted on the y-axis. The black line indicates the distribution of random control sites.
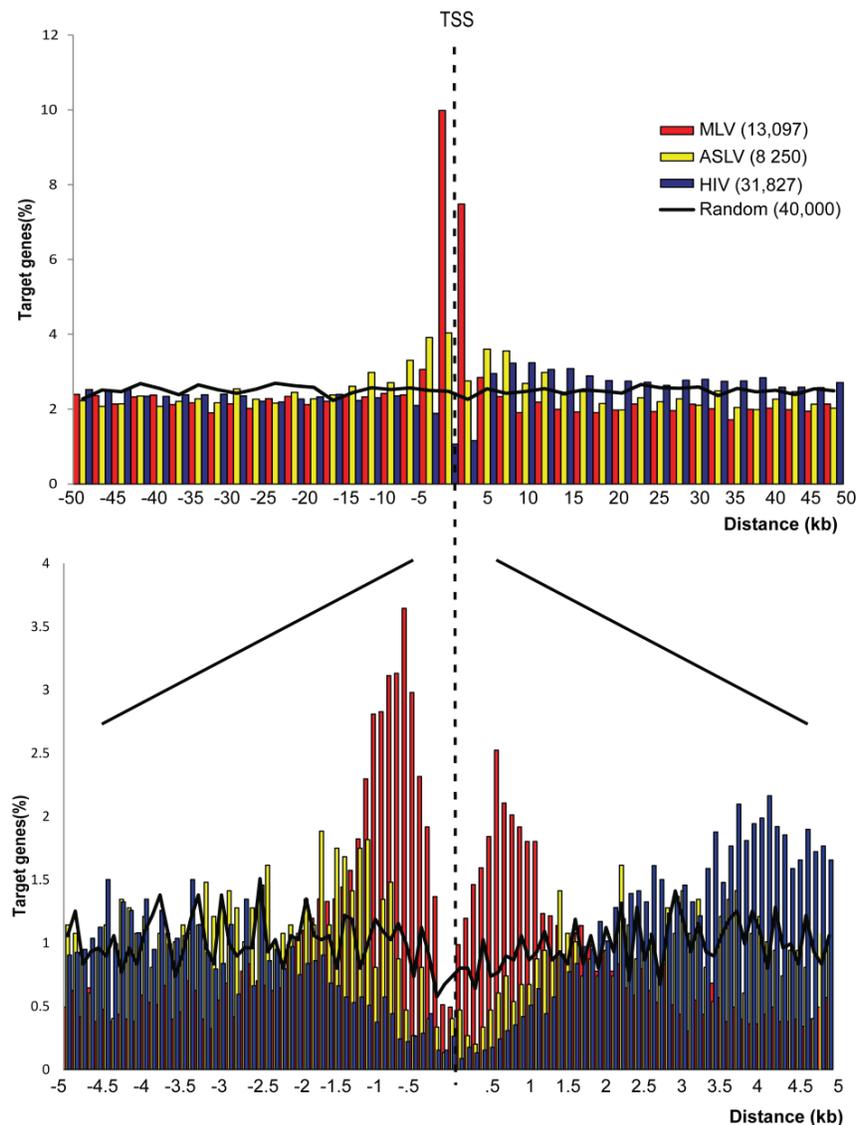
Table 2. Integrations targeting repetitive elements in the genome.

Vector	Repetitive elements (%)	LINEs (%)	SINEs (%)	Satellites (%)	LTRs (%)	Others (%)
SIN-ASLV (8,899)	51.29	19.50	19.77	0.43	6.26	5.35
SIN-MLV (13,606)	37.75	10.96	17.26	0.01	5.09	4.42
SIN-HIV (32,964)	45.78	19.45	16.86	0.05	4.08	5.35
Random (40,000)	50.96	21.27	14.69	0.35	9.57	5.10

Percentage of SIN-MLV, SIN-ASLV, SIN-HIV integrations and random sequences targeting repetitive elements and the percentage targeting each specific element: LINEs, short interspersed nuclear elements (SINEs), satellites, LTRs and all the other elements.

Overall, these data indicate a remarkably random pattern of integration for the ASLV-derived vector, which shows none of the characteristic preferences of gamma-retroviruses and lentiviruses for genes and genetic elements associated with gene function and regulation.

3.2. SIN-ASLV Integration Is Not Associated with Epigenetically-Defined Functional Genomic Regions

Many studies have reported a strong correlation between MLV and HIV integration sites and distinct epigenetic markers in different cell types (reviewed in [9]). In human CD34[+] HSPCs, MLV integrations are strongly associated with histone modifications marking transcriptionally active PolII promoters and enhancers, while HIV integrations correlate with epigenetic markers of active PolII elongation within transcription units [14,32]. We therefore investigated the association of SIN-ASLV integrations with defined epigenetic markers of functional genomic elements. Taking advantage of publicly available ChIP-Seq data in the genome of human CD34[+] HSPCs, we analyzed the association of SIN-MLV, SIN-ASLV and SIN-HIV integrations with specific histone modifications defining active or poised PolII promoters, enhancers, transcribed regions and heterochromatin (H3K4me1, H3K4me3, H3K36me3, H3K27me3 and H3K27ac).

More than 60% and 70% of SIN-MLV and SIN-HIV integrations sites, respectively, were univocally associated with a defined chromatin state, compared to only 40% of the SIN-ASLV integration sites, a frequency very close to the 30% observed for random sequences. In particular, SIN-ASLV integrations were found around regulatory regions, *i.e.*, enhancer (H3K4me1[+]) and promoters (H3K4me3[+]), at a much lower frequency compared to SIN-MLV (10% *vs.* 37% in enhancers and 6% *vs.* 26% in promoters, respectively, $p < 0.01$), a tendency comparable to that observed for SIN-HIV (10% in enhancers and 4% in promoters) and slightly higher than that of the random sample (3% and 2%, respectively). Moreover, SIN-ASLV integrations were poorly associated with a marker of transcribed gene bodies (H3K36me3) compared to SIN-HIV (15% *vs.* 38%, $p < 0.01$). All three vectors were under-represented in heterochromatic regions marked by H3K27me3 compared to random sites, with the SIN-ASLV vector showing the highest association (Figure 2A). This analysis is in agreement with the associations observed at the level of DNA sequence and genomic annotations, and confirms the preference of SIN-MLV and SIN-HIV vectors for, respectively, regulatory sequences and transcribed regions and an almost random integration pattern for SIN-ASLV. The modest bias observed for SIN-ASLV integrations in DNase I hypersensitive regions compared to the random sample (Table S1) can be explained by a certain tendency to integrate in "open" chromatin regions, as observed for most retroviruses [14,16,33,34].

The differences between SIN-ASLV and the other two vectors in targeting defined chromatin regions are highlighted by plotting the average integration densities of each vector type around each histone modification. Indeed, we clearly observe a peak of SIN-MLV integration sites in a ±2.5-kb interval from epigenetically-defined enhancers and promoters, while the distribution of the SIN-ASLV and SIN-HIV integrations around these elements is similar to that observed for random sequences (Figure 2B). The quasi-random association of ASLV integrations in regulatory element predicts a much lower genotoxic risk compared to MLV-derived vectors, whose tendency to target active regulatory elements is at the basis of their propensity to cause insertional deregulation of gene expression [9]. Most (>70%) of the genes targeted by all three vectors in HSPCs are actively expressed

(Figure S1), an expected finding, considering that retroviral target site selection is highly favored by an open chromatin state [14,16,33,34]. However, the SIN-ASLV vector targets the transcribed portion of active genes at a much lower frequency compared to HIV and is devoid of splicing signals, thus predicting a much lower risk to interfere with gene regulation at the post-transcriptional level [22,35,36].

Figure 2. Association of vector integration sites with different epigenetically-defined chromatin states. (**a**) The percentage of integration sites associated with specific, epigenetically defined genomic regions for each vector type. Chromatin states are categorized on the basis of the combination of different epigenetic marks mapped by ChIP-seq in human HSPCs. Only integration sites that are unambiguously associated with one chromatin state were used for the analysis. (**b**) The mean densities of H3K4me1, H3K4me3, H3K36me3 and H3K27me3 ChIP-seq fragments in a 5-kb window around all SIN-MLV (red), SIN-ASLV (yellow) and SIN-HIV (light blue) integration sites and random sequences (black). ac: H3K27ac.
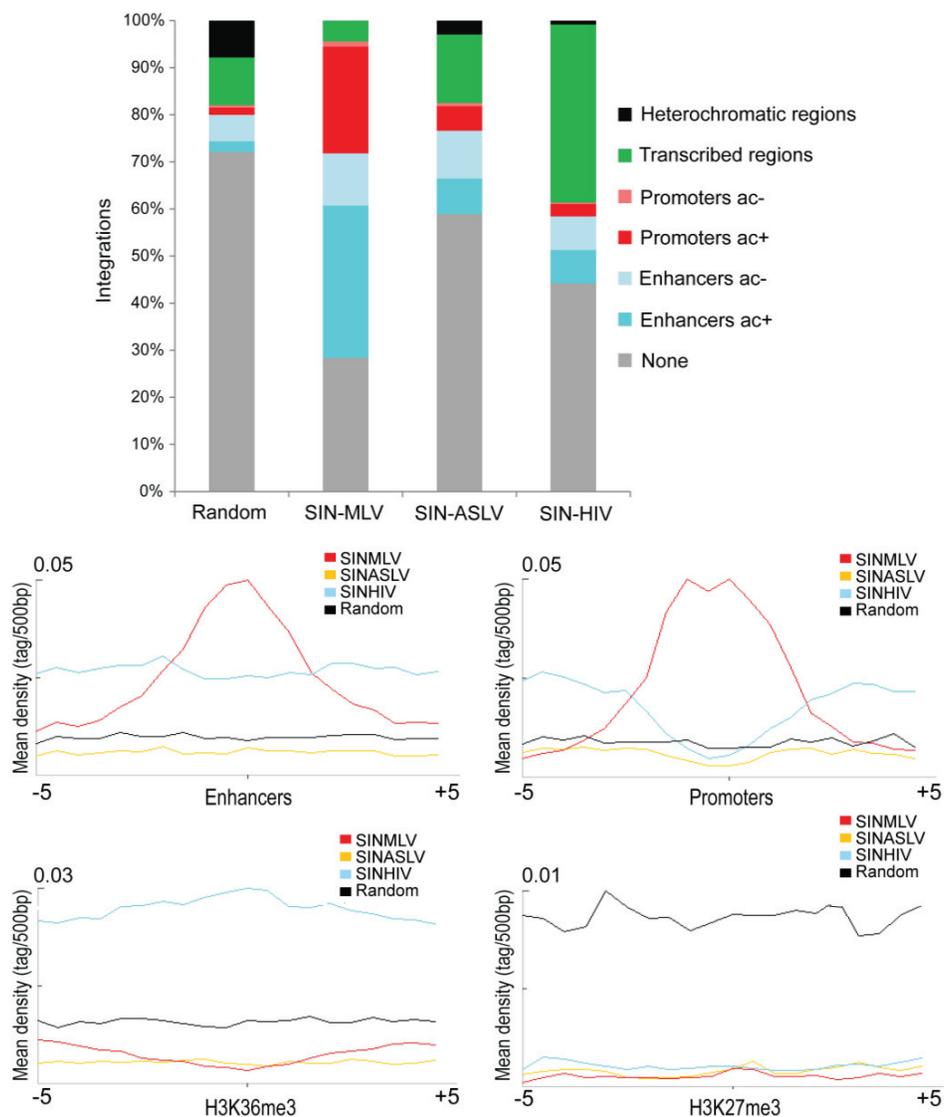
3.3. SIN-ASLV Vector Integrations Are Not Clustered in the Human Genome

The integration profile of MLV- and HIV-derived vectors in the human genome is characterized by heavy clustering into integration hot spots, where MLV forms narrow clusters overlapping active regulatory elements and HIV larger clusters targeting a subset of transcribed genes, in both cases in a cell-specific fashion [14–16,24,37]. On the contrary, the SIN-ASLV vector showed no significant clustering when we applied a statistical definition of clusters adjusted to the numerosity of the sample [14], which for the SIN-ASLV dataset was three integrations in 53,920 bp. By this threshold, we were able to identify only 484 clusters, a significantly lower frequency compared to the 1,415 and 2,724 identified for the SIN-MLV and SIN-HIV vectors, respectively ($p < 0.01$). Only 21% of all SIN-ASLV integrations are clustered, compared to 56% and 51% of SIN-MLV and SIN-HIV integrations, respectively ($p < 0.01$) (Table 3). Moreover, SIN-ASLV clusters are mostly made of few (three or four) integrations with few clusters containing up to nine integrations, while SIN-MLV and SIN-HIV clusters contain up to 37 and 122 integrations, respectively. From these data, it appears that, contrary to other retroviral vectors, SIN-ASLV integrations do not form hot spots of integrations in the human genome. Interestingly, when we looked at integration clusters at single genomic loci, we observed that the frequency of SIN-ASLV integrations at the typical MLV or HIV hot spots is very low and comparable to the frequency observed for random sequences. Figure 3 shows a comparison of the integration pattern of the three vectors in the NF1-EVI2A/B, RUNX1, LMO2 and PACS1 loci, four known hot spots for MLV or HIV integration. The same scenario is true for the EVI1/MDS1 (MECOM) locus and all other MLV or HIV integration hot spots (data not shown). The low frequency of SIN-ASLV integration in proto-oncogenes responsible for the severe adverse events observed in clinical gene therapy trials, such as LMO2, provide a further indication of its lower genotoxic profile.

**Table 3.** Clusters of integration sites in the genome of human HSPCs.

	SIN-MLV (13,097)	SIN-ASLV (8250)	SIN-HIV (31,827)
Clusters	1415	484	2724
Integrations in clusters (%)	56	21	51
Average cluster dimension	5.1	3.6	5.9

The number of SIN-MLV, SIN-ASLV and SIN-HIV clusters of integrations, the percentage of integrations in clusters and the average cluster dimension, calculated based on random sequences distribution in the genome. The threshold for cluster definition was defined at a *p*-value of <0.01 by a statistical algorithm that adjusts for the numerosity of the sample [14].

Although the SIN-ASLV integration profile shows none of the features typical of MLV- and HIV-derived vectors, it is not completely random and shows a general preference for euchromatic regions. It is now know that both MLV and HIV pre-integration complexes (PICs) are targeted to chromatin by a tethering mechanism involving the interaction of the viral integrase with host cell factors: the LEDGF/p75 chromatin component interacts with the HIV integrase and directs its integration into transcribed gene bodies [38,39], while the MLV integrase appears to bind to bromodomain-containing BET proteins specifically associated with acetylated histones around TSSs and active regulatory elements [40–42]. Although it is likely that ASLV also may adopt a tethering mechanism to direct its integration in favorable genomic regions, the details are unknown. The integration preferences

uncovered by our analysis predict an interaction with a broader range of host cell factors, which tether the PICs to open chromatin regions with unspecified or very subtle functional characteristics, thus leading to a more random profile characterized by the absence of hot spots.

Figure 3. SIN-MLV, SIN-ASLV and SIN-HIV integration sites and clusters in CD34$^+$ HSPC-specific loci. Distribution of SIN-MLV (red), SIN-ASLV (green) and SIN-HIV (blue) integration clusters (horizontal solid bars) and integrations (vertical marks) in the NF1-EVI2A/B, RUNX1, LMO2 and PACS1 loci, as displayed by the UCSC Genome Browser.
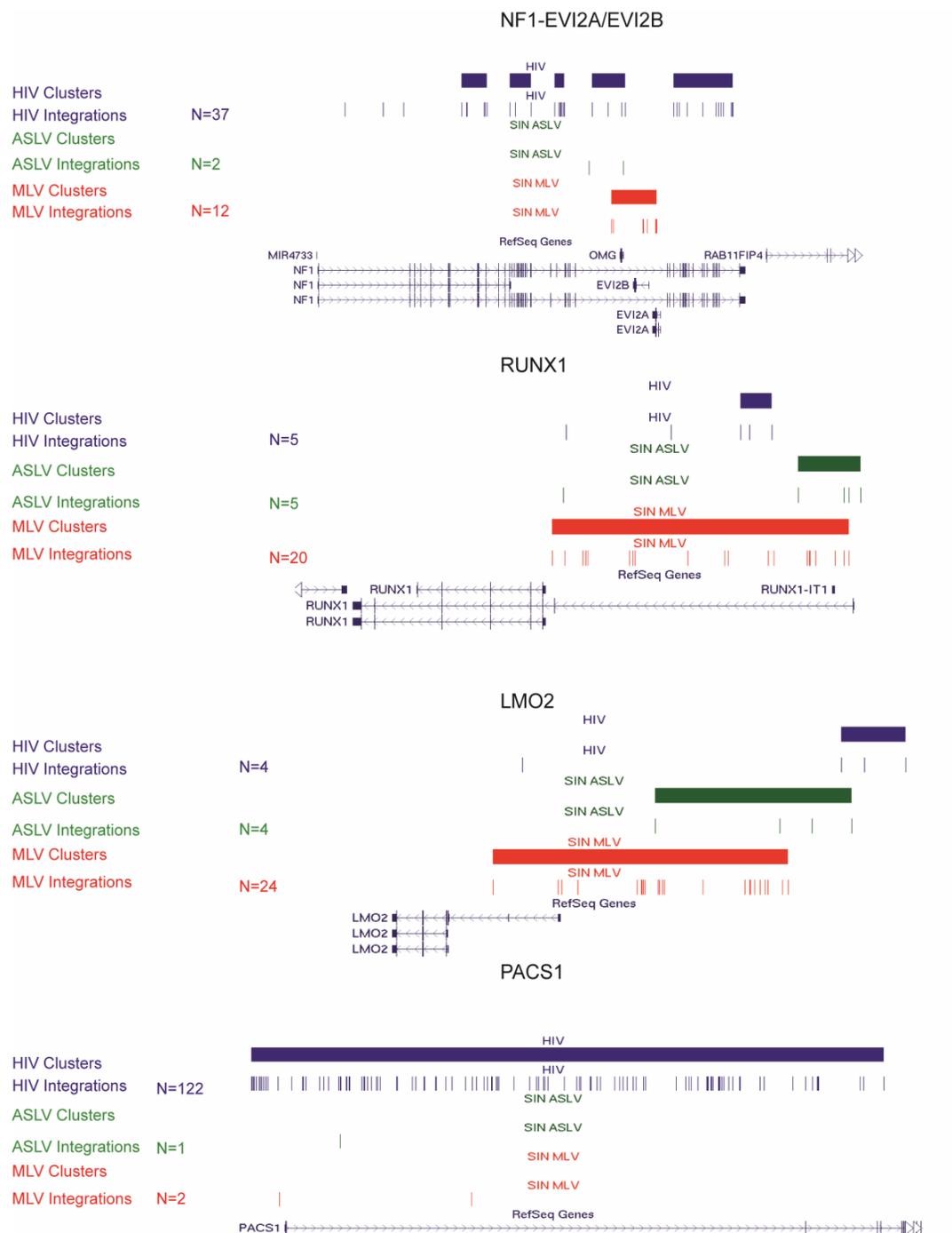
4. Conclusions

Overcoming the genotoxic consequences of retroviral vector integration in the host cell genome is one of the major issues for the application of retroviral-based gene transfer in clinical trials. The strong preference to target TSSs, active regulatory elements or transcribed genes, together with the high frequency of clustering around hot spots, is a characteristics shared by all retroviral vectors currently used in clinical gene therapy. These characteristics are at the basis of the potential of retroviral insertion to deregulate gene expression at the transcriptional or post-transcriptional level, which has been observed to cause clonal expansion and contribute to neoplastic transformation in a number of cases (reviewed by [9]). Many efforts are being made to improve the safety of currently available retroviral vectors by removing the viral transcriptional control element and avoiding dominant, long-range acting enhancers in the transgene expression cassette. Retargeting vector integration has proven more difficult and was so far unsuccessful. No strategy is obviously perfect, and even a completely random integration machinery would not abolish the risk of inducing an insertional oncogenic mutation in the host cell genome.

Based on a genome-wide analysis of >8000 integration sites in human HSPCs, we show that a SIN-ASLV vector has a quasi-random integration pattern that privileges active chromatin regions, but is not associated with active regulatory elements, like MLV, or with transcribed genes, like HIV. More importantly, the SIN-ASLV vector showed no integration hot spots and no preferences for subsets of genes with a defined ontology or genes that were previously identified as being activated by retroviral insertion into tumors. Previous evaluations of ASLV-derived vectors in pre-clinical models proved its ability to sustain long-term transgene expression in murine and human hematopoietic progenitors and to correct the pathology in a mouse model of X-linked Chronic Granulomatous Disease (X-CGD), with no evidence of post-transcriptional interference [20,22]. Combined with the use of short-range or cell-specific transcriptional regulatory elements, an ASLV vector appears to offer a very safe profile and to be an ideal candidate for *ex vivo* gene therapy applications.

Author Contributions

Conceived and designed the experiments: Arianna Moiani, Fulvio Mavilio. Performed the experiments: Arianna Moiani, Ermanno Rizzi, and Julia Debora Suerth. Analyzed the data: Arianna Moiani, Francesco Gandolfi, Marco Severgnini, and Gianluca De Bellis. Contributed reagents/materials/analysis tools: Julia Debora Suerth and Axel Schambach. Wrote the paper: Arianna Moiani and Fulvio Mavilio.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Hacein-Bey-Abina, S.; Le Deist, F.; Carlier, F.; Bouneaud, C.; Hue, C.; de Villartay, J.P.; Thrasher, A.J.; Wulffraat, N.; Sorensen, R.; Dupuis-Girod, S.; *et al.* Sustained correction of X-linked severe combined immunodeficiency by *ex vivo* gene therapy. *N. Engl. J. Med.* **2002**, *346*, 1185–1193.

2. Ott, M.G.; Schmidt, M.; Schwarzwaelder, K.; Stein, S.; Siler, U.; Koehl, U.; Glimm, H.; Kuhlcke, K.; Schilz, A.; Kunkel, H.; *et al.* Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRMD16 or SETBP1. *Nat. Med.* **2006**, *12*, 401–409.

3. Mavilio, F.; Pellegrini, G.; Ferrari, S.; di Nunzio, F.; di Iorio, E.; Recchia, A.; Maruggi, G.; Ferrari, G.; Provasi, E.; Bonini, C.; *et al.* Correction of junctional epidermolysis bullosa by transplantation of genetically modified epidermal stem cells. *Nat. Med.* **2006**, *12*, 1397–1402.

4. Aiuti, A.; Cattaneo, F.; Galimberti, S.; Benninghoff, U.; Cassani, B.; Callegaro, L.; Scaramuzza, S.; Andolfi, G.; Mirolo, M.; Brigida, I.; *et al.* Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N. Engl. J. Med.* **2009**, *360*, 447–458.

5. Cartier, N.; Hacein-Bey-Abina, S.; Bartholomae, C.C.; Veres, G.; Schmidt, M.; Kutschera, I.; Vidaud, M.; Abel, U.; Dal-Cortivo, L.; Caccavelli, L.; *et al.* Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **2009**, *326*, 818–823.

6. Boztug, K.; Schmidt, M.; Schwarzer, A.; Banerjee, P.P.; Diez, I.A.; Dewey, R.A.; Bohm, M.; Nowrouzi, A.; Ball, C.R.; Glimm, H.; *et al.* Stem-cell gene therapy for the Wiskott-Aldrich syndrome. *N. Engl. J. Med.* **2010**, *363*, 1918–1927.

7. Biffi, A.; Montini, E.; Lorioli, L.; Cesani, M.; Fumagalli, F.; Plati, T.; Baldoli, C.; Martino, S.; Calabria, A.; Canale, S.; *et al.* Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science* **2013**, doi:10.1126/science.1233158.

8. Aiuti, A.; Biasco, L.; Scaramuzza, S.; Ferrua, F.; Cicalese, M.P.; Baricordi, C.; Dionisio, F.; Calabria, A.; Giannelli, S.; Castiello, M.C.; *et al.* Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science* **2013**, doi:10.1126/science.1233151.

9. Cavazza, A.; Moiani, A.; Mavilio, F. Mechanisms of retroviral integration and mutagenesis. *Human Gene Ther.* **2013**, *24*, 119–131.

10. Montini, E.; Cesana, D.; Schmidt, M.; Sanvito, F.; Ponzoni, M.; Bartholomae, C.; Sergi Sergi, L.; Benedicenti, F.; Ambrosi, A.; di Serio, C.; *et al.* Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat. Biotechnol.* **2006**, *24*, 687–696.

11. Modlich, U.; Navarro, S.; Zychlinski, D.; Maetzig, T.; Knoess, S.; Brugman, M.H.; Schambach, A.; Charrier, S.; Galy, A.; Thrasher, A.J.; *et al.* Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Mol. Ther.* **2009**, *17*, 1919–1928.

12. Montini, E.; Cesana, D.; Schmidt, M.; Sanvito, F.; Bartholomae, C.C.; Ranzani, M.; Benedicenti, F.; Sergi, L.S.; Ambrosi, A.; Ponzoni, 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.

13. Wu, X.; Li, Y.; Crise, B.; Burgess, S.M. Transcription start regions in the human genome are favored targets for mlv integration. *Science* **2003**, *300*, 1749–1751.

14. Cattoglio, C.; Pellin, D.; Rizzi, E.; Maruggi, G.; Corti, G.; Miselli, F.; Sartori, D.; Guffanti, A.; di Serio, C.; Ambrosi, A.; *et al.* High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. *Blood* **2010**, *116*, 5507–5517.

15. Schroder, A.R.; Shinn, P.; Chen, H.; Berry, C.; Ecker, J.R.; Bushman, F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **2002**, *110*, 521–529.

16. Wang, G.P.; Ciuffi, A.; Leipzig, J.; Berry, C.C.; Bushman, F.D. HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **2007**, *17*, 1186–1194.

17. Barr, S.D.; Leipzig, J.; Shinn, P.; Ecker, J.R.; Bushman, F.D. Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* **2005**, *79*, 12035–12044.

18. Hu, J.; Renaud, G.; Gomes, T.J.; Ferris, A.; Hendrie, P.C.; Donahue, R.E.; Hughes, S.H.; Wolfsberg, T.G.; Russell, D.W.; Dunbar, C.E. Reduced genotoxicity of avian sarcoma leukosis virus vectors in rhesus long-term repopulating cells compared to standard murine retrovirus vectors. *Mol. Ther.* **2008**, *16*, 1617–1623.

19. Mitchell, R.S.; Beitzel, B.F.; Schroder, A.R.; Shinn, P.; Chen, H.; Berry, C.C.; Ecker, J.R.; Bushman, F.D. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2004**, *2*, e234.

20. Suerth, J.D.; Maetzig, T.; Brugman, M.H.; Heinz, N.; Appelt, J.U.; Kaufmann, K.B.; Schmidt, M.; Grez, M.; Modlich, U.; Baum, C.; *et al.* Alpharetroviral self-inactivating vectors: Long-term transgene expression in murine hematopoietic cells and low genotoxicity. *Mol. Ther.* **2012**, *20*, 1022–1032.

21. Suerth, J.D.; Maetzig, T.; Galla, M.; Baum, C.; Schambach, A. Self-inactivating alpharetroviral vectors with a split-packaging design. *J. Virol.* **2010**, *84*, 6626–6635.

22. Kaufmann, K.B.; Brendel, C.; Suerth, J.D.; Mueller-Kuller, U.; Chen-Wichmann, L.; Schwable, J.; Pahujani, S.; Kunkel, H.; Schambach, A.; Baum, C.; *et al.* Alpharetroviral vector-mediated gene therapy for X-CGD: Functional correction and lack of aberrant splicing. *Mol. Ther.* **2013**, *21*, 648–661.

23. Cattoglio, C.; Facchini, G.; Sartori, D.; Antonelli, A.; Miccio, A.; Cassani, B.; Schmidt, M.; von Kalle, C.; Howe, S.; Thrasher, A.J.; *et al.* Hot spots of retroviral integration in human CD34[+] hematopoietic cells. *Blood* **2007**, *110*, 1770–1778.

24. Cattoglio, C.; Maruggi, G.; Bartholomae, C.; Malani, N.; Pellin, D.; Cocchiarella, F.; Magnani, Z.; Ciceri, F.; Ambrosi, A.; von Kalle, C.; *et al.* High-definition mapping of retroviral integration sites defines the fate of allogeneic T cells after donor lymphocyte infusion. *PLoS One* **2010**, *5*, e15688.

25. Cavazza, A.; Cocchiarella, F.; Bartholomae, C.; Schmidt, M.; Pincelli, C.; Larcher, F.; Mavilio, F. Self-inactivating MLV vectors have a reduced genotoxic profile in human epidermal keratinocytes. *Gene Ther.* **2013**, doi:10.1038/gt.2013.18.

26. Moiani, A.; Miccio, A.; Rizzi, E.; Severgnini, M.; Pellin, D.; Suerth, J.D.; Baum, C.; de Bellis, G.; Mavilio, F. Deletion of the LTR enhancer/promoter has no impact on the integration profile of MLV vectors in human hematopoietic progenitors. *PLoS One* **2013**, *8*, e55721.

27. Kim, S.Y.; Pritchard, J.K. Adaptive evolution of conserved noncoding elements in mammals. *PLoS Genet.* **2007**, *3*, 1572–1586.

28. John, S.; Sabo, P.J.; Canfield, T.K.; Lee, K.; Vong, S.; Weaver, M.; Wang, H.; Vierstra, J.; Reynolds, A.P.; Thurman, R.E.; *et al.* Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol. Biol.* **2013**, doi:10.1002/0471142727.mb2127s103.

29. Ye, T.; Krebs, A.R.; Choukrallah, M.A.; Keime, C.; Plewniak, F.; Davidson, I.; Tora, L. Seqminer: An integrated chip-seq data interpretation platform. *Nucleic Acids Res.* **2010**, *39*, e35.

30. Ferrigno, O.; Virolle, T.; Djabari, Z.; Ortonne, J.P.; White, R.J.; Aberdam, D. Transposable b2 sine elements can provide mobile rna polymerase ii promoters. *Nat. Genet.* **2001**, *28*, 77–81.

31. Lunyak, V.V.; Prefontaine, G.G.; Nunez, E.; Cramer, T.; Ju, B.G.; Ohgi, K.A.; Hutt, K.; Roy, R.; Garcia-Diaz, A.; Zhu, X.; *et al.* Developmentally regulated activation of a SINE b2 repeat as a domain boundary in organogenesis. *Science* **2007**, *317*, 248–251.

32. Biasco, L.; Ambrosi, A.; Pellin, D.; Bartholomae, C.; Brigida, I.; Roncarolo, M.G.; di Serio, C.; von Kalle, C.; Schmidt, M.; Aiuti, A. 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.

33. Berry, C.; Hannenhalli, S.; Leipzig, J.; Bushman, F.D. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* **2006**, *2*, e157.

34. Lewinski, M.K.; Yamashita, M.; Emerman, M.; Ciuffi, A.; Marshall, H.; Crawford, G.; Collins, F.; Shinn, P.; Leipzig, J.; Hannenhalli, S.; *et al.* Retroviral DNA integration: Viral and cellular determinants of target-site selection. *PLoS Pathog.* **2006**, *2*, e60.

35. Cesana, D.; Sgualdino, J.; Rudilosso, L.; Merella, S.; Naldini, L.; Montini, E. Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. *J. Clin. Invest.* **2012**, *122*, 1667–1676.

36. Moiani, A.; Paleari, Y.; Sartori, D.; Mezzadra, R.; Miccio, A.; Cattoglio, C.; Cocchiarella, F.; Lidonnici, M.R.; Ferrari, G.; Mavilio, F. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J. Clin. Invest.* **2012**, *122*, 1653–1666.

37. Ambrosi, A.; Glad, I.K.; Pellin, D.; Cattoglio, C.; Mavilio, F.; di Serio, C.; Frigessi, A. Estimated comparative integration hotspots identify different behaviors of retroviral gene transfer vectors. *PLoS Comput. Biol.* **2011**, *7*, e1002292.

38. Ciuffi, A.; Llano, M.; Poeschla, E.; Hoffmann, C.; Leipzig, J.; Shinn, P.; Ecker, J.R.; Bushman, F. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* **2005**, *11*, 1287–1289.

39. Llano, M.; Vanegas, M.; Fregoso, O.; Saenz, D.; Chung, S.; Peretz, M.; Poeschla, E.M. LEDGF/p75 determines cellular trafficking of diverse lentiviral but not murine oncoretroviral integrase proteins and is a component of functional lentiviral preintegration complexes. *J. Virol.* **2004**, *78*, 9524–9537.

40. De Rijck, J.; de Kogel, C.; Demeulemeester, J.; Vets, S.; El Ashkar, S.; Malani, N.; Bushman, F.D.; Landuyt, B.; Husson, S.J.; Busschots, K.; *et al.* The BET family of proteins targets Moloney murine leukemia virus integration near transcription start sites. *Cell Rep.* **2013**, *5*, 886–894.

41. Gupta, S.S.; Maetzig, T.; Maertens, G.N.; Sharif, A.; Rothe, M.; Weidner-Glunde, M.; Galla, M.; Schambach, A.; Cherepanov, P.; Schulz, T.F. Bromo- and extraterminal domain chromatin regulators serve as cofactors for murine leukemia virus integration. *J. Virol.* **2013**, *87*, 12721–12736.

42. Sharma, A.; Larue, R.C.; Plumb, M.R.; Malani, N.; Male, F.; Slaughter, A.; Kessl, J.J.; Shkriabai, N.; Coward, E.; Aiyer, S.S.; *et al.* BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12036–12041.