THE JOURNAL OF GENE MEDICINE
J Gene Med 2006; **8**: 1197–1207.
Published online 8 September 2006 in Wiley InterScience (www.interscience.wiley.com) **DOI:** 10.1002/jgm.958

RESEARCH ARTICLE

Lentiviral vector integration sites in human NOD/SCID repopulating cells

Stephanie Laufs[1]*,†
Guillermo Guenechea[2†]
Africa Gonzalez-Murillo[2]
K. Zsuzsanna Nagy[1]
M. Luz Lozano[2]
Coral del Val[3]
Sunitha Jonnakuty[3]
Agnes Hotz-Wagenblatt[3]
W. Jens Zeller[1]
Juan A. Bueren[2]
Stefan Fruehauf[4]

[1]*Research Program Innovative Cancer Diagnostics and Therapy, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany*

[2]*Hematopoiesis and Gene Therapy Division, CIEMAT/Marcelino Botin Foundation, Avenida Complutense 22, 28040 Madrid, Spain*

[3]*Department of Molecular Biophysics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany*

[4]*Department of Internal Medicine V, University of Heidelberg, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany*

*Correspondence to:
Stephanie Laufs, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.
E-mail: s.laufs@dkfz.de

†These authors contributed equally to this work.

Received: 28 March 2006
Revised: 22 June 2006
Accepted: 26 June 2006

Abstract

Background Recent observations of insertional mutagenesis in preclinical and clinical settings emphasize the relevance of investigating comprehensively the spectrum of integration sites targeted by specific vectors.

Methods We followed the engraftment of lentivirally transduced human cord blood (CB) progenitor cells after transplantation into NOD/SCID mice using a self-inactivating HIV-1-derived vector expressing the enhanced green fluorescent protein (EGFP).

Results The mean of transduction of $CD34^+$ CB cells was 41%, as deduced from the percentage of $EGFP^+$ cells before transplantation. At 3 weeks post-transplantation, the average of $EGFP^+$ cells in the human cell population was $65 \pm 8\%$, and increased to $75 \pm 10\%$ at 12 weeks post-transplantation. In order to determine the proviral integration sites in human NOD/SCID repopulating cells (SRCs) we used the ligation-mediated polymerase chain reaction (LM-PCR) technique. Sixty-eight percent of the integrations were found to be located in RefSeq genes, most of them in intron regions. Twenty percent of these integrations occurred within a distance of 10 kb from the transcription start site; a percentage that is significantly lower compared to that observed in cells transduced by gammaretroviral vectors. Sixty-two percent of integrations occurred in genes with a biological function in cell metabolism, and four integrations were located in genes with a role in tumorigenesis.

Conclusions These investigations indicate that integration of lentiviral vectors in human repopulating cells capable of engrafting NOD/SCID mice preferentially occur in coding regions of the human genome. Nevertheless, the clustering of integrations at the transcriptional start is not as high as that observed for gammaretroviral vectors. Copyright © 2006 John Wiley & Sons, Ltd.

Keywords lentiviral vector transduction; cord blood progenitors cells; hematopoietic stem cell; scid repopulating cell; ligation-mediated PCR; insertional mutagenesis

Introduction

Gammaretroviral vectors are used as gene transfer vehicles in the majority of gene therapy protocols. The retroviral genome is encoded by a single-stranded RNA molecule, two of which homodimerize and are packaged in lipid-enveloped viral capsid particles. Following attachment and receptor-mediated entry into host cells, viral reverse transcriptase and integrase enzymes mediate

reverse transcription and integration of the virus genome into the host-cell chromatin [1–3]. Retroviral integration can lead to insertional mutagenesis and cellular transformation, as recently observed in an experimental model and also in a human gene therapy trial.

In the first case, a mouse leukemia developed as a consequence of the aberrant expression of the evi-1 proto-oncogene caused by a nearby integrated proviral genome [4]. In the corrective gene therapy trial for X-chromosomal severe combined immunodeficiency (X1-SCID) [5], integration of the retroviral vector in the LMO-2 proto-oncogene resulted in insertional mutagenesis and development of T-cell leukemia in three children [6–8].

Historically, the integration events of retroviruses were believed to be random, and the chance of accidentally disrupting or activating a gene was considered remote. Early retroviral integration studies proposed that condensed chromatin from inactive DNA regions disfavored retroviral integration, thereby concentrating the integration in more open active chromatin areas of the genome [9–11].

With the availability of the complete human genome sequence, large-scale sequence-based surveys of integration sites have become possible [12–14]. Schroder *et al*. [12] investigated the targeting of human immunodeficiency virus (HIV) and HIV-based vectors in a human lymphoid cell line (SupT1) and found that genes were strongly favored as integration acceptor sites. Global analysis of cellular transcription in SupT1 cells indicated that active genes were preferential integration targets [12]. Recently, Wu *et al*. [14] examined the targeting of pseudotyped murine leukemia virus (MLV) in human HeLa cells and found that MLV preferred integration near the start of transcriptional units (either upstream or downstream). In contrast to this report they observed that HIV-1 was integrated anywhere in the transcriptional unit [14]. Very recently, Mitchell *et al*. [15] reported contrasting results on the integration targeting mediated by retroviral vectors derived from avian sarcoma-leukosis virus (ASLV), HIV and MLV. These authors confirmed in two different primary cell types (peripheral blood mononuclear cells and IMR90 lung fibroblasts) that MLV integration is favored near the transcription start sites and that active genes are preferential targets for HIV [15].

Concerning the integration pattern of gammaretroviral vectors in human hematopoietic stem cells (HSCs), previous studies traced the transduced NOD/SCID repopulating cells (SRCs) by means of Southern blot clonal analyses [16,17]. Using a more sensitive assay – the linear amplification-mediated polymer chain reaction (LAM-PCR) – Ailles *et al*. [18] analyzed the lentiviral vector integrations using the same human-NOD/SCID model. In contrast to previous studies showing that human engraftment was predominantly produced by a few repopulating clones [16,17], LAM-PCR analyses of lentivirally transduced SRCs showed a polyclonal pattern of repopulation.

Further LAM-PCR analyses in SRCs subjected to lentiviral transduction showed multiple vector integrations per transduced precursor [19]. In that study, five integration sites were determined, one of which was located in the tumor-suppressor gene BRCA1 [19]. Using the same NOD/SCID mouse model, we showed for the first time with the ligation-mediated PCR (LM-PCR) method that the integration of gammaretroviral vectors in human SRCs (mobilized peripheral blood SRCs) is essentially nonrandom [20]. In this study the oncoretroviral vector (SF-vector) showed an increased integration frequency near the transcription start site [20]. Additionally, we observed that integrations occurred with a significantly increased frequency into chromosomes 17 and 19 [13].

Here we have investigated the kinetics of engraftment of lentivirally transduced SRCs, and determined the integration site profile of the lentiviral inserts in the genome of the SRCs. New data showing the pattern of integration of lentiviral vectors in the genome of human hematopoietic repopulating cells is presented.

Material and methods

Lentiviral vector production

Replication-defective self-inactivating HIV-based vectors expressing the enhanced green fluorescent protein (EGFP) under the control of an internal promoter, the immediate-early human cytomegalovirus (CMV), were pseudotyped with VSV-G and packaged by a conditional expression system that only uses a fractional set of HIV-1 genes and provides enhanced biosafety [21–23]. This vector also contains a central polypurine tract (cPPT) and the woodchuck hepatitis virus post regulatory element (Wpre). The cPPT sequence enhanced transduction to a higher frequency and mean fluorescence intensity (MFI) [24]. To increase EGFP expression in the transduced cells, a Wpre sequence was inserted downstream of EGFP. Briefly, vector stocks of VSV-G-pseudotyped lentivectors were prepared by calcium phosphate mediated four-plasmid transfection of 293T cells essentially as described. Briefly, 9 μg of the self-inactivating (SIN) transfer vector construct pRRLsin18.PPT.CMV.eGFP.Wpre [23,25], 5.85 μg and 2.25 μg of the third-generation packaging constructs pMDLg-pRRE and pRSV-Rev, respectively [21] and 3.15 μg of the VSV-G expressing construct pMD2.VSV.G were used for transfection of 3×10^6 293T cells in a p90 plate for 8 h in the presence of chloroquine diphosphate salt (97%, Sigma-Aldrich, St. Louis, MO, USA) at a final concentration of 33.3 mM in Dulbecco's modified Eagle's medium (DMEM, Gibco, Grand Island, NY, USA) with 10% heat-inactivated fetal bovine serum (FBS, Serum Supreme, Bio-Whittaker Europe, Verviers, Belgium). After the transfection, medium was replaced by fresh DMEM supplemented with 10% FBS, and the next day medium was replaced again. Supernatants were collected 40 h after transfection. The constructs were kindly provided by Dr. L. Naldini (San Raffaele Telethon Institute for Gene Therapy – "Vita-Salute San Raffaele" University Medical School, Milano, Italy). High-titer viral vector stocks were

DOI: 10.1002/jgm

prepared by two rounds of ultracentrifugation before freezing and storing at −70 °C. The functional titers of viral vectors were determined by infection of HeLa cells with serial dilutions of the vector stocks, followed by cytometric analysis for $EGFP^+$ cells. Viral titer was 1.2×10^8 transduction units (TU)/ml.

Isolation and lentiviral transduction of $CD34^+$ cells

Cells were obtained from human umbilical cord blood (CB) after normal full-term deliveries, according to the protocol approved by the Ethical Committee of the Centro de Transfusiones de la Comunidad de Madrid. $CD34^+$ cells were purified by an immunomagnetic method (Minimacs, Miltenyi Biotec, Gladbach, Germany). The purity of the $CD34^+$ population ranged from 95% to 99% as evaluated by flow cytometry using an EPICS Elite ESP analyzer (Coulter, Hialeah, FL, USA). $CD34^+$ cells were cryopreserved in liquid nitrogen, with 10% dimethyl sulfoxide (DMSO) and 20% FBS. After thawing, $CD34^+$ cells were diluted 1 : 10 with medium and centrifuged at 800 *g* for 15 min.

Purified CB $CD34^+$ cells were transduced for 24 h in flat-bottomed P24-well plates (Nunc, Burlington, ON, Canada). Cells were well mixed with the lentiviral supernatant at a density of 5×10^5 cells/ml in StemSpan serum-free medium (Stem Cell Technologies Inc., Vancouver, BC, Canada) supplemented with 100 ng/ml of thrombopoietin (TPO, kindly provided by Kirin Brewery, Tokyo, Japan) and progenipoietin (ProGP, kindly provided by Monsanto Co., St. Louis, MO, USA), and 300 ng/ml of stem cell factor (SCF, kindly provided by Amgen, Thousand Oaks, CA, USA). The multiplicity of infection (MOI) was 10 TU/cell. One hour after the mixture had been incubated at 37 °C and 5% CO_2, the plates were centrifuged at 2500 rpm for 90 min. Next day, before harvesting the cells, centrifugation was performed once more. After 24 h of infection, cells were harvested with the help of cell dissociation buffer (Gibco, Grand Island, NY, USA). To eliminate the contamination with the lentiviral vector, cells were washed twice and resuspended in IMDM.

Scid repopulating cell (SRC) assay

NOD/LtSz-scid/scid (NOD/SCID) mice (deficient in Fc receptors, complement function, natural killer, B-, and T-cell function) were used as recipients of the human hematopoietic cells. Mice were purchased from The Jackson Laboratory (Bar Harbor, ME, USA). All animals were handled under sterile conditions and maintained under microisolators. Before transplantation, 6- to 8-week-old mice were total body irradiated with 2.5 to 3.0 Gy of X-rays. The dose rate was 1.03 Gy/min using a Philips MG 324 X-ray equipment (Philips, Hamburg, Germany), at 300 kV, 10 mA. Mice were transplanted with 3×10^5 unsorted cells obtained immediately after the 24 h transduction protocol.

Flow cytometric analysis

Phenotype and EGFP expression analyses in transduced cells and in transplanted mice were conducted essentially as previously described [26–28]. A minimum of 1×10^5 cells was incubated with 0.1% bovine serum albumin (BSA) in phosphate-buffered saline, and stained with phycoerythrin (PE)-conjugated anti-CD34 monoclonal antibody (mAb) (Anti-HPCA-2, Becton Dickinson, San Jose, CA, USA) for 20 min at room temperature. Cells were washed and analyzed using an EPICS ELITE-ESP cytometer (Coulter). Cells within a forward versus side scatter gate were analyzed for the expression of EGFP and CD34 antigen. PE- and FITC-conjugated mouse isotypic mAbs served as controls. For analyzing NOD/SCID recipients, hematopoietic samples were aspirated from the femoral bone marrow (BM) at periodic intervals following transplantation, as previously described [26–28]. At the end of the experiments, generally 90 to 120 days after transplantation, mice were killed and BM, peripheral blood (PB), spleen (Sp), and thymus (Thy) cells were analyzed by flow cytometry for the presence of human cells and EGFP expression. Aliquots containing 0.5 to 1×10^5 cells were stained with anti-human CD45-PECy5 mAb (Clone J33, Immunotech, Marseille, France) in combination with either anti-human CD34-PE mAb (Clone 8G12, Becton Dickinson), anti-human CD33-PE mAb (P67.6, Becton Dickinson), anti-human CD19-PE (J4.119, Immunotech), or anti-human CD3-PE (Clone UCHT1, Immunotech). Thereafter, red blood cells were lysed by adding 2.5 ml lysis solution (0.155 mol/l NH_4Cl, 0.01 mol/l $KHCO_3$, 1024 mol/l EDTA) and incubating at room temperature for 10 min. Cells were then washed in PBA (phosphate-buffered salt solution with 0.1% BSA and 0.01% sodium azide), resuspended in PBA + 2 μg/ml propidium iodide added to the cells before the analysis, and analyzed by flow cytometry. In all instances, mAbs were titrated with reference cells and used at saturating concentrations. Cells labeled with conjugated nonspecific isotypic mAbs were used as controls. In addition, BM cells from nontransplanted NOD/SCID mice were also stained with the same anti-human mAbs. Off-line analysis was done with the WinMDI (http://www.cyto.purdue.edu/flowcyt/software/Winmdi.htm) free software package (a kind gift from Dr. J. Trotter, The Scripps Research Institute, La Jolla, CA, USA). Aliquots from all these samples (BM, Sp, PB and Thy) were collected for DNA extraction and LM-PCR analysis.

Ex vivo expansion of NOD/SCID bone marrow

At the end of the experiments, when mice were killed, an aliquot of the BM was *ex vivo* cultured in

DOI: 10.1002/jgm

StemSpan serum-free medium and in the presence of TPO (100 ng/ml), ProGP (100 ng/ml) and SCF (300 ng/ml) for 7 days. Cell concentration ranged from 1 to 2.5 $\times$ 10^6 cells/ml. After expansion, BM cells were analyzed for human engraftment and gene transfer efficiency by flow cytometry. DNA was also extracted from an aliquot of these samples to perform LM-PCR.

Ligation-mediated PCR

For detection of retroviral integration sites, DNA was extracted from two NOD/SCID mouse chimeric BM preparations and corresponding spleen (QiaAmp blood kit, Qiagen, Hilden, Germany). As shown in Table 1, an average number of 4–13 integration sites were detected per each LM-PCR. Ligation-mediated (LM) PCR was performed as described previously [13]. Briefly, isolated BM DNA was digested with the restriction enzyme PvuII. Retroviral-human DNA junctions were marked using a biotinylated long terminal repeat (LTR)-specific primer followed by enrichment of the biotin-marked fragments (Dynal). Next, an adapter oligo-cassette was ligated blunt-end to the LTR-distant portion of enriched fragments to create binding sites for forward primers. Nested PCR was performed on the purified fragments. PCR products were analyzed on agarose gels, and gel blocks were excised and purified using a gel extraction kit (Qiagen). DNA of each excised gel block was cloned into pCR4 plasmid vector (Topo TA Cloning Kit, Invitrogen) according to the manufacturer's instructions. Colonies of each cloning reaction were screened for insert length by direct PCR of bacterial colonies with standard vector-primers T3 and T7.

Sequence analysis

Sequencing of the plasmids containing LM-PCR amplicon inserts was performed using an ABI Prism genetic analyzer 310 (PE Applied Biosystems, Weiterstadt, Germany) according to the manufacturer's instructions.

The sequences were first viewed using Chromas 2.23 software (Technelysium Pty Ltd., Tewantin, Australia). Sequence matches were judged to be authentic only if the matching part of the human query sequence was surrounded by the 5'LTR-sequence on the one side and the adapter-sequence on the other side. To find out identical sequences from different cloning reactions within one mouse the sequences were aligned with Clustal W [29] integrated in W2H/HUSAR [30]. For sequence analysis the BLAT tool with the UCSC Human Genome Project Working Draft was used (Assembly May 2004, http://genome.ucsc.edu/). The next criteria was that the match to the human genome extended over the length of the high-quality sequence with average identity >99.5%. The UCSC Genome Browser was used to determine the location of the match in respect to exons and introns. As described before an integration was defined as having landed in a gene only if it was between the transcriptional start and transcriptional stop boundaries of one of the 20471 RefSeq genes [31] (Freeze of 5 February, 2004; SRS server at DKFZ, Heidelberg http://genius.embnet.dkfz-heidelberg.de/menu/srs/databanks.html). The sequence segments of the human genome accepted as matches always started exactly at the last base of the viral DNA.

Table 1. Human transduced repopulating clones detected in NOD/SCID mice in the BM 90 days post-transplantation (dpt) and 7 days *ex vivo* expanded (EE)-BM from mice A and B. Clones from the spleen of mouse B are also presented. Common clones between fresh and EE samples are underlined. Common clones between 1st and 2nd LM-PCRs are marked with an asterisk (*)

	Mouse A		Mouse B		
	BM		BM		Spleen
Sample	Fresh	EE	Fresh	EE	Fresh
Clones 1. LM-PCR	**A1**	**A1**	**B1**	**B1***	B13
	A2	A6	**B2**	**B2**	B14
	A3	A7	**B3***	**B3**	B15
	A4	A8	B4	B5	B16
	A5			B6	B17
				B7	B18
				B8	B19
				B9	
				B10	
				B11	
				B12	
Clones 2. LM-PCR	C1	**C2**	B3*	B1*	**D11**
	C2	C6	**D1**	**D1**	
	C3	C7	D2	**D7**	
	C4	C8	D3	**D8**	
	C5	C9	D4	**D10**	
		C10	D5	D12	
		C11	D6	D13	
		C12	**D7**	D14	
		C13	**D8**	D15	
		C14	D9		
		C15	**D10**		
			D11		
N° of different clones (LM-PCR 1 and 2)	**10**	**15**	**15**	**19**	**8**
Common clones	**2**		**7**		**1**
N° of different clones per mouse	**23**		**34**		

Gene ontology (GO) annotations

Using the vocabularies developed by the Gene Ontology™ (GO) Consortium [32] the functional annotation was established. The Gene Ontology organizes the GO terms in a directed acyclic graph (DAG) in a hierarchical manner. GO provides controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. Therefore, genes with similar functions or genes involved in the same biological process can be readily identified by common GO annotations. To obtain the organism-specific GO functional annotation 'GO slims' (the slimmed-down versions of the three ontology categories) were constructed. The mapping of genes to GO terms in the

DOI: 10.1002/jgm

context of the DAG structure allows viewing the three GO ontologies at higher levels of the GO DAG.

Statistical analysis

To test whether the number of integrations is equally distributed along the chromosomes we used a chi-squared goodness-of-fit test to analyze whether the observed number of integrations (oi) arose from a multinomial distribution with specified expected integrations (ei) for the 24 chromosomes (22 autosomes as well as the sex chromosomes X and Y). Expected integration counts were computed assuming a discrete uniform distribution but also correcting for the chromosome size distribution [ei = (determined bases of chromosome (bp)/determined bases of genome (bp)) × total number of mapped integrations]. For example, for chromosome 11, we found 6 integrations (oi = 6). For the same chromosome we calculated ei = (determined bases of chromosome (134,48 MB)/determined bases of genome (3070,13)) ×28) = 1.2. Chromosome and genome size were calculated with the Human genome Release 24.35e.1. The obtained differences are highly significant as evidenced by the RISC score (**r**etroviral **i**nsertion **s**ite in **c**hromosome) defined as: $(oi\text{-}ei)^2/ei$ values [oi < ei then RISC score x (−1)] [13]. The RISC score for chromosome 11 is thus calculated as following: $(6–1.2)^2/1.2 = 19.2$. In the case of nonpreferential integration a RISC score of 0 would be expected. We used $\alpha = 0.05$ as significance level of the test. For the detection of preferred genomic integration sites a cut-off of $(oi\text{-}ei)^2/ei \geq 3$ was set.

Results

In this study we aimed to analyze the distribution of proviral integrations in the genome of human

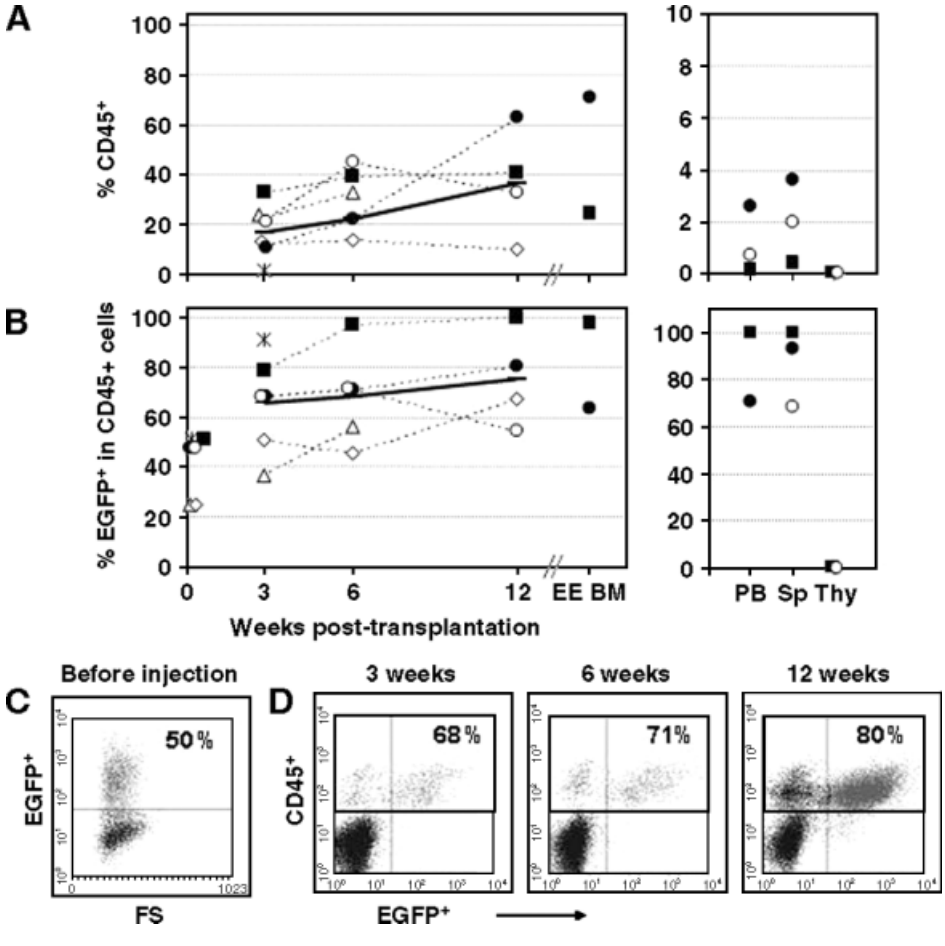

Figure 1. Kinetics of human engraftment (A) and gene transfer efficiency (B) of individual NOD/SCID mice transplanted with transduced CD34$^+$ cord blood (CB) cells. At 3 and 6 weeks post-transplantation femoral bone marrow (BM) samples were obtained from the recipients by BM aspiration. Twelve weeks post-transplantation the mice were sacrificed and BM, peripheral blood (PB), spleen (Sp) and thymus (Thy) were analyzed by flow cytometry. Data from *ex vivo* expanded (EE)-BM are also presented. Closed symbols correspond to mice further analyzed by LM-PCR: mouse A (■) and mouse B (●). Mean values are presented by ▬ symbol (n = 6). Data at day 0 represent the percentage of EGFP$^+$ cells in the total population 3 days after transduction in liquid culture. (C) Flow cytometric analysis of CD34$^+$ CB cells 3 days after transduction. (D) Kinetics of the engraftment of transduced CD34$^+$ human umbilical CB cells in mouse A at different times post-transplantation. Dot plots show analysis of cells labeled with anti-CD45-Cy5 antibody to determine the proportion of human cells (D). Quadrants were set according to isotype-matched negative controls. The percentage indicates the proportion of EGFP$^+$ cells within the human cells

DOI: 10.1002/jgm

hematopoietic SRCs transduced with lentiviral vectors. For this purpose, human $CD34^+$ cord blood cells were transduced with a self-inactivating HIV-1-derived vector expressing the enhanced green fluorescent protein (EGFP) [24], and transplanted into NOD/SCID mice. Hematopoietic cells generated by the human SRCs were analyzed by flow cytometry at different times after transplantation. At the end of the experiments, the pattern of lentiviral insertion sites in human hematopoietic cells was determined by LM-PCR.

Kinetics of human engraftment and EGFP expression *in vivo*

$CD34^+$ samples subjected to lentiviral transduction showed a range of 25–50% $EGFP^+$ cells. These data were obtained from cell aliquots maintained for 3 days in culture after the transduction period (Figures 1B and 1C). Immediately after the 24 h transduction procedure, samples consisting of 3×10^5 cells were transplanted into sublethally irradiated NOD/SCID mice. Between 3–12 weeks post-transplantation, femoral bone marrow (BM) samples were aspirated to evaluate the kinetics of engraftment of the human transduced samples. Flow cytometric analyses of $CD45^+$ cells in the NOD/SCID mouse BM showed a mean level of engraftment of $CD45^+$ cells of 17% at 3 weeks post-transplantation. A progressive increase in the level of human engraftment was observed in most animals at longer times post-transplantation, resulting in a mean engraftment of 37% after 12 weeks (Figure 1A). As shown in Figure 1B, high levels of human $EGFP^+$-transduced cells were always detected in the NOD/SCID mice. The proportion of $EGFP^+$ cells within the $CD45^+$ population was significantly higher after transplantation, compared to *in vitro* data obtained 3 days after transduction (mean value: $41 \pm 4\%$). Moreover, in most animals, the proportion of transduced cells increased along the post-transplantation period (mean values at 3 and 12 weeks post-transplantation $\pm$standard error of the mean (SEM): $65 \pm 8\%$ and $75 \pm 10\%$, respectively; see representative analyses in Figures 1C and 1D).

At the end of the experiments, femoral BM, as well as peripheral blood (PB), spleen (Sp) and thymus (Thy), were obtained and analyzed by flow cytometry. Additionally, an aliquot of BM cells was subjected to a 7-day *ex vivo* expansion (EE) process, as indicated in Materials and Methods. Consistent with our previous observations [26], the level of human engraftment in PB, Sp and Thy was markedly lower compared with the engraftment observed in BM (Figure 1A). In spite of this observation, the proportion of $EGFP^+$ cells within the $CD45^+$ cells in these tissues was similar to that observed in fresh or EE-BM (Figure 1B). Additionally, the presence of $EGFP^+$ cells in the different hematopoietic lineages was also confirmed in all BM samples (data not shown), consistent with the notion that very primitive multipotent repopulating cells had been transduced with the lentiviral vectors.

Detection and genomic analysis of lentiviral integration sites

Aiming to investigate the integration pattern of lentiviral vectors in SRCs, mice with the highest levels of $EGFP^+$ cells (mice A and B in Figure 1) were selected for conducting LM-PCR integration site analysis (Figure 2). In addition to fresh samples harvested 12 weeks after transplantation, BM cells subjected to EE were also analyzed to facilitate the identification of potentially nonproliferating transduced progenitors (Table 1). Each sample was subjected to two LM-PCRs, except the EE-BM from mouse B, on which three LM-PCRs were performed. A total of 57 different integrations from these donors were detected by LM-PCR.

In mouse A, ten clones were detected in fresh BM (A1–A5, C1–C5). Two of these clones were also found in EE-BM (A1 and C2), while 13 clones not found in fresh BM were observed after EE (A6-A8, C6-C15). In mouse B, 15 clones were identified in fresh BM (B1–B4, D1–D11). Seven of these clones were also detected in the EE-BM (B1–B3, D1, D7, D8, D10), while 12 new clones became apparent after EE (B5–B12, D12–D15). The spleen of mouse B was also subjected to insertional analysis and seven new clones were detected (B13–B19), while one clone was already detected in fresh BM of this mouse (D11).

Integration sites were only analyzed when they met the quality criteria, i.e. detection of the sequence of the corresponding LTR on one end of the PCR product and the adapter sequence on the other end. We also detected integrations consisting of one or more LTRs back to back,

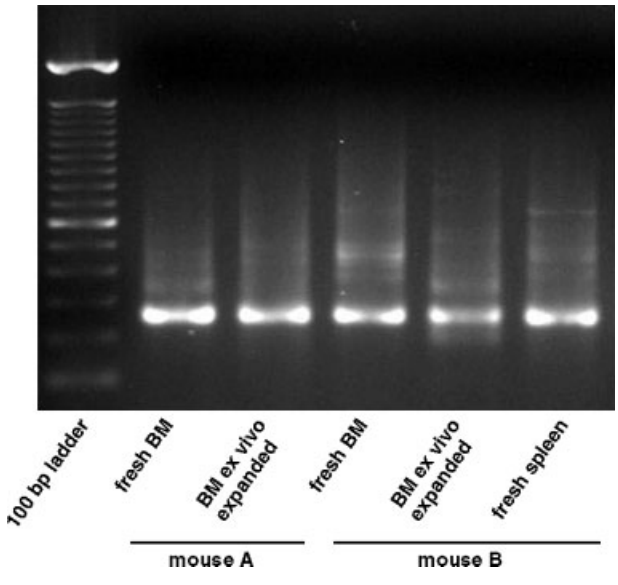

Figure 2. LM-PCR gel electrophoresis. DNA was extracted from different samples (mouse A: fresh bone marrow (BM), *ex vivo* expanded (EE)-BM; mouse B: fresh BM, EE-BM, and fresh spleen) and LM-PCR was performed (marker: 100 bp ladder)

DOI: 10.1002/jgm

Table 2. Chromosomal localization of HIV-derived vector integration sites with corresponding hit RefSeq gene (Dist. to TSS = distance to transcription start site)

Integration ID	Chrom. localization	Accession number	RefSeq	Hit gene	Exon/Intron	Dist. to TSS (kb)
A1 (×2)	–	–	–	–	–	–
A2	3q27.1	NT_005612	NM_032047	beta-1 3-N-acetylglucosaminyltransferase bGnT-5	3′UTR	−81
A3	20q11.22	NT_028392	NM_018677	acyl-CoA synthetase short-chain family member 2	Intron 2	1213
A4	–	–	–	LTR: family ERV1	–	–
A5	11p11.12	NT_009237	NM_004476	folate hydrolase 1 isoform 1	Intron 3	64
A6	18q11.2	NT_010966	NM_015461	early hematopoietic zinc finger	Intron 4	−258
A7	9q34.3	NT_024000	NM_152286	chromosome 9 open reading frame 111	Intron 22	−71
A8	16q12.1	NT_010498	NM_133443	alanine aminotransferase 2	Exon 5	−21
B1(×2)	11q13.1	NT_033903	NM_031471	UNC-112 related protein 2 long form	Intron 6	501
B2 (×2)	2q31.3	NT_005403	–	–	–	220
B3 (×2)	16q22.1	NT_010498	NM_145059	fucokinase	Intron 1	7
B4	Xq27.3	NT_011681	–	–	–	−550
B5	21q22.2	NT_011512	NM_018963	bromodomain and WD repeat domain	Intron 14	−50
B6	20p12.1	NT_011387	NM_017714	taspase 1	Intron 10	−181
B7	–	–	–	SINE: family Alu Sg	–	–
B8	5q31.2	NT_034772	NM_199189	matrin 3	Intron 5	36
B9	21q21.1	NT_011512	–	–	–	−315
B10	3q21.3	NT_005612	NM_004637	RAB7 member RAS oncogene family	Intron 1	35
B11	6p23	NT_007592	NM_005493	RAN binding protein 9	Intron 1	−14
B12	7q11.22	NT_007758	NM_015570	autism susceptibility candidate 2	Intron 5	1064
B13	1q32.1	NT_034410	–	–	–	−344
B14	11p15.5	NT_035113	NM_006755	transaldolase 1	Intron 3	12
B15	–	–	–	SINE: family Alu Sx	–	–
B16	1q21.3	NT_004487	–	–	–	59
B17	10q22.2	NT_008583	NM_145170	tetratricopeptide repeat domain 18	Intron 7	−22
B18	–	–	–	LINE: family AluJb	–	–
B19	20q13.32	NT_011362	–	–	–	95
C1	–	–	–	SINE: family AluJb	–	–
C2 (×2)	1p34.1	NT_032977	NM_006369	MUF1 protein	Intron 6	−22
C3	6q16.2	NT_025741	NM_032870	hypothetical protein LOC25957	Intron 7	−10
C4	8p21.3	NT_023666	NM_018411	hairless protein isoform b	Intron 2	−2
C5	11p15.4	NT_009237	–	–	–	75
C6	–	–	–	LTR: family ERVK, LTR5 Hs	–	–
C7	–	–	–	LINE: family L1PA13	–	–
C8	–	–	–	LINE: family L1MA7	–	–
C9	–	–	–	LINE: family L1ME	–	–
C10	–	–	–	SINE: family AluSq	–	–
C11	–	–	–	SINE: family AluSx	–	–
C12	–	–	–	SINE: family AluY	–	–
C13	14q11.2	NT_026437	–	–	–	−7
C14	–	–	–	LINE: family L2	–	–
C15	12q13.12	NT_029419	NM_023071	spermatogenesis associated serine-rich 2	Intron 1	74
D1 (×2)	22q12.1	NT_011520	–	–	–	−264
D2	8p11.21	NT_007995	NM_006766	MYST histone acetyltransferase	Intron 2	−11
D3	15q22.31	NT_010194	NM_006660	ClpX caseinolytic protease X homolog	Intron 3	−15
D4	19q13.33	NT_011109	–	–	–	25
D5	14q24.1	NT_026437	NM_003861	Breakpoint cluster region protein uterine	Intron 6	31
D6	17q11.2	NT_010799	NM_015194	myosin ID	Intron 16	−220
D7 (×2)	7p21.3	NT_007819	–	–	–	−708
D8 (×2)	6q16.2	NT_025741	NM_032870	hypothetical protein LOC25957	Intron 7	−10
D9	–	–	–	SINE: family AluSx	–	–
D10 (×2)	11q13.3	NT_033927	NM_177423	PTPRF interacting protein alpha 1 isoform a	Intron 2	3
D11 (×2)	5q35.2	NT_023133	NM_020444	hypothetical protein LOC57179	Intron 4	7
D12	21q22.2	NT_011512	NM_018963	bromodomain and WD repeat domain	Intron 17	−50
D13	6p21.32	NT_007592	–	–	–	45
D14	–	–	–	LINE: family L1	–	–
D15	11q13.1	NT_033903	–	–	–	3

which were not included in this analysis (data not shown). One detected insertion site was too short to be mapped (A1) and 15 integration sites were found to be in repetitive elements (SINES, LINES, LTR, etc.) of the genome so that the exact location of these insertion sites could not be ascertained (Table 2). The remaining 41 integrations could be unambiguously mapped to the human genome and were further analyzed.

To test whether the number of integrations is equally distributed along the chromosomes we used a chi-squared goodness-of-fit test to analyze whether the observed number of integrations (oi) arose from a multinomial distribution with specified expected integrations (ei) for the 24 chromosomes (22 autosomes as well as the sex chromosomes X and Y). The chromosomal integration analysis as given in Table 2 shows that vector integrations occurred

DOI: 10.1002/jgm

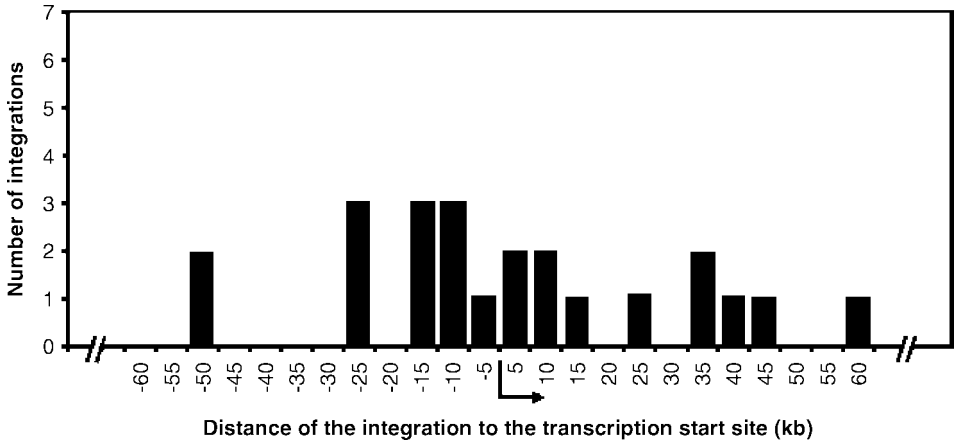

Figure 3. Distance of the detected integration site to the transcription start. We selected windows of varying sizes over 5 kb each upstream and downstream (0 to ±60 kb; n = 23) of the transcriptional start site for all integrations landing in RefSeq genes. The total number of integrations of each window is shown in the graph

with significantly increased frequency into chromosome 11. On chromosome 11, a total of 6 different integrations were observed (clones A5, B1, B14, C5, D10, D15), while 1.8 integrations were expected (ei). Expected integration counts were computed assuming a discrete uniform distribution but also correcting for the chromosome size distribution (see Material and Methods). These findings suggest that lentiviral vector integration into chromosomes of hematopoietic SRCs is nonrandom. While this seemingly nonrandom chromosomal distribution of integrations is intriguing, the results here have to be judged with caution because analysis comprised only 57 different NOD/SCID repopulating, lentivirally transduced clones.

In a next step, detailed gene analysis was performed on fresh BM cells from the NOD/SCID mice, and in some instances after EE of these cells. A total of 57 integration sites were identified (Table 2 and Figure 3). It turned out that 8 of 41 (20%) unambiguously mapped integrations occurred within a distance of 10 kb up- and downstream from the transcription start site (Figure 3). Twenty-eight out of the 41 integrations (68%; see Table 2) were found to be located within human RefSeq genes. Among these 28 integrations, 26 integrations (93%) landed in intron regions of the genes and 4 integrations (14%) occurred in the first intron (Table 2). No evident differences between integration analyses conducted in fresh and EE samples were obtained, indicating that, at least after a short EE of 7 days, there is no significant selection of genetically marked clones.

Analysis of targeted genes

A detailed list of the 28 hit RefSeq genes is shown in Table 2. Two integration sites were located in genes with a role in tumorigenesis, which are the RANBP9 (ran binding protein 9 gene, NM_005493) and the EHZF (early hematopoietic zinc finger gene, NM_015461). One integration site occurred in the WDR22 gene, encoding for a break point cluster region protein. Another integration site was detected in the MYST3 gene. A chromosomal aberration involving MYST3 may be a cause of acute myeloid leukemias [33].

The annotation with Gene Ontology (GO) of the molecular function shows that the most frequent integrations occurred in genes with hydrolase and transferase functions (20% each), followed by metal ion binding (15%), nucleic acid binding (15%) and nucleotide binding (Figure 4A). Regarding the distribution of insertions in genes participating in the different biological processes, 62% of the targeted genes are involved in metabolic processes, 19% in cell growth and/or maintenance and 13% in cell communication (Figure 4B). Almost all genes with hydrolase activity are participating in metabolic processes while the genes with transferase activity are the only ones involved in morphogenesis processes.

Discussion

Here we report the integration characteristics of lentiviral vectors in human hematopoietic precursors that progressively repopulated the hematopoiesis of NOD/SCID mice. Human engraftment increased over a period of 12 weeks just as the proportion of $EGFP^+$ cells. These results prove a long-term engraftment and show that the transduced cells are functionally primitive precursors. Only in one mouse did the percentage of $EGFP^+$ cells decrease with time, showing that the contribution of transduced cells was not equal at different times post-transplantation. This result also suggests that the transduced SRCs detected at 12 weeks after transplantation had lower proliferative potential than the others detected in the rest of the mice.

The progressive increase in the level of human engraftment was associated with an increase in gene marked cells. However, hematopoiesis was polyclonal at the end of the study (12 weeks, Table 1) and there was no indication of clonal dominance (Figure 2) as has been described with different vectors and longer-term follow-up in mice [34] and patients [35].

DOI: 10.1002/jgm

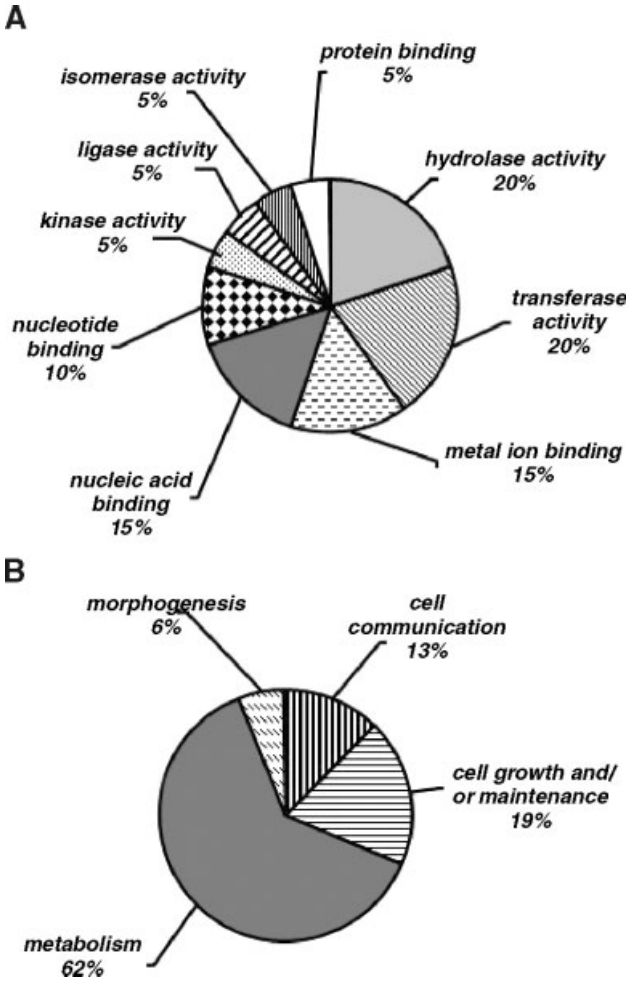

Figure 4. Gene ontology (GO) annotation of hit RefSeq genes shown in the categories of molecular function (A) and biological process (B)

Twenty-eight out of the 41 integrations that mapped to the human genome (68%) occurred in RefSeq genes. It is apparent that this proportion is significantly higher from that deduced of a computer-simulated random data set in which 22% of 10 000 simulated integrations landed in RefSeq genes [14]. Results published in established cell lines and human hematopoietic cells showed that a high proportion of MLV and HIV integrations landed in RefSeq genes [14,15,20,36].

When a detailed gene analysis of integrations determined in our study was performed, 93% (26/28) of the analyzed integrations in RefSeq genes landed in intron sequences. These results are in line with data from Imren *et al*. [37], showing that 88% of lentivirally transduced human cord blood cells are intragenic integration events which occurred in introns [37]. This proportion is significantly higher compared to the 21% determined as intron sequences in the human genome analysis by Venter *et al*. [38].

Recent data has shown the different integration pattern of MLV and HIV vectors within transcribed genes [39]. Whereas MLV-based traps showed a strong bias for promoter-proximal integration – leading to efficient reporter expression, HIV-based traps integrated throughout transcriptional units, and were limited for expression by the distance from the promoter and the reading frame of the targeted gene [39]. Interestingly, the fraction of HIV integrations resulting in efficient reporter expression were those occurring proximal to a promoter or within the first intron [39]. A previous *in vitro* study in HeLa cells and human H9 cells revealed that integration of MLV is favored in regions near the transcription start sites, but no such bias was seen for HIV [14]. Furthermore, Mitchell *et al*. [15] recently showed in peripheral blood mononuclear cells, and also in human H9 cells, that the integration of HIV vectors is not markedly favored in regions close to the transcription start site, as seen with MLV. In our study, 20% of integrations mapped to the human genome occurred within a distance of ±10 kb of the transcription start site. To compare with previously published data we calculated also the ±5 kb distance to the transcription start site which was 7% for our data. Although this proportion (20% for ±10 kb or 7% for ±5 kb) is higher than the expected proportion in the case of random integration (4% for ±5 kb) [14], it is significantly lower than that observed in cell lines (20% for ±5 kb) [14], human SRCs (19% for ±5 kb) [20] or human T cells (30–36% ±10 kb) [40] previously transduced with gammaretroviral vectors. Whether or not a preferential expansion of NOD/SCID repopulating clones with insertions close to the transcription start partially account for our results, as recently observed in retrovirally transduced mouse HSCs [34], is currently unknown.

Regarding the functionality of targeted genes, in mouse A one common clone was too short to be mapped within the human genome (clone A1). The second common clone in fresh and EE-BM (C2) was located in LRRC41 (leucine rich repeat containing 41) which is involved in the ubiquitin cycle [41]. In mouse B seven different clones were detected in fresh as well as in EE-BM. While three of them did not land within RefSeq genes (B2, D1, D7), one of the integration sites was in a gene the function of which has not yet been described (B1) and a further occurred in a gene defined as hypothetical protein (D8). One common integration occurred in the gene coding for the PTPRF interacting protein alpha 1 isoform a (D10). A further integration site was in the gene coding for the L-fucokinase (B3). Fucokinase participates in a salvage pathway for reutilization of fucose from oligosaccharide degradation [42]. The sugar fucose is present in glycoproteins and glycolipids and is involved in blood group antigen recognition [43], inflammation [44], and metastasis [45]. The integration site of the clone B1 occurred in the UNC-112 related protein 2 gene (URP2). The expression of URP2 appears to be confined primarily to tissues of the immune system [46]. We were able to detect one common clone in the fresh BM of mouse B as well as in the spleen on the same mouse (D11, occurred in a gene defined as hypothetical protein).

Further integration sites were located in genes with a role in tumorigenesis, e.g. RAB7 (Ras related protein 7) which is localized on 3q21, a region in which

DOI: 10.1002/jgm

translations or deletions were frequently described in association with leukemia. Additionally, if the results are compared with integration results obtained by analysis of oncoretroviral vector (SF91m3)-transduced human peripheral progenitor cells following transplantation and engraftment into NOD/SCID immunodeficient mice, common integration sites between both studies can be detected, such as the early hematopoietic zinc finger (EHZF, NM_015461) and the N-acetylated alpha-linked acidic dipeptidase-like gene (NM_005468). EHZF expression is abundant in human $CD34^+$ progenitors and declines rapidly during cytokine-driven differentiation [47]. EHZF has 96% identity to mouse evi3, a recently identified gene associated with the retroviral integration site in AKXD-27 B-cell lymphomas [48]. Bond and colleagues [47] found significant mRNA levels in the majority of acute myelogenous leukemias. It has been suggested that EHZF is likely to play a relevant role in the control of human hematopoiesis. Furthermore, EHZF might be implicated in the development of hematopoietic malignancies.

Interestingly, three integrations occurred in the genomic region 11q13. This region has been confirmed as the breakpoint region in multiple myeloma (MM) [49]. These authors showed that 11q13 breakpoints in MM are scattered along the 11q13. These data might allow the hypothesis that retroviral integrations occur preferably in such breakpoint regions. Whether these regions might have a steric advantage for lentiviral integration has to be investigated in greater detail.

Taken together, our findings offer new data showing the pattern of lentiviral vector integration in human HSCs. A preferential integration within intronic sequences of Refseq genes is deduced from our observations. Nevertheless, compared to previous studies with gammaretroviral integrations, the proportion of integration within a ± 10 kb distance of the transcription start site is lower. However, if transactivation has been shown over more than 100 kb, this does not mean that lentiviral vectors can be *a priori* considered to be safer than gammaretroviral vectors. The knowledge of the preferential integration sites of viral vectors in stem cells will help to understand the mechanisms of integration and/or clonal selection, and eventually facilitate the development of new vectors with targeted integration sites to improve the safety of human gene therapy trials.

Acknowledgements

The technical assistance of Bernhard Berkus, Hans-Jürgen Engel, Sigrid Heil and the support of the animal facility team of the German Cancer Research Center are gratefully acknowledged. The authors also want to thank Jesús Martínez, Israel Orman, Aurora de la Cal, Maria Dolores López, Sergio García and Elena López for their expert experimental assistance. We thank Dr. L. Naldini for the lentiviral vector. Supported by grant M 20.4 of the H.W. & J. Hector-Stiftung, by grant FR 1732/3-1of the Deutsche Forschungsgemeinschaft, and by grants from the *Dirección General de Investigación de la Comunidad de Madrid* (Ref.08.6/0013.1/2001); Sixth Framework RTD Program of Life Sciences, Genomics and Biotechnology for Health (LSHB-CT-2004-005242); *Comisión Interministerial de Ciencia y Tecnología* (SAF2005-00058). A.G-M. is a recipient of a fellowship from CIEMAT.

References

1. Brown PO. In *Retroviruses*, Coffin JM, Hughes SH, Varmus HE (eds). Harbor Laboratory Press: New York, 1997; 161–203.
2. Bushman FD. *Lateral DNA Transfer: Mechanisms and Consequences*, Cold Spring Harbor Laboratory Press: New York, 2001.
3. Bushman FD. Targeting retroviral integration? *Mol Ther* 2002; **6**: 570–571.
4. Li Z, Dullmann J, Schiedlmeier B, *et al*. Murine leukemia induced by retroviral gene marking. *Science* 2002; **296**: 497.
5. Cavazzana-Calvo M, Hacein-Bey S, de Saint Basile G, *et al*. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* 2000; **288**: 669–672.
6. Hacein-Bey-Abina S, von Kalle C, Schmidt M, *et al*. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 2003; **302**: 415–419; Erratum in: *Science* 2003; **302**: 568.
7. Hacein-Bey-Abina S, von Kalle C, Schmidt M, *et al*. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* 2003; **348**: 255–256.
8. Schmidt M, Hacein-Bey-Abina S, Wissler M, *et al*. Clonal evidence for the transduction of CD34+ cells with lymphomyeloid differentiation potential and self-renewal capacity in the SCID-X1 gene therapy trial. *Blood* 2005; **105**: 2699–2706.
9. Panet A, Cedar H. Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. *Cell* 1977; **11**: 933–940.
10. Vijaya S, Steffen DL, Robinson HL. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J Virol* 1986; **60**: 683–692.
11. Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* 1987; **61**: 336–343.
12. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 2002; **110**: 521–529.
13. Laufs S, Gentner B, Nagy KZ, *et al*. Retroviral vector integration occurs into preferred genomic targets of human hematopoietic stem cells with long-term bone marrow repopulating ability. *Blood* 2003; **101**: 2191–2198.
14. Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 2003; **300**: 1749–1751.
15. Mitchell RS, Beitzel BF, Schroder AR, *et al*. Retroviral DNA Integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2004; **2**: E234.
16. Barquinero J, Segovia JC, Ramirez M, *et al*. Efficient transduction of human hematopoietic repopulating cells generating stable engraftment of transgene-expressing cells in NOD/SCID mice. *Blood* 2000; **95**: 3085–3093.
17. Guenechea G, Gan OI, Dorrell C, Dick JE. Distinct classes of human stem cells that differ in proliferative and self-renewal potential. *Nat Immunol* 2001; **2**: 75–82.
18. Ailles L, Schmidt M, Santoni de Sio FR, *et al*. Molecular evidence of lentiviral vector-mediated gene transfer into human self-renewing, multi-potent, long-term NOD/SCID repopulating hematopoietic cells. *Mol Ther* 2002; **6**: 615–626.
19. Woods NB, Muessig A, Schmidt M, *et al*. Lentiviral vector transduction of NOD/SCID repopulating cells results in multiple vector integrations per transduced cell: risk of insertional mutagenesis. *Blood* 2003; **101**: 1284–1289.
20. Laufs S, Nagy KZ, Giordano FA, Hotz-Wagenblatt A, Zeller WJ, Fruehauf S. Insertion of retroviral vectors in NOD/SCID repopulating human peripheral blood progenitor cells occurs preferentially in the vicinity of transcription start regions and in introns. *Mol Ther* 2004; **10**: 874–881.

21. Dull T, Zufferey R, Kelly M, *et al*. A third-generation lentivirus vector with a conditional packaging system. *J Virol* 1998; **72**: 8463–8471.
22. Naldini L, Blomer U, Gallay P, *et al*. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* 1996; **272**: 263–267.
23. Zufferey R, Dull T, Mandel RJ, *et al*. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol* 1998; **72**: 9873–9880.
24. Follenzi A, Ailles LE, Bakovic S, Geuna M, Naldini L. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat Genet* 2000; **25**: 217–222.
25. Zufferey R, Donello JE, Trono D, Hope TJ. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J Virol* 1999; **73**: 2886–2892.
26. Guenechea G, Segovia JC, Albella B, *et al*. Delayed engraftment of nonobese diabetic/severe combined immunodeficient mice transplanted with ex vivo-expanded human CD34(+) cord blood cells. *Blood* 1999; **93**: 1097–1105.
27. Drize NJ, Keller JR, Chertkov JL. Local clonal analysis of the hematopoietic system shows that multiple small short-living clones maintain life-long hematopoiesis in reconstituted mice. *Blood* 1996; **88**: 2927–2938.
28. Verlinden SF, van Es HH, van Bekkum DW. Serial bone marrow sampling for long-term follow up of human hematopoiesis in NOD/SCID mice. *Exp Hematol* 1998; **26**: 627–630.
29. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**: 4673–4680.
30. Senger M, Flores T, Glatting K, Ernst P, Hotz-Wagenblatt A, Suhai S. W2H: WWW interface to the GCG sequence analysis package. *Bioinformatics* 1998; **14**: 452–457.
31. Pruitt KD, Katz KS, Sicotte H, Maglott DR. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet* 2000; **16**: 44–47.
32. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**: 25–29.
33. Borrow J, Stanton VP Jr, Andresen JM, *et al*. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet* 1996; **14**: 33–41.
34. Kustikova O, Fehse B, Modlich U, *et al*. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* 2005; **308**: 1171–1174.
35. Ott MG, Schmidt M, Schwarzwaelder K, *et al*. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* 2006; **12**: 401–409.
36. Hematti P, Hong BK, Ferguson C, *et al*. Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol* 2004; **2**: e423.
37. Imren S, Fabry ME, Westerman KA, *et al*. High-level beta-globin expression and preferred intragenic integration after lentiviral transduction of human cord blood stem cells. *J Clin Invest* 2004; **114**: 953–962.
38. Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science* 2001; **291**: 1304–1351.
39. De Palma M, Montini E, de Sio FR, *et al*. Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. *Blood* 2005; **105**: 2307–2315.
40. Recchia A, Bonini C, Magnani Z, *et al*. Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. *Proc Natl Acad Sci U S A* 2006; **103**: 1457–1462.
41. Kamura T, Burian D, Yan Q, *et al*. Muf1, a novel elongin BC-interacting leucine-rich repeat protein that can assemble with Cul5 and Rbx1 to reconstitute a ubiquitin ligase. *J Biol Chem* 2001; **276**: 29748–29753.
42. Hinderlich S, Berger M, Blume A, Chen H, Ghaderi D, Bauer C. Identification of human L-fucose kinase amino acid sequence. *Biochem Biophys Res Commun* 2002; **294**: 650–654.
43. Lloyd KO. The chemistry and immunochemistry of blood group A, B, H, and Lewis antigens: past, present and future. *Glycoconjate J* 2000; **17**: 531–541.
44. Lasky LA. Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science* 1992; **258**: 964–969.
45. Marionneau S, Cailleau-Thomas A, Rocher J, *et al*. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* 2001; **83**: 565–573.
46. Weinstein EJ, Bourner M, Head R, Zakeri H, Bauer C, Mazzarella R. URP1: a member of a novel family of PH and FERM domain-containing membrane-associated proteins is significantly over-expressed in lung and colon carcinomas. *Biochim Biophys Acta* 2003; **1637**: 207–216.
47. Bond HM, Mesuraca M, Carbone E, *et al*. Early hematopoietic zinc finger protein (EHZF), the human homologue to mouse Evi3, is highly expressed in primitive human hematopoietic cells. *Blood* 2003; **103**: 2062–2070.
48. Warming S, Liu P, Suzuki T, *et al*. Evi3, a common retroviral integration site in murine B-cell lymphoma, encodes an EBFAZ-related kruppel-like zinc finger protein. *Blood* 2003; **101**: 1934–1940.
49. Ronchetti D, Finelli P, Richelda R, *et al*. Molecular analysis of 11q13 breakpoints in multiple myeloma. *Blood* 1999; **93**: 1330–1337.

DOI: 10.1002/jgm