Stem Cell Marking With Promotor-deprived Self-inactivating Retroviral Vectors Does Not Lead to Induced Clonal Imbalance

Kerstin Cornils[1,2], Claudia Lange[2], Axel Schambach[3], Martijn H Brugman[3], Regine Nowak[1], Michael Lioznov[2], Christopher Baum[3,4] and Boris Fehse[1]

[1]Experimental Pediatric Oncology and Hematology, Pediatric Clinic III, University Hospital of the Johann Wolfgang Goethe-University, Frankfurt am Main, Germany; [2]Clinic for Stem Cell Transplantation, University Cancer Center, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [3]Department of Experimental Hematology, Hannover Medical School, Hannover, Germany; [4]Division of Experimental Hematology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

Stable genetic modification of stem cells holds great promise for gene therapy and marking, but commonly used γ-retroviral vectors were found to influence growth/survival characteristics of hematopoietic stem cells (HSCs) by insertional mutagenesis. In this article, we show that promoter-deprived γ-retroviral self-inactivating (pd-SIN) vectors allow stable genetic marking of serially reconstituting murine HSC. In contrast to findings with γ-retroviral long terminal repeat (LTR) vectors, serial transplantation of pd-SIN-marked HSC in a sensitive mouse model was apparently not associated with induced clonal imbalance of gene-marked HSC. Furthermore, insertions of pd-SIN into protooncogenes, growth-promoting and signaling genes occurred significantly less frequent than in control experiments with LTR vectors. Also, transcriptional dysregulation of neighboring genes potentially caused by the pd-SIN insertion was rarely seen and comparatively weak. The integration pattern of promotor-deprived SIN vectors in reconstituting HSC seems to depend on the transcriptional activity of the respective gene loci reflecting the picture described for LTR vectors. In conclusion, our data strongly support the use of SIN vectors for gene-marking studies and suggest an increased therapeutic index for vectors lacking enhancers active in HSC.

Received 9 September 2008; accepted 30 September 2008; published online 11 November 2008. doi:10.1038/mt.2008.238

INTRODUCTION

Because of their ability to stably integrate in their host cells' genomes, retroviral vectors have been successfully used for gene marking[1–3] and gene therapy,[4–6] primarily in the hematopoietic system. However, growing evidence from animal[7,8] as well as clinical studies[6,9,10] indicates a significant influence of γ-retroviral vector insertions in the vicinity of growth-regulatory genes (*i.e.*, insertional mutagenesis) on the *in vivo* behavior of the affected clones potentially resulting in induced clonal dominance or leukemia development. The actual prevalence of those side effects may depend on additional factors, *e.g.*, target cell population,[11] transgene[12] or even underlying disease.[13,14] The latter has recently been demonstrated by several groups that performed large-scale analysis of insertion sites in several clinical gene therapy studies.[6,15–17] Obviously, the potential impact of insertional mutagenesis has important implications for (i) the interpretation of marking experiments and (ii) the planning of clinical gene therapy studies which are based on this vector type.

Molecular analysis of retroviral vector–mediated malignant transformation has revealed activation of growth-promoting genes as the main mechanism of insertional mutagenesis.[6–10,12–14,17,18] In fact, the strong enhancer and promotor sequences located in the retroviral long terminal repeats (LTRs) may cause transactivation of neighboring genes resulting in their aberrant expression. In accord, self-inactivating (SIN) vectors lacking the relevant sequences in the U3 region of their LTRs have been shown to be significantly less mutagenic both *in vitro* and *in vivo*.[19,20] However, there is strong evidence that even γ-retroviral SIN vectors containing internal viral promotor sequences are eventually able to transform hematopoietic stem cells (HSCs) by insertional mutagenesis.[19–21]

We asked here whether gene marking of HSC would be feasible with a retroviral vector lacking any promotor/enhancer elements. We therefore constructed a promotor-deprived γ-retroviral SIN (pd-SIN) vector and used this for HSC marking in a serial bone marrow transplantation (BMT) assay, previously shown to represent a sensitive model for detecting insertional clonal imbalance and leukemogenesis.[7,8,22]

RESULTS

Development and testing of pd-SIN

The promotor-deprived vector pd-SIN was generated by excising the internal SF-promotor/enhancer sequences from SINSFeGFP (**Figure 1a**), a SIN γ-retroviral vector described previously.[20,23] To test whether pd-SIN is still able to express enhanced green fluorescent protein (eGFP), SC-1 cells were transduced in parallel with pd-SIN and SINSFeGFP at comparable efficiencies (59% for pd-SIN

Correspondence: *Boris Fehse, Experimental Pediatric Oncology and Hematology, Pediatric Clinic III, University Hospital of the Johann Wolfgang Goethe-University, D-60590 Frankfurt am Main, Germany. E-mail: b.fehse@kinderkrebsstiftung-frankfurt.de*

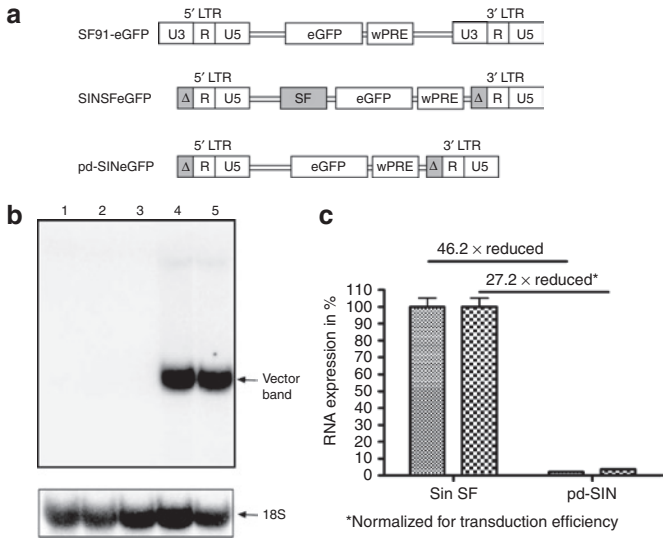

Figure 1 Promoter-deprived γ-retroviral self-inactivating (pd-SIN) vector design and transcriptional activity. (**a**) Construction of the vector SF91-eGFP was described earlier.[22] pd-SIN was derived from SINSFeGFP (refs. 19, 23) by deleting the SF-promotor/enhancer sequences. It thus does not contain any promotor/enhancer element. (**b**) No enhanced green fluorescent protein (eGFP) transcription was detectable in murine SC-1 fibroblasts after transduction with pd-SIN (performed in duplicates) by northern blot analysis using an eGFP probe (lanes 2,3). In contrast, SC-1 cells transduced in parallel with SINSFeGFP (in duplicates) revealed strong eGFP transcription (arrow, lanes 4,5). Lane 1 shows a mock control. RNA load was controlled by hybridization with an 18S probe (lower blot, arrow). (**c**) Since pd-SIN-transduced cells were weakly eGFP-positive in fluorescence-activated cell sorting analysis, quantitative reverse transcription–PCR analyses were performed. As shown, normalized transcription levels were reduced by a factor of >25.

and 64% for SINSFeGFP). Surprisingly, in about one per three pd-SIN transduced cells, weak expression of eGFP was detectable by fluorescence-activated cell sorting (FACS) analysis, which was, however, reduced as compared to SINSFeGFP-transduced cells by a factor of >50, as determined by mean fluorescence intensities (data not shown).

We next measured RNA levels for both pd-SIN and SINSFeGFP. In northern blot analysis no eGFP message was evident in pd-SIN as well as mock-transduced cells, whereas a strong signal was seen in SC-1 cell transduced with SINSFeGFP (**Figure 1b**). However, with a more sensitive quantitative reverse transcription–PCR technique,[19] eGFP transcription was detectable in pd-SIN-transduced cells (**Figure 1c**). The low-level expression of eGFP observed in SC-1 cells, as well as in primary HSC probably indicates the presence of some cryptic promotor activity in the retroviral leader.

Serial transplantation of HSC marked with pd-SIN

Two independent mouse experiments (**Figure 2**) were subsequently performed, each with observation periods of ~1 year. In both experiments, the γ-retroviral pd-SIN vector was used for gene marking in the experimental group and a standard γ-retroviral vector (SF91-eGFP) bearing the strong SFFV U3 promotor/enhancer region in the control group.[24] Earlier studies with the latter and similar LTR vectors have shown a significant mutagenic potential which manifested in induced clonal dominance or even malignant transformation.[6–10,12–15,25]

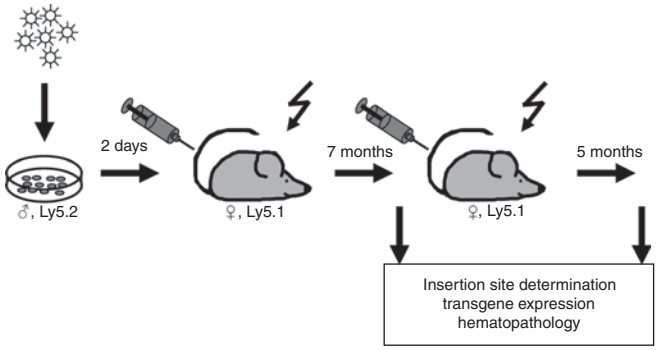

Figure 2 Experimental setup. The design of the mouse study follows the well established, sex-mismatched diallelic (CD45.2/Ly5.2 into CD45.1/Ly5.1) serial bone marrow transplantation model.[7,8,21] This model had enabled us to discover serious side effects of γ-retroviral gene transfer before their occurrence in clinical gene therapy studies, in particular the phenomenon of induced clonal dominance. Shortly, Lin$^-$ bone marrow cells from male donors were transduced with the promoter-deprived γ-retroviral self-inactivating vector and transplanted into female mice after TBI (10 Gy). Seven months later, hematopoietic organs of these mice were thoroughly investigated before secondary transplantation into irradiated mice took place. After a further observation period of 5 months, final analysis of hematopoietic organs, including molecular analysis of vector insertion sites, was performed.

Gene transfer efficiency into HSC for both pd-SIN and SF91-eGFP was in the range of 50% as estimated by flow cytometry and quantitative PCR (data not shown). Transduced HSC were serially transplanted into lethally irradiated mice,[7,8,22] which were followed up for the indicated periods of time (**Figure 2**).

In the pd-SIN group, 14 animals from the first cohorts (Exp. I: 7; Exp. II: 7) and 13 from the second cohorts (Exp. I: 5; Exp. II: 8) survived and were available for detailed investigation. The corresponding numbers in the SF91-eGFP control group were 14 for the first cohorts (Exp. I: 7; Exp. II: 7) and 17 for the second cohorts (Exp. I: 8; Exp. II: 9).

FACS analysis[22] showed hematopoietic donor chimerism in the peripheral blood (**Materials and Methods**) and normal cell counts for all blood lineages in all animals (n = 58); also, final pathological inspection of all 58 available animals did not reveal any abnormalities (data not shown).

Limited numbers of dominant clones in both the pd-SIN and the SF91-eGFP groups

In order to preferentially identify insertion sites in clones dominating hematopoiesis we used a modified ligation-mediated PCR (LM-PCR) protocol.[26] To concentrate on long-term HSC,[27,28] we focused our analysis on clones dominating hematopoiesis after serial BMT (as illustrated for pd-SIN in **Figure 3a**).[25,26] A retroviral vector insertion identified by LM-PCR was considered to represent a dominant clone only if the signal was retrieved in at least two independent LM-PCRs. Using this definition, we found a total of 25 "dominant" pd-SIN insertions in the 13 animals of the two second cohorts available for analysis (Exp. I: 12 insertions, 5 mice; Exp. II: 13 insertions, 8 mice; **Table 1**). For SF91-eGFP, 30 HSC clones were retrieved from 14 secondary transplanted animals (**Table 2**).

These limited numbers confirm the impact of serial transplantation on the establishment of clonal dominance,[2,28] and may

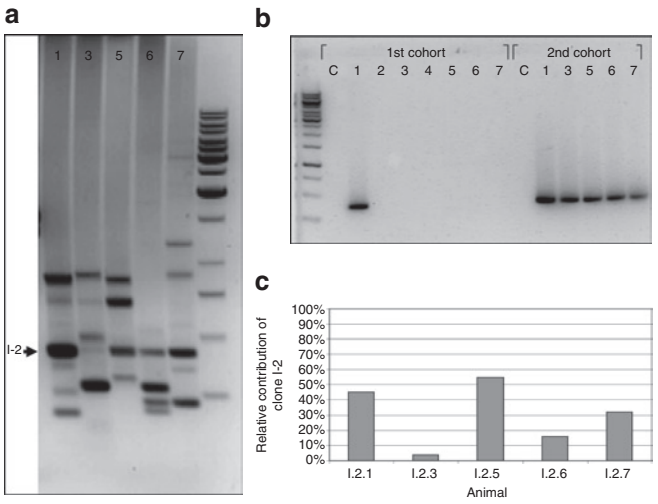

Figure 3 Analysis of promoter-deprived γ-retroviral self-inactivating insertion sites by different PCR techniques. (**a**) A modified ligation-mediated PCR (LM-PCR) method[26] was used to assess genomic locations of integrated vector copies in reconstituting hematopoietic cells (~12 months after initial transduction and 5 months after serial bone marrow transplantation). A representative agarose gel analysis showing the five secondary recipients from Exp. 1 is depicted. Twelve different insertions representing dominant clones could be identified. One insertion (I-2) was dominant in four out of five secondary animals (arrow). (**b**) Insertion site specific (is-PCR) revealed the distribution of hematopoietic clones bearing specific insertion in individual animals of the first and second cohorts of transplanted mice. As a representative example, is-PCR to detect the hematopoietic clone marked by insertion I-2 is shown. As illustrated, this clone took part in the reconstitution of hematopoiesis in all secondary animals. Insertion I-2 had occurred into the *Efcab2* gene not known to be involved in hematopoietic stem cell (HSC) growth regulation. (**c**) Shot-gun cloning of LM-PCR products followed by sequencing of individual clones was performed to assess the relative contribution of specifically marked HSC clones to hematopoiesis in individual animals. Again, insertion I-2 has been selected as a representative example for this type of analysis. Quantitative data are in very good agreement with data from (**a**) LM-PCR and (**b**) is-PCR.

also reflect impaired survival of long-term repopulating HSC under our culture conditions. Additional selection steps seem to further reduce HSC clone number as noted in our previous study with mutagenic γ-retroviral marking LTR vectors, in which only 12 "dominant insertions" were retrieved from 16 secondary recipients.[8]

Using insertion-specific PCR (is-PCR) (**Supplementary Materials and Methods**) we were able to establish comprehensive figures of the distribution of HSC clones in primary and secondary recipients marked by pd-SIN (**Figure 3b**, **Table 1**) or SF91-eGFP (**Table 2**), respectively. For pd-SIN we found just two insertions (II.3 and II.5) that showed a fully identical distribution pattern by both LM-PCR and is-PCR analysis indicating that these two insertions had occurred within one and the same cell, thus representing only one dominant clone. For the control vector at least three clones apparently contained two insertion sites (KI-1 and KI-8, KII-5 and KII-7, KII-9 and KII-10) and one clone was marked by three insertions (KII-11, KII-12 and KII-13). It might be speculated that in HSC clones with multiple vector integrations only one insertion was involved in the establishment of induced clonal dominance (*e.g.*, KII-10: *Foxc2* or KII-11: *Sesn2*), whereas the other insertion(s) may just represent marking events or "bystander insertions." This may lead to an overestimation of relative numbers of insertion sites in genes previously not associated with clonal dominance (gene class 3 in **Figure 4**).

Notably, LM-PCR and is-PCR analysis, *i.e.*, two independent techniques, led to fully identical results during analysis of insertion site patterns (**Tables 1** and **2**). Importantly, results for different insertions within one and the same clone were highly consistent. For instance, if a "dominant" clone contains two vector insertions, one would expect that both insertions will independently be detectable by our LM-PCR technique (focusing on "dominant insertions"). This was indeed the case for all respective clones and in all tested animals (compare, for instance, insertions II-3 and II-5, in animals II.2.0, II.2.1, II.2.2, II.2.6, II.2.7, II.2.8). Accordingly, if a clone contributed in a minor fashion to hematopoiesis, all insertions present in that clone were independently retrieved by the more sensitive is-PCR, but none of them by LM-PCR (compare the same insertions II-3 and II-5 in animals II.1.4 and II.2.4). Together these data confirm the validity of the used definition for clonal dominance.[8,26]

All other insertion sites revealed a distinct distribution pattern between primary and secondary animals which strongly suggests that they represented HSC clones marked by single integrations. Altogether, a total of 24 HSC clones (23 with one, and 1 with two insertions) marked by pd-SIN and 25 clones (21 with one, and 3 with two and 1 with three insertions) marked by SF91-eGFP were analyzed in this study.

Relative contribution of single clones to hematopoiesis

Next we aimed at determining the relative contribution of defined (dominant) clones to hematopoiesis in single animals. To do so, we established a high-throughput sequencing approach for insertion sites retrieved by LM-PCR based on shot-gun cloning of whole LM-PCR products. For each secondary recipient, 96 colonies were sequenced; thus 1,248 (13 × 96) additional sequences were obtained for the 13 secondary recipients in the two pd-SIN experiments. Nine hundred and ninety of these sequences represented pd-SIN insertion sites which were unambiguously locatable in the murine genome; the mean number of identifiable sequences was 76 ± 12 per mouse.

In **Figure 3c**, an example of such data is provided for one insertion (I-2) detected by LM-PCR in animals I.2.1; I.2.5, I.2.6, I.2.7 (compare **Figure 3a**) and additionally by is-PCR in animal I.2.3 (compare **Figure 3b**). Remarkably, the quantitative sequencing data are in excellent agreement with the LM-PCR data (**Figure 3c**). Indeed, only in those cases in which the given clone accounted for >15% of the total detectable transduced clones did it become a "dominant clone" in the LM-PCR analysis. In contrast, in animal I.2.3, where the same clone was only detectable by is-PCR, it constituted <5% of the transduced hematopoietic cells detectable with our assay (**Figure 3a**). Similar results were also obtained for the other insertions (**Table 1**), providing data with regard to the relative contribution of single clones to the hematopoiesis in a given animal.

An analogous analysis was performed for the SF91-eGFP control group. To avoid sequencing of internal vector control

Table 1 Insertion sites of pd-SIN in dominant clones

	Recipients										
		2°BMT									
I[a]	1°BMT[g]	Animal	%[h]	Gene	Chromosome	Band	Distance to TSS[b] (bp)	Orientation[c]	Gene class[d]	RTCGD[e]	IDDb[f]
I-1	I.1.7	**I.2.1**	4	LIM domain only 2 (*Lmo2*)	2	E2	−40,883	R	1	6	1
I-2	I.1.1	**I.2.1**	71	EF-hand calcium binding domain 2 (*Efcab2*)	1	H3	4,251 (Intron [I] 1)	F	3	—	—
		I.2.3	5	Heterogeneous nuclear ribonucleoprotein U (*Hnrpu*)			−72,420	R	3	—	—
		I.2.5	56								
		I.2.6	17								
		I.2.7	29								
I-3	I.1.2	**I.2.1**	2	Adrenergic receptor kinase, β 1 (*Adrbk1*)	19	A	−25,839	R	2	—	—
				F-box snd leucine-rich repeat protein 11 (*Fbxl11*)			−65,283 (I 16)	R	3	—	—
I-4	I.1.6	**I.2.1**	0	Cut-like 1 (*Cutl1, CUX*)	5	G2	139,786 (I 2)	F	3	4	1
				SH2B adaptor protein 2 (*Sh2b2, APS*)			−183,000	F	2	3	—
I-5	I.1.7	**I.2.1**	10	Bri3 binding protein (*Bri3bp*)	5	F	24,105	R	3	—	—
				Acetoacetyl-Coa synthetase (*Aacs*)			−10,200	R	3	—	—
I-6	I.1.5	**I.2.3**	1	Dedicator of cytokinesis 4 (*Dock4*)	12	C1	127,525 (I 1)	F	2	—	—
		I.2.7	1	Zinc finger protein 277 (*Zfp277*)			−127,525	R	3	—	—
I-7	I.1.6	**I.2.3**	0	Notch gene homolog 3 (*Notch3*)	17	B1	62,021	R	2	—	—
				Similar to pyridoxal kinase (*LOC435518*)			−4,545	R	U	—	—
I-8	I.1.5	**I.2.3**	85	Baculoviral IAP repeat–containing 5 (*Birc5*)	11	E2	31,882	F	2	—	—
		I.2.6	60	Threonine aldolase 1 (*Tha1*)			−7,638	R	3	—	—
		I.2.7	0								
I-9	I.1.4	**I.2.5**	3	Potassium channel tetramerization domain containing 12 (*Kctd12*)	14	E2.3	174,000	R	3	—	—
I-10	I.1.2	**I.2.1**	0	Predicted gene EG626897 (*EG626897*)	1	E2.2	96,000	R	U	—	—
		I.2.3	4								
		I.2.5	25								
I-11	I.1.3	**I.2.5**	13	SCY1-like 1 (*Scyl1*)	19	A	−31,435	F	2	4	—
I-12	I.1.5	**I.2.6**	23	ELK3, member of ETS oncogene family (*Elk3*)	10	C2	14,623 (I 4)	F	1	—	—
	I.1.6	I.2.7	67	Predicted gene EG667314 (*EG667314*)			−96,807	F	3	—	—
II-1	II.1.1	**II.2.0**	17	Transglutaminase 2, C polypeptide (*Tgm2*)	2	H2	−10,826	R	3	—	—
	II.1.3	II.2.6	3	RIKEN cDNA D630003M21 gene (*D630003M21Rik*)			72,004	R	U	—	—
II-2	II.1.6	**II.2.0**	55	Predicted gene OTTMUSG00000000971 (*OTTMUSG00000000971*)	11	B5	−1,519	F	U	—	—
		II.2.1	0								
		II.2.4	44	Extracellular proteinase inhibitor (*Expi, Kal1*)			−6,467	F	3	—	—
		II.2.5	49								
		II.2.7	27								
II-3	II.1.4	**II.2.0**	4	Bcl3-binding protein (*B3bp, N4bp2*)	5	D	2,263 (I 1)	R	2	—	—
		II.2.1	26	Mammary gland RCB-0527 JygMC(B) cDNA, RIKEN full-length enriched library, clone:G930037E09 product (*LOC384261*)			2,450 (I 1)	F	U	—	—
		II.2.2	6								
		II.2.4	1								
		II.2.6	17								
		II.2.7	11								
		II.2.8	17								

II-4	II.1.7	**II.2.0**	8	cDNA sequence BC021523 (*BC021523*)	15	E3	4,822 (I 1)	F	U	—	—
		II.2.2	0								
		II.2.6	8								
II-5	II.1.4	**II.2.0**	9	Similar to gem (nuclear organelle) associated protein 8 (*LOC100043887*)	17	E2	94,987	R	U	—	—
		II.2.1	14								
		II.2.2	5								
		II.2.4	1								
		II.2.6	11								
		II.2.7	0								
		II.2.8	4								
II-6	II.1.4	**II.2.1**	6	Baculoviral IAP repeat–containing 5 (*Birc5*)	11	E2	46,310	F	2	—	—
		II.2.2	16	Threonine aldolase 1 (*Tha1*)			−22,066	R	3	—	—
		II.2.4	1								
		II.2.7	13								
		II.2.8	18								
II-7	II.1.1	**II.2.1**	23	Purinergic receptor P2X, ligand-gated ion channel 4 (*P2rx4*)	5	F	24,188	F	2	—	1
		II.2.2	10								
		II.2.7	1	Purinergic receptor P2X, ligand-gated ion channel 7 (*P2rx7*)			87,834	F	2	—	1
		II.2.8	10								
				Calcium/calmodulin-dependent protein kinase kinase 2, β (*Camkk2*)			32,549	R	3	4	—
				F-box and leucine-rich repeat protein 10 (*Fbxl10*)			138,923	R	3	6	—
II-8	II.1.1	**II.2.1**	26	RIKEN cDNA E430004N04 gene (*E430004N04Rik*)	10	A4	−149,045	F	U	—	—
		II.2.2	60								
		II.2.4	0								
		II.2.5	21								
		II.2.7	0								
		II.2.8	48								
II-9	II.1.6	II.2.1	0	Nucleotide-binding oligomerization domain containing 1 (*Nod1*)	6	B3	32,785 (I 4)	F	3	2	—
		II.2.2	0								
		II.2.4	44								
		II.2.5	19								
		II.2.7	37								
		II.2.8	0								
II-10	II.1.3	**II.2.5**	0	RIKEN cDNA C13009O11 gene (*C130092O11Rik*)	6	C1	−18,817	R	U	—	—
II-11	II.1.7	II.2.4	0	Predicted gene, 100041590 (*LOC100041590*)	13	A3–A5	−18,078	R	U	—	—
		II.2.5	11	Phenylalanine-tRNA synthetase 2, mitochondrial (*Fars2*)			84,901 (I 1)	F	3	—	—
II-12	II.1.5	**II.2.6**	5	Predicted gene, EG668164 (*LOC100043729*)	2	F1	3,492 (I 2)	F	U	—	—
II-13	n.d.	**II.2.6**	12	Ena-vasodilator stimulated phosphoprotein (*Evl*)	12	F2	83,600 (I 1)	F	3	—	—

Abbreviation: n.d., not detected.

[a]I: insertion site—insertions identified by ligation-mediated PCR in secondary recipients (bold) are listed with I for the first and II for the second experiment and using arabic numbers; insertions in all primary and some secondary recipients were subsequently detected using insertion-specific PCR (see **Figure 3**). [b]TSS: transcription start site (the mRNA start site according to the National Center for Biotechnology Information database, www.ncbi.nlm.nih.gov). [c]Vector orientations with regard to the respective gene. [d]Genes are classified as 1: (proto-)oncogenes, (2) signaling genes, (3) other, and (U) unknown genes/functions (ref. 25). [e]RTCGD: retrovirus tagged cancer genes database (ref. 29). [f]IDDb: insertional dominance database (ref. 25). [g]1°BMT: recipients analyzed after primary bone marrow transplantation (BMT), individual mice are numbered: I.1.2, I.1.2 and so forth for the first experiment, II.1.0, II.1.1 and so forth for the second experiment, analogously for secondary BMT (2°BMT, *i.e.*, I.2.1 and II.2.0, respectively). [h]Relative contribution of the given clone in the respective animal as estimated based on the presence of plasmid clones bearing the identified sequence after shot-gun cloning.

Table 2 Insertion sites of SF91-eGFP in dominant clones

	Recipients										
	1°BMT[g]	2°BMT					Distance to				
KI[a]	(animal)	Animal	%	Gene	Chromosome	Band	TSS[b] (bp)	Orientation[c]	Gene class[d]	RTCGD[e]	IDDb[f]
KI-1	KI.1.8	**KI.2.1**	0	Genetic suppressor element 1 (*Gse1*)	8	E1	−159,000	F	3	2	—
		KI.2.3	22								
		KI.2.4	1								
		KI.2.8	0								
KI-2	KI.1.4	**KI.2.1**	49	Topoisomerase (DNA) I (*Top1*)	2	H2	55,263 (I 9)	F	3	—	—
	KI.1.6	KI.2.3	0	Phospholipase C, γ 1			−30,150	F	2	—	—
		KI.2.4	0								
		KI.2.8	48								
KI-3	KI.1.7	**KI.2.1**	6	Hypothetical protein LOC100042178 (*LOC100042178*)	4	E2	1,382 (I 2)	F	3	—	—
		KI.2.2	3	PR domain containing 16 (*Prdm16*)			−1,423	R	1	8	—
		KI.2.8	0								
KI-4	KI.1.7	**KI.2.2**	3	LOC100039735 similar to HN1 (*LOC100039735*)	12	C1	17,391	F	3	—	—
	KI.1.8	KI.2.4	4	LOC673582 similar to ribosomal protein L31 (*LOC673582*)			−26,690	R	3	—	—
		KI.2.7	9	Sorting nexin 6 (*Sxn6*)			67,230	R	2	—	—
KI-5	KI.1.8	**KI.2.2**	1	WNK lysine deficient protein kinase 2 (*Wnk2*)	13	A5–B1	−32,648	R	2	—	—
		KI.2.8	0	Ninjurin 1 (*Ninj1*)	13		−6,885	F	3	—	—
				Sushi domain containing 3 (*Susd3*)	13		67,501	R	3	4	1
KI-6	KI.1.3	**KI.2.2**	7	Ferritin heavy chain 1 (*Fth1*)	19	A–C	−6,605	R	2	—	—
KI-7	KI.1.7	**KI.2.2**	34	Interferon dependent positive acting transcription factor 3 γ	14	C3	621 (E 2, CDS)	R	2	—	—
	KI.1.8	KI.2.3	0	(*Isgf3g*, *Irf-9*)							
		KI.2.4	19								
		KI.2.7	3								
		KI.2.8	3								
KI-8	KI.1.8	KI.2.1	1	Lamin A (*Lmna*)	3	E3–F1	−8,305	F	3	—	—
		KI.2.3	16	Sema domain, immunoglobulin domain (Ig), transmembrane domain			−55,933	F	2	—	—
		KI.2.4	5	(TM) and short cytoplasmic domain, (semaphorin) 4A (*Sema4a*)							
		KI.2.8	0	RAB25, member RAS oncogene family (*Rab25*)			36,667	F	1	—	—
KI-9	KI.1.8	KI.2.1	0	Homeobox A7 (*Hoxa7*)	6	B3	−87	R	1	25	2
		KI.2.3	17	Homeobox A6 (*Hoxa6*)			−10,036	R	1	—	—
		KI.2.4	1	Homeobox A9 (*Hoxa9*)			8,710	R	1	25	—
		KI.2.6	0								
		KI.2.8	0								
KI-10	KI.1.8	**KI.2.4**	1	Peptidoglycan recognition protein 2 (*Pglyrp2*)	17	B1–B2	1,260 (I 1)	F	3	—	—
				A kinase (PRKA) anchor protein 8 (*Akap8*)			−101,754	F	2	—	—
KI-11	KI.1.6	**KI.2.4**	4	Cyclin-dependent kinase 6 (*Cdk6*)	5	A2–A3	89,048 (I 3)	R	1	2	1
KI-12	KI.1.7	KI.2.1	0	Similar to ribosomal protein S18 (*LOC666782*)	4	A1	−33,646	R	3	—	—
		KI.2.3	0								
		KI.2.4	0								
		KI.2.5	25								
		KI.2.8	0								

KI-13	KI.1.7	KI.2.1	n.d.	Septin 9 (*Seqt9*)	11	E2	−22,645	F	2	7	—
		KI.2.2									
		KI.2.3									
		KI.2.4									
		KI.2.6									
		KI.2.8									
KI-14	KI.1.6	KI.2.1	n.d.	Cyclin-dependent kinase (CDC2-like) 10 (*Cdk10*)	8	E1	−60,126	F	2	—	—
		KI.2.2		Sulfotransferase family 5A, member 1 (*Sult5a1*)			−6,447	R	3	—	—
		KI.2.3									
		KI.2.6									
KI-15	KI.1.8	KI.2.2	0	Transmembrane protein 63b (*Tmem63b*)	17	B3	233 (I 1)	R	U	—	—
		KI.2.4	5	Mitochondrial ribosomal protein L14 (*Mrpl14*)			−467	F	3	—	—
		KI.2.7	35								
KI-16	KI.1.3	**KI.2.7**	3	G protein-coupled receptor 81 (*Gpr81*)	5	F	−17,546	F	2	—	—
	KI.1.6			Density-regulated protein (translational initiation) (*Denr*)			−9,726	R	3	—	—
KII-1	KII.1.4	**KII.2.0**	9	B-cell CLL/lymphoma 7B (*Bcl7b*)	5	G1	−142,862	R	2	—	—
		KII.2.2	11	Claudin 3 (*Cldn3*)			39,416	R	2	—	—
		KII.2.3	0	Syntaxin 1A (brain) (*Stx1a*)			2,058 (I 1)	R	3	—	—
		KII.2.6	0								
KII-2	KII.1.0	**KII.2.0**	11	Myocyte enhancer factor 2C (*Mef2c*, *Evi121*)	13	C3	69,625 (I 2)	F	1	11	
KII-3	KII.1.4	**KII.2.0**	6	Glucosamine-6-phosphate deaminase 1 (*Gnpda1*)	18	B3	−10,415	R	3	—	—
		KII.2.2		Nedd4 family interacting protein 1 (*Ndfip1*)			−69,609	F	2	—	—
KII-4	n.d.	**KII.2.0**	3	Similar to Tubb2a protein (*LOC666700*)	4	A1	−116,869	R	3	—	—
KII-5	KII.1.6	KII.2.0	0	MAP kinase-interacting serine/threonine kinase 2 (*Mknk2*)	10	C1	−10,356	R	2	—	—
		KII.2.1	3	MOB1, Mps One Binder kinase activator-like 2A (yeast) (*Mobkl2a*)			19,488	R	3	—	—
		KII.2.2	11								
		KII.2.3	3								
		KII.2.4	12								
		KII.2.6	9								
		KII.2.7	1								
		KII.2.8									
KII-6	KII.1.5	**KII.2.1**	13	Sortilin-related receptor, LDLR class A repeats containing (*Sorl1*)	9	B	212,000	R	3	—	—
				RIKEN cDNA 6720432D03 gene (*6720432D03Rik*)			−21,653	R	3	—	—
KII-7	n.d.	KII.2.0	7	HECT domain containing 2 (*Hectd2*)	19	C3	−19,599	R	3	—	—
		KII.2.1	15								
		KII.2.2	16								
		KII.2.3	21								
		KII.2.4	21								
		KII.2.6	14								
		KII.2.7	8								
		KII.2.8									
KII-8	KII.1.6	KII.2.0	28	Docking protein 3 (*Dok3*)	13	B2	163 (I 1)	R	2	—	—
		KII.2.1	23								
		KII.2.2	16								
		KII.2.3	16								
		KII.2.4	26								
		KII.2.6	24								
		KII.2.7									
		KII.2.8									

Table 2 Continued on next page

Table 2 Insertion sites of SF91-eGFP in dominant clones (continued)

	Recipients										
	1°BMT[g]	2°BMT					Distance to				
I[a]	(animal)	Animal	%[h]	Gene	Chromosome	Band	TSS[b] (bp)	Orientation[c]	Gene class[d]	RTCGD[e]	IDDb[f]
KII-9	KII.1.6	KII.2.0	5	Zinc finger protein 750 (*Zfp750*)	11	E2	−39,974	R	U	—	—
		KII.2.2	15	Tubulin-specific chaperone d (*Tbcd*)			107,274 (I 16)	F	U	—	—
		KII.2.3	12								
		KII.2.4	7								
		KII.2.6	4								
		KII.2.7	5								
		KII.2.8									
KII-10	KII.1.6	KII.2.0	3	Forkhead box C2 (*Foxc2*)	8	E1	−119,157	R	2	3	—
		KII.2.2	11	Forkhead box L1 (*Foxl1*)			−130,259	R	2	2	—
		KII.2.3	13								
		KII.2.4	15								
		KII.2.6	23								
		KII.2.7	14								
		KII.2.8									
KII-11	KII.1.6	KII.2.0	8	Regulator of chromosome condensation 1 (*Rcc1*)	4	D2.3	−8,282	R	2	—	—
		KII.2.2	8	Sestrin 2 (*Sesn2*)			156,000	R	2	—	3
		KII.2.3	8	Small nucleolar RNA host gene (nonprotein coding) 3 (*Snhg3*)			−311	R	3	—	—
		KII.2.4	12								
		KII.2.6	9								
		KII.2.7									
		KII.2.8									
KII-12	KII.1.6	KII.2.0	11	Aspartate-β-hydroxylase (*Asph*)	4	A1	+247,000	R	3	—	—
		KII.2.2	16								
		KII.2.3	6								
		KII.2.4	11								
		KII.2.6	3								
		KII.2.7									
		KII.2.8									
KII-13	KII.1.6	KII.2.0	n.d.	Cux1/Cutl1 cut-like homeobox 1 (*Cutl1*, *Cux1*, *Evi71*)	5	G2	157,000	R	2	4	—
		KII.2.2									
		KII.2.3									
		KII.2.4									
		KII.2.8									
KII-14	n.d.	**KII.2.5**	5	Prominin 1 (*Prom1*, *CD133*)	5	B3	−32,625	R	3	—	—
				Transmembrane anterior posterior transformation 1 (*Tapt1*)			92,300	R	3	—	—

Abbreviation: n.d., not detected.

[a]I: insertion site—insertions identified by ligation-mediated PCR in secondary recipients (bold) are listed with KI for the first and KII for the second experiment and using arabic numbers. [b]TSS: transcription start site (the mRNA start site according to the National Center for Biotechnology Information database, ref. 19). [c]Vector orientation with regard to the respective gene. [d]Genes are classified as 1: (proto-)oncogenes, (2) signaling genes, (3) other and (U) unknown genes (ref. 25). [e]RTCGD: retrovirus tagged cancer genes database (ref. 29). [f]IDDb: insertional dominance database (ref. 25). [g]21°BMT: recipients analyzed after primary BMT, individual mice are numbered: KI.1.2, KI.1.2 and so forth for the first experiment, KII.1.0, KII.1.1 and so forth for the second experiment, analogously for secondary BMT (2°BMT, *i.e.*, KI.2.1 and KII.2.0, respectively).

bands, those were excluded based on a newly developed PCR approach. A total of 1,632 sequences were investigated for 17 secondary BMT recipients, of which 1,229 represented SF91-eGFP insertions (72 ± 7 per mouse). The obtained clonal quantification data are presented in **Table 2**.

It is important to note, however, that the above quantification procedure has certain biases. Both the LM-PCR (influenced, for example, by the length of the amplified fragment) and the cloning (not all pieces of eukaryotic genomes are equally "clonable" in bacteria) steps are not reliably quantitative methods. This may particularly apply, if several clones are present in a given animal resulting in a competitive situation. Those restrictions may explain our observation that not all insertions that represented dominant clones in the gel analysis quantitatively contributed to shot-gun cloning (**Table 1**). This finding underlines the importance of directly cutting out and sequencing dominant bands from the agarose gel after LM-PCR in order to avoid loss of insertion sites during the cloning step.

Genomic locations of insertion sites

An important question was whether the genomic location pattern of pd-SIN insertion sites qualitatively differs from that found with LTR vectors in clones dominating hematopoiesis after serial BMT. To answer this question all identified insertions were blasted against the National Center for Biotechnology Information mouse build 37 genome database (www.ncbi.nlm.nih.gov); genes in a 150-kb window around the insertion were included into analysis (**Tables 1** and **2**).

Six of the 25 pd-SIN vector insertions found in reconstituting HSC were located in intergenic regions with the next (predicted) gene ~100 kb or further away. In the remaining 19 dominant clones, only two insertions were located close to or within established protooncogenes (*Lmo2, Elk3*), eight nearby signaling genes (including the putative growth–promoting gene *Notch3* and the apoptosis inhibitor *Birc5* which was found to be independently hit in both mouse experiments). Fifteen integrations had occurred in the proximity of "other," unknown or predicted genes (**Table 1**; **Figure 4**). In line with that, only a minority of four clones (17%) was found to be dominant in more than half of the secondary recipients (Exp. I: I-2; Exp. II: II-3, II-5, II-8). Two of the latter (II-5, II-8) bore insertions which were located far away from the next gene (95 and 149 kb, respectively). The other two were located in the proximity of *Efcab2* (5 kb downstream, 1st intron; I-2) and *B3bp* (2 kb downstream, 1st intron; II-3), two genes not known to be involved in growth regulation in HSC or other cell types. These data strongly argue against a role of insertional mutagenesis during establishment of clonal dominance of pd-SIN-marked clones.

In contrast, the distribution of insertion sites of the SF91-eGFP control vector well resembled earlier results from our and many other laboratories reported for this vector type[25] (**Figure 4**). In fact, as in the IDDb (ref. 25), >70% of all insertions were found in signaling and/or protooncogenes. Five different protooncogenes (**Table 2**) were identified as insertion sites in dominant clones, four of them previously shown to be involved in hematopoietic malignancies. Most remarkably, one of them, *Prdm16*, is not only represented in the Retrovirus Tagged Cancer Gene Database

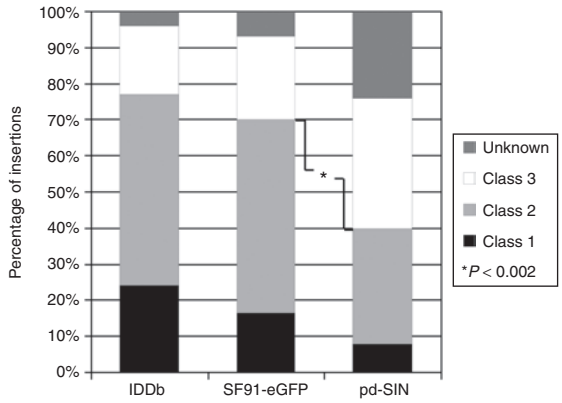

Figure 4 Specific distribution of promoter-deprived γ-retroviral self-inactivating (pd-SIN) insertion sites in reconstituting hematopoietic stem cell. Insertion sites were categorized based on the genes present in a 150-kb window at both the 5′ and the 3′ sites of each individual insertion. Genes were classified into (putative) protooncogenes (class 1), signaling genes (class 2), other genes (class 3), and genes of unknown function as previously described.[25] For control, respective data obtained in the two parallel control experiments with the γ-retroviral long terminal repeat (LTR) vector SF91-eGFP are shown. As evident, pd-SIN insertions show a strongly different insertion pattern as compared to γ-retroviral LTR vectors. Based on the frequency of hits into putative growth–promoting genes (classes 1 and 2), this difference is highly significant ($P < 0.002$; Fisher's exact test). In contrast, no significant differences were found between insertion patterns for SF91-eGFP (control group) and historical data from the IDDb.[25]

(RTCGD),[29] but its human counterpart has also been associated with induced clonal dominance in a clinical gene therapy study.[6] In line, three other protooncogenes (*HoxA7*, *Cdk6*, *Mef2c/Evi121*) were identified before by the Copeland group as common retroviral insertion sites in retrovirus-induced malignancies.[29] Moreover, *HoxA7* and *Cdk6* had already been linked to induced clonal dominance in the IDDb.[25]

Besides those protooncogenes, several other genes hit by SF91-eGFP were previously suggested to be involved in malignant transformation and/or induced clonal dominance (**Table 2**). These include the *FoxL1* and *FoxF1a* genes (one genomic locus) listed in the RTCGD[29] and IDDb,[25] respectively. Another common retroviral insertion sites in both RTCGD and IDDb is *Susd3*, which is particularly interesting because its function is so far largely unknown. Two other genes hit by SF91-eGFP in dominant clones could be found in the RTCGD: *Cux1/Evi71* (also an insertion site of pd-SIN, **Table 1**) and *Sept9*.

In summary, the overrepresentation of SF91-eGFP integrations in protooncogenes already shown to be able to induce benign and/or malignant clonal outgrowth after retroviral insertion–mediated upregulation (*Prdm16*, *Hoxa7*, *Cdk6*, *Mef2c/Evi121*) strongly argues for a potential role of insertional mutagenesis in the affected clones. The actual role of these, and also the other named candidate genes, in the establishment of induced clonal dominance is currently under investigation.

In conclusion, the insertion pattern observed for pd-SIN was strikingly different from that seen in our previous studies and with the control vector SF91-eGFP (**Figure 4**).[8] In fact, based on Fisher's exact test, integrations of pd-SIN into oncogenes, growth-promoting and signaling genes (gene groups 1 and 2) occurred

significantly ($P < 0.002$) less frequent than with the SF91-eGFP control vector.

No strong transcriptional dysregulation of genes at pd-SIN insertion sites

Although pd-SIN would not be expected to transactivate neighboring genes, its integration may still have resulted in transcriptional dysregulation which principally might have facilitated later dominance of an affected clone. The latter might particularly be expected with the insertions in the proximity of the protooncogenes (*e.g.*, *Lmo2*, *Elk3*) or for those clones which in a dominant way contributed to hematopoiesis in more than half of the secondary recipients. As stated above, only for two (out of four) of those clones genes were detectable in the proximity of the vector integration (*Efcab2*, *B3bp*).

To address this question we measured transcription levels for 32 adjacent genes in splenocytes of affected animals using quantitative reverse transcription–PCR (**Figure 5**). Changes in transcription levels compared to control splenocytes for most genes analyzed (19/32, 59%) were below a factor of 2. For almost 90% (28/32), the changes were <3, a threshold often used for defining significant transcriptional upregulation in array analysis (**Figure 5**). In particular, transcriptional upregulation of *Lmo2* was hardly detectable. In line, integration into the *Lmo2* gene had occurred far away from the described negative regulatory element (data not shown).[30]

Three of the genes with higher upregulation (*Aacs*: factor 3.0 ± 0.9; *EG667228*: 3.1 ± 2.0; *P2xr4*: 6.0 ± 1.2) have no described function in cell survival or hematopoiesis. Also, we found no large changes in the mean transcription of *Efcab2* (1.9 ± 0.8) and *B3bp* (1.4 ± 0.1) in the secondary animals reconstituted by those two clones comprising a high reconstitution potential (see above) (**Figure 5**).

A comparatively strong transcriptional upregulation was found for *Elk3*, although the big SDs indicate large inter-animal variations in the transcription of this gene (**Figure 5**). *Elk3* is a member of the *Ets* family of transcription factors. Because a potential role of *Elk3*, which represents an effector of mitogen-activated protein kinase pathways, in growth regulation of HSC could not be excluded, its upregulation and a potential contribution to clonal survival need further investigation. However, after serial BMT only one animal was reconstituted by a dominant *Elk3* clone (**Table 1**) arguing against strong induced clonal dominance.

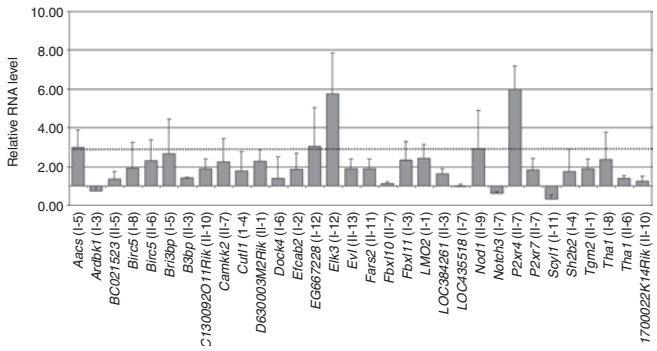

Figure 5 Minor transcriptional dysregulation in gene loci targeted by promoter-deprived γ-retroviral self-inactivating (pd-SIN). Real-time reverse transcription–PCR using QuantiTect Primer Assays were used to assess transcription levels of genes adjacent to pd-SIN vector insertions. RNA levels as compared to healthy control mice splenocytes were established using the $\Delta\Delta C_t$ method.[8,49] Only comparatively slight changes in transcription levels were detected. In particular, genes adjacent to insertion sites found to mark clones dominant in more than half of the secondary recipients did not show relevant dysregulation of transcription (I-2: *Efcab2*; II-3: *B3bp*). Relative error bars indicate SDs from at least three independent assays.

Integrations preferentially occur into active genomic loci

The above data indicate that pd-SIN marking of HSC is not associated with induced clonal dominance. Still, the distribution seems to be different from what would be expected in a pure chance setting. In fact, among the 25 vector insertions we found two insertions into protooncogenes (~8%) (**Figure 4**). On the other hand,

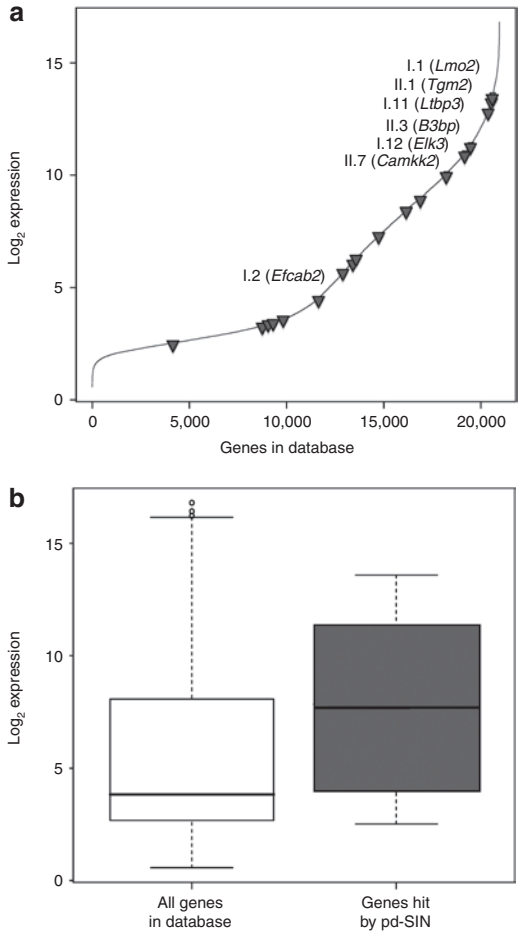

Figure 6 Promoter-deprived γ-retroviral self-inactivating (pd-SIN) preferentially integrates into transcriptionally active regions. (**a**) Insertion sites of pd-SIN were correlated with transcription levels of the targeted gene loci. To do so we made use of an expression database for murine hematopoietic stem cells established by Goodell and colleagues.[33] All genes contained in this database were plotted against their respective expression values. The shown line consists of >20,000 individual data points representing the highest expression values for each gene on the array. As shown, most genes hit by pd-SIN (gray triangles) were transcriptionally highly active in hematopoietic stem cell (HSC). (**b**) The relative transcriptional activity of genes targeted by pd-SIN was compared with that of all genes contained in the HSC expression database.[29] Clearly, pd-SIN preferred actively transcribed gene loci for integration.

such preference is well in line with the reported integration bias of γ-retroviral vectors in mixed cultures of HSC and progenitor cells.[31]

It has been suggested that this integration bias may be based on the expression levels of the targeted gene loci in HSC.[31,32] To proof this possibility, we made use of an expression database for murine HSCs established by the Goodell lab.[33] On the basis of that database, we were able to determine expression levels for 18 genes adjacent to the integration sites. Genes located 5′ from the insertion site were preferentially used since many insertions were found in introns thus obviously depending on the activity of the promoter located 5′. However, if there was no gene detectable in the 5′ proximity of an insertion, we used the next 3′ gene for analysis. In some cases, closely located genes were identified on both sites of the vector integration. It is remarkable that $\log_2$-transformed expression values for those genes were nearly identical, independent of the expression strength of the given locus (insertion I-11: *Malat1* – 13.59 versus *Ltbp3* – 12.82; II-2: *Expi* – 3.58 versus *Ccl4* – 3.41). These data underline the validity of our approach.

In order to assess transcriptional activity of the targeted gene loci in HSC we plotted the expression values of the hit genes against all genes in the Goodell database (**Figure 6a**). As evident, most targeted gene loci show intermediate-to-high-expression levels, whereas only one (II-13, *Degs2*) of the 18 analyzed genes (5.6%) is located in the lowest third. This is also reflected in the overall comparison depicted in **Figure 6b** clearly indicating comparatively higher expression in the targeted loci.

Remarkably, the two protooncogenes identified as insertion sites (*Lmo2*, *Elk3*) are amongst those gene loci showing the highest expression level. The same is true for *B3bp*, an insertion site found in one of the clones (II-3) dominant in more than half of the secondary HSCT recipients. Another dominantly repopulating clone (I-2) had an insertion in *Efcab2*, a gene which is characterized by intermediate expression in HSC.

DISCUSSION

Different animal models and also several clinical trials have clearly shown the great potential of γ-retroviral vectors for gene marking and gene therapy.[1–6,34–37] Gene marking has allowed investigation of the reconstitution dynamics of the hematopoietic system after HSCT,[1,2] the impact of *ex vivo* culture and conditioning regimens on engraftment[34,35] and the role of contaminating malignant cells for relapse development after autologous HSCT in clinical settings.[3] At the same time, the successful treatment of inherited monogenic diseases of the hematopoietic system, such as severe combined immunodeficiencies (SCID-X1, ADA-SCID)[4,5,36] or chronic granulomatous disease,[6] using γ-retroviral vector–mediated gene correction has provided the long-awaited proof of principle for genetic therapy.

Unfortunately, further progress has been thwarted by the occurrence of severe side effects including induced dominance as well as malignant transformation of affected clones in both animal models[7,8,37] and clinical studies.[6,9,10] Thorough molecular analysis allowed attributing those side effects to insertional mutagenesis.[6–10,12–14,37] Although insertional mutagenesis had been reported with replication-competent retroviruses decades ago,[38] it was thought to be quite unlikely with nonreplicating vectors given the low probability of hitting a protooncogene in a susceptible (stem) cell. In accord, no evidence for insertional oncogenesis after retroviral vector–mediated gene transfer had been found in numerous animal studies.[12–14,39]

In most instances, insertional mutagenesis was shown to be conferred by transactivation of neighboring genes from the strong promotor/enhancer elements located in the U3 region of retroviral LTRs.[6–14] However, even SIN vectors may still be mutagenic, given that they contain a sufficiently strong internal promotor. In fact, γ-retroviral SIN vectors containing internal retroviral promotors were still able to facilitate *in vitro* immortalization[19] as well as leukemia development[21] in suitable models, albeit at decreased incidences. In addition, lentiviral vectors equipped with the strong SFFV-LTR revealed genotoxicity in a tumor-prone mouse model.[40] Moreover, even SIN lentivectors harboring physiological promotor/enhancer sequences derived from the β-globin locus control region were shown to cause insertional dysregulation of cellular genes in erythroid cells over long distances and at high frequencies.[41]

Given those results, we explored the usefulness of an almost totally "neutral" gene-marking vector lacking all promotor/enhancer elements. This pd-SIN vector would not be expected to directly transactivate neighboring genes. Using pd-SIN vectors to mark HSC in our sensitive serial BMT model,[7,8] we were able to address a number of questions. First, we investigated whether pd-SIN vectors are still associated with induced clonal dominance. The significantly lower incidence (as compared to LTR vectors) of pd-SIN insertions into putative growth–promoting genes and the lack of evidence for very strong transcriptional upregulation of adjacent genes (in contrast to that observed with LTR vectors[8]) argue against this possibility. Indeed, most observed transcriptional changes were below factor 2. Moreover, given the high SDs, most differences seem to rather represent natural interindividual variations than induced upregulation or downregulation (compare SDs, **Figure 5**). However, the varying contribution of single clones to hematopoiesis of individual mice complicates reliable quantitative statements.

It should be taken into account that even a promotor-deprived vector is potentially genotoxic, because its integration may disrupt or truncate open-reading frames or gene-regulatory regions. Intriguingly, for the insertion into the *Elk3* protooncogene we observed a sixfold upregulation (though with large inter-animal deviations). Although the affected clone reconstituted only one animal in a dominant fashion, we cannot fully rule out the involvement of insertional mutagenesis in its fate.

Still it is safe to conclude that the picture obtained with pd-SIN reflects mainly an unbiased insertion profile of γ-retroviral vectors in long-term reconstituting HSC. Indeed, our data relating the identified vector integration sites with the expression characteristics of the targeted loci (based on the expression profile of the best-characterized murine HSC population as established by the Goodell group[33]) suggest that integration site selection largely depended on the activity of a given gene locus. These data confirm the preference of γ-retroviral vectors for actively transcribed genomic regions[42,43] and are well in line with earlier *in vitro* results for cultured progeny of human $CD34^+$ cells.[31] Other data[44,45] suggest that the apparent preference of retroviral vectors for certain genomic loci ("hot spots") may be caused by specific interactions of retroviral preintegration complexes with certain transcription factor binding sites. Interestingly, Recchia

et al.[45] have provided evidence that, in addition to the retroviral integrase, direct interactions of transcription factors with LTR enhancer sequences play a pivotal role in directing proviral integrations toward transcription factor binding site–rich genomic regions. To investigate this further, a large-scale insertion site database for the pd-SIN vector lacking any enhancers should be established (using novel sequencing techniques[46]) which will allow comparing its insertion site preferences with the data obtained for LTR vectors.[15–17,31,43–45]

The above findings lead to the assumption that the clonal dominance seen in our secondary transplanted animals represents a direct result of serial BMT and thus reflects intrinsic properties of serially transplanted HSC. This observation indicates that most conclusions of earlier marking studies with γ-retroviral LTR vectors[1,2] still hold true. At the same time it underlines that the concept of using promotor-deprived SIN vectors for gene marking may overcome limitations of γ-retroviral LTR vectors, namely their strong mutagenicity leading to a rare, but unfortunately overriding phenotype of induced clonal dominance.[6,8,25]

In summary, to investigate gene marking with the virtually neutral pd-SIN, we have performed two independent long-term animal experiments with very similar results. Based thereon it seems safe to draw the following conclusions: (i) γ-retroviral, pd-SIN vectors ensure stable genetic marking of reconstituting HSC; (ii) the insertion pattern of pd-SIN vectors is significantly different from that of γ-retroviral LTR vectors, but shows a similar dependence on transcriptional activity of targeted loci. This might explain the slight preference for growth-regulatory genes;[31] (iii) insertion of pd-SIN vectors has primarily a neutral affect on adjacent genes, *i.e.*, major transcriptional dysregulation of adjacent genes seems to be a rare event; (iv) in accord, gene marking with γ-retroviral pd-SIN apparently usually does not result in induced clonal dominance.

Our observations may be useful for the design of future gene marking[1,2] and competitive engraftment[34,35] studies. Also, the use of vectors that do not influence clonal outgrowth will permit the analysis of the insertion bias of different vector types in various types of target cells. Finally, our results underline the great promise of using improved SIN vectors for gene therapy. Indeed, as seen here, the mutagenicity of SIN vectors containing tissue-specific promotors that are silent in stem cells significantly lower the likelihood of severe side effects of stem cell gene therapy.

MATERIALS AND METHODS

Retroviral vector production. In the animal experiments we used the retroviral vectors pd-SIN (see below), and SF91-eGFP/wPRE ("SF91-eGFP") containing the SFFV U3 promotor/enhancer region and a wPRE element.[24] Cell-free supernatants containing ecotropic viral particles were generated by transient transfection of PhoenixGP cells (kindly provided by G Nolan) as described.[47,48] Titers (as established on NIH3T3 cells) for pd-SIN were ~3×10^5 per ml (Exp. I) and 5×10^6 per ml (Exp. II). For SF91-eGFP titers were ~1×10^6 per ml in Exp. I and 4×10^6 per ml in Exp. II.

Serial BMT model. Two independent serial HSC transplantation experiments were set up according to our established C57Bl/6J mouse model (**Figure 2**).[7,8,22] In brief, murine BM isolated from young adult male donors was lineage-depleted and taken in culture for *ex vivo* transduction.[48] Lin^- BM cells were transduced with vector-containing supernatant at a multiplicity of infection of three to obtain gene transfer of ~50–60%.[48] Transduced cells (1×10^6 per recipient) were transplanted into female recipient mice that had received total body irradiation (9.5 Gy) (Exp. I: $n = 8$; Exp. II: $n = 10$). FACS analysis[22] showed donor-derived hematopoiesis in the peripheral blood of all animals (pd-SIN – Exp. I: mean 87%, 78–93%; Exp. II: mean 31%, 9–69%; SF91-eGFP – Exp. I: mean 80%, 64–93%; Exp. II: mean 32%, 13–57%). In the eight transplanted groups (two experiments each including serial BMT, two vectors) a total of 13 mice were lost because of graft failure or unknown reasons. This is in line with our previous experience in the given model.

After 7 months, transplanted mice were killed. At autopsy gross evaluation of hematopoietic organs (spleen, thymus, lymph nodes) was performed before they were conserved. Cells from BM, PB, and thymus were submitted to detailed FACS analysis of different hematopoietic lineages. Splenocytes were used to isolate DNA and RNA for molecular analysis. Blood smears were investigated for presence of leukemic blasts.

Harvested BM was pooled and transplanted into a second cohort of female recipient mice (Exp. I: $n = 8$; Exp. II: $n = 10$) after total body irradiation (9.5 Gy) at 1×10^6 per mouse. Remaining BM cells, as well as cells isolated from other hematopoietic organs, were conserved for and/or directly applied to FACS and molecular analysis.

Mice from the second cohort were observed for at least 5 months before final analysis (as above) took place.

Molecular analysis of insertion sites in dominant clones. For PCR assays, DNA and RNA were isolated from hematopoietic organs as described.[8] For northern blot analysis, RNA was isolated from SC-1 cultures by standard procedures.[23] Northern blots were performed using a ^{32}P-labeled eGFP probe.[23]

To identify insertion sites in dominant clones we made use of a modified LM-PCR protocol.[26] Once "dominant insertion sites" were identified (see below), primers were designed for is-PCR (see **Supplementary Materials and Methods** and **Supplementary Table S1**).

Transcription levels of genes adjacent to pd-SIN vector insertions were comparatively analyzed for splenocytes from mice bearing those insertions (**Table 1**) versus control splenocytes (cells from nontransplanted mice and/or from secondary recipients that did not bear the clone of interest) using QuantiTect Primer Assays (Qiagen, Hilden, Germany). Differences were determined based on the $\Delta\Delta C_t$ method.[8,49]

To quantitatively determine the contribution of dominant and nondominant hematopoietic clones, whole LM-PCR products were shot-gun cloned into pCRII-TOPO (Invitrogen, Karlsruhe, Germany) according to the manufacturer's protocol. For SF91-eGFP, bacterial colonies were first screened using a vector-specific PCR (primers in **Supplementary Table S2**) to exclude LM-PCR products representing the internal vector control from further analysis (there is no internal PCR signal with the LM-PCR primers used for pd-SIN).[19] Ninety-six cloned insertion sites per individual animal were sequenced using the terminator dye approach (GATC, Heidelberg, Germany).

Expression levels of targeted gene loci in HSC. In order to assess expression levels of targeted gene loci in HSC we made use of a database established by Goodell and colleagues based on microarray technology.[33] According to that work, murine HSC are defined as side population cells which are $Sca\text{-}1^+$, $c\text{-}Kit^+$, and Lin^-.[33] To evaluate the expression level of those loci where vector integrations had happened, we first identified the closest 5′ located gene for each single insertion site. If no 5′ gene was found to be present in the Goodell database, the closest 3′ gene was used instead. For all selected genes, $\log_2$-transformed expression values (representing the mean for two different HSC preparations) were retrieved from the database.[33] If the database contained more than one entry, the highest value was included into further analyses. Those expression values were set into relation to the expression values of all other genes (again only the highest values were taken for each individual gene). Analysis and plotting were performed using R-2.7.0.[50]

ACKNOWLEDGMENTS

We thank Gökhan Arman-Kalcek, and Melanie Engel for their expert technical assistance. We are grateful to Carol Stocking for critical reading of the manuscript. This work has been supported by the Deutsche Forschungsgemeinschaft DFG [FE568/9-1 (SPP1230) and BA1837/7-1 (excellence cluster REBIRTH)], the Roggenbuck-Stiftung and the Frankfurter Stiftung für Krebskranke Kinder. B.F.'s position is funded by the Deutsche Krebshilfe, A.S.'s by the Else-Kröner-Stiftung. This work is part of K.C.'s PhD thesis. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIAL

Table S1. Primer Sequences used for insertion-specific (is-)PCR.
Table S2. Primer Sequences for vector-specific PCR.
Materials and Methods.

REFERENCES

1. Dick, JE, Magli, MC, Huszar, D, Phillips, RA and Bernstein, A (1985). Introduction of a selectable gene into primitive stem cells capable of long-term reconstitution of the hemopoietic system of W/Wv mice. *Cell* **42**: 71–79.
2. Lemischka, IR, Raulet, DH and Mulligan, RC (1986). Developmental potential and dynamic behavior of hematopoietic stem cells. *Cell* **45**: 917–927.
3. Tey, SK and Brenner, MK (2007). The continuing contribution of gene marking to cell and gene therapy. *Mol Ther* **15**: 666–676.
4. Hacein-Bey-Abina, S, Le Deist, F, Carlier, F, Bouneaud, C, Hue, C, De Villartay, JP *et al.* (2002). Sustained correction of X-linked severe combined immunodeficiency by *ex vivo* gene therapy. *N Engl J Med* **346**: 1185–1193.
5. Aiuti, A, Slavin, S, Aker, M, Ficara, F, Deola, S, Mortellaro, A *et al.* Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning (2002). *Science* **296**: 2410–2413.
6. Ott, MG, Schmidt, M, Schwarzwaelder, K, Stein, S, Siler, U, Koehl, U *et al.* (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* **12**: 401–409.
7. Li, Z, Düllmann, J, Schiedlmeier, B, Schmidt, M, von Kalle, C, Meyer, J *et al.* (2002). Murine leukemia induced by retroviral gene marking. *Science* **296**: 497.
8. Kustikova, O, Fehse, B, Modlich, U, Yang, M, Düllmann, J, Kamino, K *et al.* (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* **308**: 1171–1174.
9. Hacein-Bey-Abina, S, von Kalle, C, Schmidt, M, McCormack, MP, Wulffraat, N, Leboulch, P *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* **302**: 415–419.
10. Howe, SJ, Mansour, MR, Schwarzwaelder, K, Bartholomae, C, Hubank, M, Kempski, H *et al.* (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* **118**: 3143–3150.
11. Newrzela, S, Cornils, K, Li, Z, Baum, C, Brugman, MH, Hartmann, M *et al.* (2008). Resistance of mature T cells to oncogene transformation. *Blood* **112**: 2278–2286.
12. Baum, C, Düllmann, J, Li, Z, Fehse, B, Meyer, J, Williams, DA *et al.* (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood* **101**: 2099–2114.
13. Baum, C, Kustikova, O, Modlich, U, Li, Z and Fehse, B (2006). Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Hum Gene Ther* **17**: 253–263.
14. Baum, C, von Kalle, C, Staal, FJ, Li, Z, Fehse, B, Schmidt, M *et al.* (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther* **9**: 5–13.
15. Deichmann, A, Hacein-Bey-Abina, S, Schmidt, M, Garrigue, A, Brugman, MH, Hu, J *et al.* (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J Clin Invest* **117**: 2225–2232.
16. Aiuti, A, Cassani, B, Andolfi, G, Mirolo, M, Biasco, L, Recchia, A *et al.* (2007). Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. *J Clin Invest* **117**: 2233–2240.
17. Schwarzwaelder, K, Howe, SJ, Schmidt, M, Brugman, MH, Deichmann, A, Glimm, H *et al.* (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution *in vivo*. *J Clin Invest* **117**: 2241–2249.
18. Nienhuis, AW, Dunbar, CE and Sorrentino, BP (2006). Genotoxicity of retroviral integration in hematopoietic cells. *Mol Ther* **13**: 1031–1049.
19. Modlich, U, Bohne, J, Schmidt, M, von Kalle, C, Knöss, S, Schambach, A *et al.* (2006). Cell culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood* **108**: 2545–2453.
20. Zychlinski, D, Schambach, A, Modlich, U, Maetzig, T, Meyer, J, Grassman, E *et al.* (2008). Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Mol Ther* **16**: 718–725.
21. Modlich, U, Schambach, A, Brugman, MH, Wicke, DC, Knoess, S, Li, Z *et al.* (2008). Leukemia induction after a single retroviral vector insertion in Evi1 or Prdm16. *Leukemia* **8**: 1519–1528.
22. Li, Z, Fehse, B, Schiedlmeier, B, Düllmann, J, Frank, O, Zander, AR *et al.* (2002). Persisting multilineage transgene expression in the clonal progeny of a hematopoietic stem cell. *Leukemia* **16**: 1655–1663.
23. Schambach, A, Mueller, D, Galla, M, Verstegen, MM, Wagemaker, G, Loew, R *et al.* (2006). Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene Ther* **13**: 1524–1533.
24. Schambach, A, Wodrich, H, Hildinger, M, Bohne, J, Kräusslich, HG and Baum, C (2000). Context dependence of different modules for posttranscriptional enhancement of gene expression from retroviral vectors. *Mol Ther* **2**: 435–445.
25. Kustikova, O, Geiger, H, Li, Z, Brugman, MH, Chambers, SM, Shaw, CA *et al.* (2007). Retroviral vector insertion sites associated with dominant hematopoietic clones mark "stemness" pathways. *Blood* **109**: 1897–1907.
26. Kustikova, OS, Baum, C and Fehse, B (2008). Retroviral integration site analysis in hematopoietic stem cells. *Methods Mol Biol* **430**: 255–267.
27. Purton, LE and Scadden, DT (2007). Limiting factors in murine hematopoietic stem cell assays. *Cell Stem Cell* **1**: 263–270.
28. McKenzie, JL, Gan, OI, Doedens, M, Wang, JC and Dick, JE (2006). Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment. *Nat Immunol* **7**: 1225–1233.
29. Akagi, K, Suzuki, T, Stephens, RM, Jenkins, NA and Copeland, NG (2004). RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res* **32**: D523–D527.
30. Hammond, SM, Crable, SC and Anderson, KP (2005). Negative regulatory elements are present in the human LMO2 oncogene and may contribute to its expression in leukemia. *Leuk Res* **29**: 89–97.
31. Cattoglio, C, Facchini, G, Sartori, D, Antonelli, A, Miccio, A, Cassani, B *et al.* (2007). Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood* **110**: 1770–1778.
32. Mitchell, RS, Beitzel, BF, Schroder, AR, Shinn, P, Chen, H, Berry, CC *et al.* (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* **2**: E234.
33. Chambers, SM, Boles, NC, Lin, KY, Tierney, MP, Bowman, TV, Bradfute, SB *et al.* (2007). Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell* **1**: 578–591.
34. Fraser, C, Szilvassy, S, Eaves, C and Humphries, R (1992). Proliferation of totipotent hematopoietic stem cells *in vitro* with retention of long-term competitive *in vivo* reconstituting ability. *Proc Natl Acad Sci USA* **89**: 1968–1972.
35. Goebel, WS, Yoder, MC, Pech, NK and Dinauer, MC (2002). Donor chimerism and stem cell function in a murine congenic transplantation model after low-dose radiation conditioning: effects of a retroviral-mediated gene transfer protocol and implications for gene therapy. *Exp Hematol* **30**: 1324–1332.
36. Gaspar, HB, Bjorkegren, E, Parsley, K, Gilmour, KC, King, D, Sinclair, J *et al.* (2006). Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning. *Mol Ther* **14**: 505–513.
37. Seggewiss, R, Pittaluga, S, Adler, RL, Guenaga, FJ, Ferguson, C, Pilz, IH *et al.* (2006). Acute myeloid leukemia is associated with retroviral gene transfer to hematopoietic progenitor cells in a rhesus macaque. *Blood* **107**: 3865–3867.
38. van der Putten, H, Quint, W, van Raaij, J, Maandag, ER, Verma, IM and Berns, A (1981). M-MuLV-induced leukemogenesis: integration and structure of recombinant proviruses in tumors. *Cell* **24**: 729–739.
39. Kohn, DB, Sadelain, M, Dunbar, C, Bodine, D, Kiem, HP, Candotti, F *et al.* (2003). American Society of Gene Therapy (ASGT) ad hoc subcommittee on retroviral-mediated gene transfer to hematopoietic stem cells. *Mol Ther* **8**: 180–187.
40. Montini, E, Cesana, D, Schmidt, M, Sanvito, F, Ponzoni, M, Bartholomae, C *et al.* (2008). The genotoxic potential of integrative vectors is strongly modulated by vector design and the profile of integration site selection. *Mol Ther* **16** (suppl. 1): S362.
41. Hargrove, PW, Kepes, S, Hanawa, H, Obenauer, JC, Pei, D, Cheng, C *et al.* (2008). Globin lentiviral vector insertions can perturb the expression of endogenous genes in beta-thalassemic hematopoietic cells. *Mol Ther* **16**: 525–533.
42. Mooslehner, K, Karls, U and Harbers, K (1990). Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J Virol* **64**: 3056–3058.
43. Wu, X, Li, Y, Crise, B and Burgess, SM (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**: 1749–1751.
44. Berry, C, Hannenhalli, S, Leipzig, J and Bushman, FD (2006). Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput Biol* **2**: e157.
45. Recchia, A, Cittaro, D, Cattoglio, C, Felice, B, Miccio, A, Ferrari, G *et al.* (2008). Integration of retroviral vectors into the human genome is biased by specific subsets of transcription factor binding sites. *Mol Ther* **16** (suppl. 1): S360.
46. Wang, GP, Garrigue, A, Ciuffi, A, Ronen, K, Leipzig, J, Berry, C *et al.* (2008). DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. *Nucleic Acids Res* **36**: e49.
47. Beyer, W, Westphal, M, Ostertag, W and von Laer, D (2002). Oncoretrovirus and lentivirus vectors pseudotyped with lymphocytic choriomeningitis virus glycoprotein: generation, concentration, and broad host range. *J Virol* **76**: 1488–1495.
48. Li, Z, Schwieger, M, Lange, C, Kraunus, J, Sun, H, van den Akker, E *et al.* (2003). Predictable and efficient retroviral gene transfer into murine bone marrow repopulating cells using a defined vector dose. *Exp Hematol* **31**: 1206–1214.
49. Pfaffl, MW (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: e45.
50. R Development Core Team (2008). R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.