Seminars in Cancer Biology 20 (2010) 269–278

Contents lists available at ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Review

The genomic risk of somatic gene therapy

Olga Kustikova, Martijn Brugman, Christopher Baum*

Department of Experimental Hematology, Hannover Medical School, Carl-Neuberg-Straße 1, D-30625 Hannover, Germany

ARTICLE INFO

Keywords:
Retrovirus
Lentivirus
Transposon
Insertional mutagenesis
Oncogene
Tumor
Hematopoiesis

ABSTRACT

Gene vectors with an untargeted insertion profile have been explored in preclinical models and clinical trials for the transfer of potentially therapeutic genetic information into somatic cells that have a high replicative potential. The gene-modified cell population can be viewed as a genetic mosaic whose complexity depends upon the target cell type, the number of transduced cells, the average number of insertions per cell, the genetic stability and composition of the transgene, and the integration pattern of the vector. Selection by the environment encountered in the patient may support the preferential survival of clones with insertional deregulation of genes that are involved in the control of engraftment, proliferation or differentiation, in the worst case initiating oncogenic progression. Rapid scientific and technological progress has shed much light onto this dark side of untargeted vector integration. New approaches to unbiased and highly sensitive "integromics" promise a precise documentation of stable polyclonality, clonal fluctuation or clonal imbalance of gene-modified cell populations. Evidence has been obtained for a number of approaches to potentially reduce the genomic risk of gene therapy: targeting cells that lack sustained replicative potential, using vectors with a more neutral integration spectrum, reducing the number of vector copies per cell, designing gene expression cassettes that avoid long-distance enhancer interactions or fusion transcripts, and reducing, as far as possible, the risk of secondary mutations. The genomic risk of gene therapy can thus be prevented by the collective targeting of all contributing factors.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The basic principle of gene therapy is the genetic modification of somatic cells, excluding cells contributing to the germline, to prevent or cure disease. Based on insights into the molecular mechanisms of pathogenesis, gene transfer may correct or replace a defective gene function, or instruct somatic cells to adopt novel therapeutic properties. Numerous approaches target post-mitotic cells or use gene vectors that have a low probability of integration into the host genome. With a single exception discussed further below [1], functional consequences of genomic disturbance have not been described in this context. Another strategy is the genetic modification of cells with a high replicative potential, such as somatic stem and progenitor cells, by vectors that mediate efficient chromosomal integration for stable transgene inheritance by the progeny of the targeted population. In this case, gene transfer may introduce a high degree of genomic diversity, depending on the chromosomal integration pattern of the vector used. The complexity of the genetic mosaic that originates from the use of integrating gene vectors depends upon numerous variables, such as the target cell type, the number of transduced cells, the average number of insertions per cell, the genetic stability and composition of the transgene, and the integration pattern of the vector [2–6].

Insertional mutagenesis refers to the alteration of the cellular genome by the integration of viral or recombinant DNA, and is the default outcome of transgene integration, unless homologous recombination is achieved. Whether or not insertional mutagenesis may be overtly genotoxic and induce adverse events has been debated since the early days of gene therapy. Before 2002, the likelihood of cancer induction by insertional mutagenesis had mostly been considered negligible. However, since 2002, animal models and unfortunately also clinical trials have revealed that the genomic risk associated with insertional mutagenesis originating from replication-defective gene vectors can be profound (Fig. 1) [1,7–18]. Evidence for the curative potential of gene therapy is increasing though [19–22], and the support for this fundamentally new medical approach may be substantially higher if the genomic risk could be controlled.

A puzzling observation made over the past years was that even when using very similar gene vectors in related target cells, the severity of genotoxic adverse events manifested in a highly context-dependent manner. For example, transducing hematopoietic stem/progenitor cells with gammaretroviral vectors designed to express *Gp91phox*, the gene defective in an X-linked form of Chronic Granulomatous Disease (CGD), induced only minor alter-

* Corresponding author. Tel.: +49 511 532 6067; fax: +49 511 532 6068.
E-mail address: baum.christopher@mh-hannover.de (C. Baum).

1044-579X/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.semcancer.2010.06.003

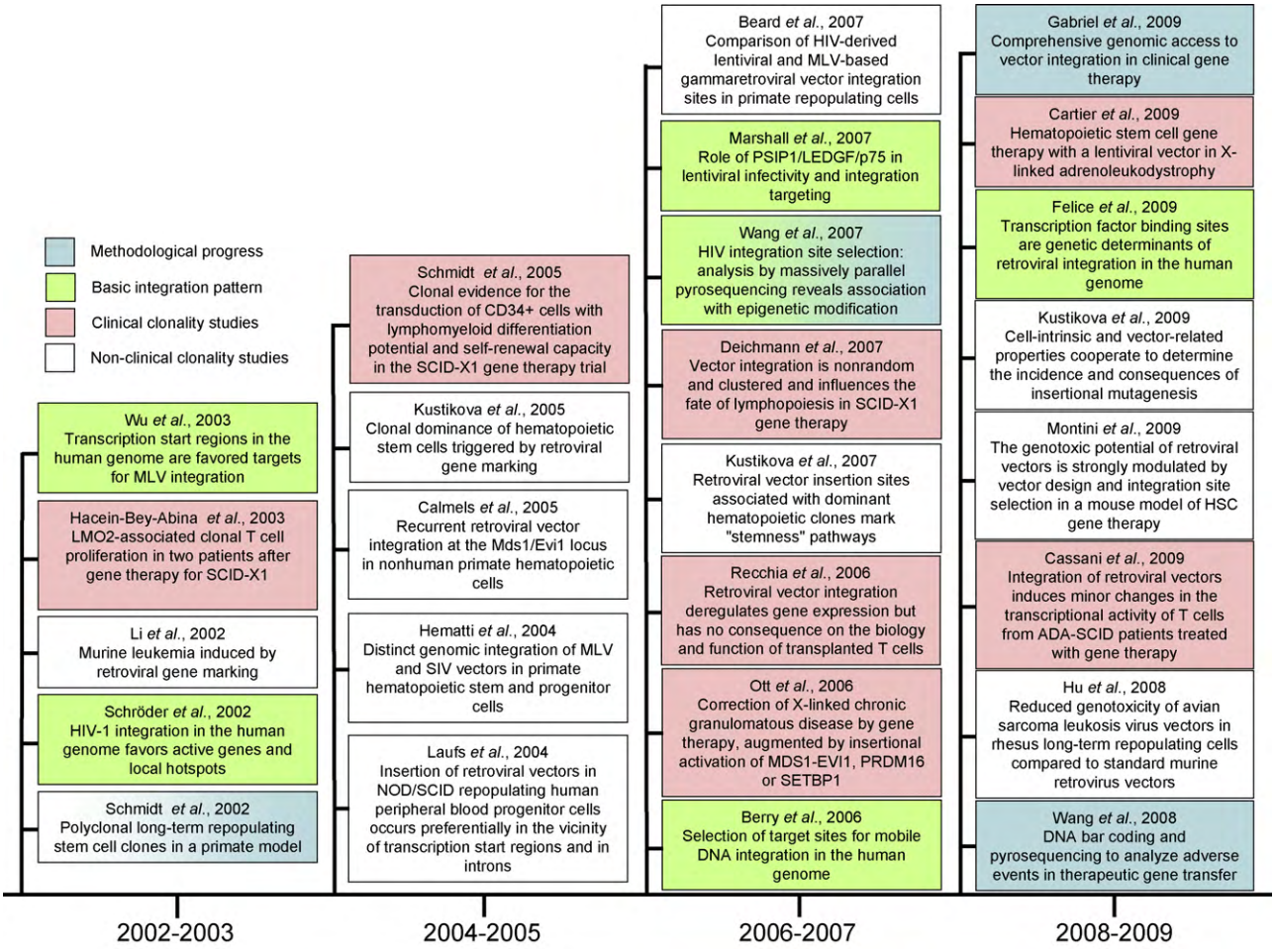

Fig. 1. Overview of studies describing key observations in the emerging field of integromics.

ations in hematopoietic clonality in murine models [23–26] and caused no overt clonal imbalance in rhesus monkeys (despite insertions close to the proto-oncogenes *PIM2* and *MDS/EVI1*) [27], but led to a substantial proliferative advantage of clones with insertions in the proto-oncogenes *MDS/EVI1* and *PRDM16* in two adult CGD patients [7]. In addition, a relatively high incidence of malignant lymphoproliferation was observed in a pioneering clinical trial to correct SCID-X1 disease [18,28,29], while a related trial performed by other investigators with minor modifications of the experimental protocol had a trend to a reduced rate of insertional complications [30,31]. The use of similar "first generation" gammaretroviral vectors with highly active enhancer sequences in the long terminal repeats (LTRs) to treat patients suffering from severe combined immunodeficiency caused by mutations of adenosine deaminase (SCID-ADA) has so far remained free of insertional leukemia or preleukemia even at the molecular level [20,32,33].

Research over the past years has made considerable progress in the understanding of the variables that contribute to the genomic risk associated with insertional mutagenesis in gene therapy. In the present review, we discuss key observations related to insertional adverse events in human gene therapy and preclinical models, address the contributing factors, describe ways to refine the clonal tracking of gene-modified cells, and consider recent progress and the remaining challenges encountered in the methodology of integromics. The concluding chapter develops an outline of important research questions and avenues for future investigations.

2. Adverse events related to genotoxicity and contributing factors

Since at least 1% of the gene repertoire may be involved in the various forms of cancer, untargeted transgene integration has a relatively high potential to affect genes regulating cell growth and survival. However, the induction of cancer by insertional mutagenesis in gene therapy was considered unlikely, as cancer is a multistep process that requires quite specific forms of oncogene collaboration [34]. When the first cases of oncogenic adverse events following transgene insertion close to cellular proto-oncogenes were reported in a murine model and a clinical trial [8,9], the potential collaboration of insertional mutagenesis with signalling alterations caused by the expressed transgene was considered. In a murine gene marking study, originally designed to address potential adverse events related to the expression of cell-surface proteins, a rather unusual manifestation of monocytic leukemia was observed after a vector encoding a truncated form of the low affinity nerve growth factor receptor (dLNGFR) had inserted into and upregulated the proto-oncogene *Evi1* [8]. In a gene therapy trial to treat CGD, the corresponding human locus was involved in two cases of progressive clonal myeloproliferation, with secondary genetic instability contributing to the manifestation of a myelodysplastic syndrome [35]. In 4 of the 5 patients that suffered from X-linked severe combined immunodeficiency (SCID-X1) and developed leukemogenic complications a few years after infusion of gene-modified hematopoietic stem/progenitor cells, insertions of the gammaretroviral vector encoding the interleukin 2 receptor

common gamma chain (*IL2RG*) have been observed in or upstream of the proto-oncogene *LMO2*, again leading to its transcriptional upregulation and consequent malignant progression that was associated with secondary mutations [9,18,31].

In both situations (the murine marking study and the SCID-X1 clinical trial) evidence for a collaboration of the insertional proto-oncogene upregulation with potential signalling alterations induced by the expressed receptor molecules remained indirect. While a leukemogenic potential for activated neurotrophin signalling was clearly described [36], a causal involvement of dLNGFR was not demonstrated although it may well have contributed to the unusual morphology of the leukemia. It rather appears as if the upregulation of *Evi1* or the related gene *Prdm16* can be sufficient to induce leukemia, at least in mice [37]. As indicated above, upregulation of *EVI1* was associated with premalignant clonal dominance and progressive myelodysplasia in CGD patients [7,35], but not in the rhesus model [27,38]. This discrepancy points to the existence of disease-associated cofactors (such as stress hematopoiesis caused by cytokine conditions or associated with infections) that may contribute to the preferential selection of insertional mutants. Furthermore, although murine leukemias have been identified in which replication-competent retroviruses had hit both *Lmo2* and *Il2rg* in the same malignant clone [39], experiments conducted to reveal a potential direct collaboration remained inconclusive. While some studies favored a substantial role for ectopic expression of *IL2RG* in insertional leukemogenesis, others pointed to a dominant role for *LMO2* [10,40–42]. Again, disease-specific cofactors such as an increased replicative stress or immunodeficiency may have contributed to the acquisition of secondary mutations and the failure to reject the leukemic clone [43].

Examples that suggest a direct interaction of the "therapeutic" transgene with insertional upregulation of a proto-oncogene include cases of myeloid leukemia observed in large animal models after introducing a retroviral vector expressing the HOXB4 transcription factor to amplify gene-modified hematopoietic cells [11], and cases of erythroid leukemia after expressing the Thrombopoietin receptor gene, *Mpl*, from gammaretroviral vectors [44]. The latter constellation is reminiscent of the paradigmatic collaboration of activated erythropoietic receptor signalling with insertional activation of transcription factors, as observed in erythroleukemias induced by the replication-competent Friend retrovirus complex [45]. It appears safe to conclude that the type of gene expressed by the gene transfer vector may, but does not necessarily, affect the incidence, latency, and cell-type specificity of insertional oncogenesis.

Importantly, both insertional leukemias and liver tumors have been observed, at least in animal models, following the use of gene vectors encoding genes for which a direct interference with cell survival pathways is unlikely: A gammaretroviral vector encoding a mutant form of dehydrofolate reductase induced myeloid sarcoma in a monkey by insertion into an antiapoptotic gene [12], gammaretroviral vectors encoding fluorescent proteins induced insertional leukemias in mice following the transplantation of transduced hematopoietic stem/progenitor cells [37,46], both gammaretroviral and lentiviral vectors encoding fluorescent proteins accelerated oncogenesis in a tumor-prone mouse model, with incidence and kinetics depending on details of the vector design [47] and liver tumors with insertional deregulation of a miRNA gene cluster were observed after treating neonatal (but not adult) mice with Adeno-associated vectors (AAV) vectors encoding a metabolic gene, β-glucuronidase [1]. Although the latter case is unique in that it occurred with the use of a vector that has a relatively low likelihood of integration, the number of insertion events occurring after administration *in vivo* was obviously still high enough to hit dangerous loci, and oncogenic transformation may have been supported by secondary mutations related to the substantial postnatal proliferation of the transduced tissue. Based on these findings, expression of a gene involved in cell signalling pathways is not necessary to induce cancer by gene vector-dependent insertional mutagenesis.

To decipher the mechanisms for risk reduction, it is thus of great importance to understand the role of the vector insertion pattern and cis-regulatory components of the transgene cassette. Over the past years it has become apparent that each vector system has a characteristic genomic insertion pattern. This is mostly dependent upon features of the integrase or respective DNA insertion enzyme and associated cellular factors such as the paradigmatic interaction with p75/LEDGF in the case of HIV-1 and derived lentiviral vectors [48]. Sequences in the U3-region of the retroviral LTR may also contribute to the fine tuning of the insertion pattern, most likely by tethering the preintegration complex to chromatin via interaction with transcription factors that recognize their genomic binding sites [49]. The most frequently referenced generic differences affect the proximity of insertions to the transcriptional start site (TSS) and other regulatory sequences such as CpG islands, and the association with transcribed gene regions. While regulatory sequences are preferentially targeted by gammaretroviruses and derived vectors, transcribed gene regions are a preferred location for HIV-1 and related lentiviral vectors [49–53]. Alpharetroviruses and derived vectors as well as the Sleeping Beauty transposon appear to have a reduced affinity for active genes or regulatory regions [54–56]. Foamy virus (spumavirus) stands somewhat in between gammaretroviruses and alpharetroviruses [57]. Furthermore, gammaretroviruses have been shown to have a propensity for growth-regulatory genes [51], which may be related to their dependence on cell mitosis for gaining access to chromosomes. Even without selection for a competitive growth advantage of transduced cells, gammaretroviral vectors were thus shown to have a greater propensity to insert near proto-oncogenes than lentiviral vectors [51].

A selection of studies addressing these genomic integration patterns is provided in Table 1 [7,15,16,49,50,52–65]. This overview, as well as individual studies contained therein, illustrate the fact that the characteristic differences related to the type of integrase (or transposase) are further modified by other experimental conditions such as vector architecture and content, cell type, and observation time. For future investigations and interlaboratory comparisons, it would be helpful to appoint uniform experimental conditions and bioinformatical approaches.

What are the functional consequences of these insertional preferences? A survey of studies designed to address this important question (some of which listed in the key references that we summarized in Fig. 1 [7–9,15,19,29,46,48–50,53,56,58–60,66–77]) reveals that the numerical differences noted in the insertion pattern obtained from freshly transduced cells do not directly predict the risk of insertional cell transformation.

Extensive work performed in a tumor-prone murine model ($Cdkn2a^{-/-}$ mice) indicated that the lentiviral insertion pattern is about 10 times safer than the gammaretroviral [47,77]. This became apparent when calculating a hazard function which included the latency of tumor development and the number of vector copies per tumor. Because of the relatively rapid development of spontaneous tumors in this model, the genotoxic impact of vectors in which gene expression is driven from an internal promoter rather than from the LTR could not be determined. Furthermore, a cell line in which transformation events occurring in association with intragenic and fusion transcript-generating insertion events were selected for, demonstrated the mutagenic potential of integrations in active genes [78]. Another experimental system, which selects for transformation of cultured primary hematopoietic cells, has indicated that lentiviral vectors are ~3-fold less likely than similarly configured gammaretroviral vectors

Table 1
Comparison of studies investigating vector integration site profiles.

Study	Vector basis	Cell type	Species (time of analysis)	Number of VIS	Proximity to transcriptional start sites		Proximity to CpG islands		% of integration into RefSeq (random control)	% of integration within/near POG/RTCGD CIS (random control)
					Window (kb)	% of events (random control)	Window (kb)	% of events (random control)		
Wu et al. [50]	MLV	HeLa	Human (48 h *in vitro*)	903	±5	20.2 (4.3)	±1	16.8 (2.1)	34.2 (22.4)	N.A.
Lewinski et al. [52]	MLV	HeLa[a]	Human (2 weeks *in vitro*)	544	±5	26.1 (5.0)	±1	11.8 (1.0)	44.3 (33.9)	N.A.
Kustikova et al. [60]	MLV	Lin$^-$BM	Mouse (3–7 months *in vivo*)	280[b]	−1 to 0 −2.5 to −1 +1 to +2.5	5.5[c] (0.5) 5.9[c] (0.5) 2.9[c] (0.7)	N.A.	N.A.	N.A.	22.5
Laufs et al. [59]	MLV	CD34$^+$	Human in mouse (6–8 weeks *in vivo*)	207	±5	19 (4.3[d])	±1	10 (2.1[d])	40 (22.4[d])	N.A.
Hematti et al. [58]	MLV	CD34$^+$	Rhesus monkey (>5 months *in vivo*)	432	±2 −30 +30	11 (N.A.) 20 (N.A.) 27 (N.A.)	N.A.	N.A.	49.1 (31.6[e])	N.A.
Ott et al. [7]	MLV	CD34$^+$	Human, Patient 1 (21–542 days *in vivo*)	435	±5	20	N.A.	N.A.	47	N.A.
Ott et al. [7]	MLV	CD34$^+$	Human, Patient 2 (24–343 days *in vivo*)	330	±5	21	N.A.	N.A.	52	N.A.
Beard et al. [53]	MLV	CD34$^+$	Canine (95–600 days *in vivo*)	82	±10 ±50	34.2 (14.2) 73.2 (44.7)	±5	3.7 (1.6)	48.8 (37.3)	7.3 (2.4)
Schwarzwaelder et al. [16]	MLV	CD34$^+$	Human, 5 Patients (9–30 months *in vivo*)	304	±5	23 (29[f])	±1	13 (14)	41 (49[f])	N.A.
Deichmann et al. [15]	MLV	CD34$^+$	Human, 9 patients (4–41 months *in vivo*)	572	±5	27	±1	17	38	N.A.
Hu et al. [56]	MLV	CD34$^+$	Rhesus monkey (>5 months *in vivo*)	396	−10 −30	10.6 (5.5) 23.7 (14.2)	Within CpG	0.8 (0.9)	45.7 (33.8)	N.A.
Felice et al. [49]	MLV	CD34$^+$	Human (5–12 days *in vitro*)	829	±5	16.6 (4.5[g])	±1	6.5 (0.9[g])	46.4 (35.7[g])	N.A.
Felice et al. [49]	MLV (ΔU3)	CD34$^+$	Human (5–12 days *in vitro*)	200	±5	12.5 (4.5[g])	±1	4.5 (0.9[g])	40.0 (35.7[g])	N.A.
Felice et al. [49]	MLV (SFFV)	CD34$^+$	Human (5–12 days *in vitro*)	195	±5	19.0 (4.5[g])	±1	8.2 (0.9[g])	39.0 (35.7[g])	N.A.
Wu et al. [50]	HIV-1	HeLa/H9	Human (48 h *in vitro*)	379	±5	10.8 (4.3)	±1	2.1 (2.1)	57.8 (22.4)	N.A.
Lewinski et al. [52]	HIV-1	HeLa[a]	Human (2 weeks *in vitro*)	525	±5	6.9 (5.0)	±1	0.2 (1.0)	77.9 (33.9)	N.A.
Lewinski et al. [52]	HIV-1 mGag	HeLa[a]	Human (2 weeks *in vitro*)	493	±5	3.9 (5.0)	±1	0.0 (1.0)	66.7 (33.9)	N.A.
Lewinski et al. [52]	HIV-1 mIN	HeLa[a]	Human (2 weeks *in vitro*)	352	±5	20.7 (5.0)	±1	9.4 (1.0)	38.9 (33.9)	N.A.
Lewinski et al. [52]	HIV-1 mGag-mIN	HeLa[a]	Human (2 weeks *in vitro*)	526	±5	22.4 (5.0)	±1	9.9 (1.0)	42.6 (33.9)	N.A.
Laufs et al. [61]	HIV-1	CD34$^+$	Human in mouse (>12 weeks *in vivo*)	57	±10	20	N.A.	N.A.	68	N.A.
Beard et al. [53]	HIV-1	CD34$^+$	Canine (363–498 days *in vivo*)	210	±10 ±50	15.2 (14.2) 44.3 (44.7)	±5	0.0 (1.6)	58.6 (37.3)	4.3 (2.4)
Felice et al. [49]	HIV-1	CD34$^+$	Human (5–12 days *in vitro*)	403	±5	8.4 (4.5[g])	±1	0.0 (0.9[g])	63.5 (35.7[g])	N.A.
Felice et al. [49]	HIV-1 (ΔU3-CMV)	CD34$^+$	Human (5–12 days *in vitro*)	445	±5	7.4 (4.5[g])	±1	0.2 (0.9[g])	66.1 (35.7[g])	N.A.

Felice et al. [49]	HIV-1 (ΔU3-MLV)	$CD34^+$	Human (5–12 days *in vitro*)	200	±5	9.5 (4.5[g])	±1	0.0 (0.9[g])	66.0 (35.7[g])	N.A.
Felice et al. [49]	HIV-1 (MLV U3)	$CD34^+$	Human (5–12 days *in vitro*)	400	±5	10.0 (4.5[g])	±1	0.5 (0.9[g])	64.0 (35.7[g])	N.A.
Hematti et al. [58]	SIV	$CD34^+$	Rhesus monkey (>6 months *in vivo*)	328	±2 −30 +30	2 (N.A.) 9 (N.A.) 37 (N.A.)	N.A.	N.A.	73.5 (31.6[e])	N.A.
Hu et al. [56]	SIV	$CD34^+$	Rhesus monkey (>6 months *in vivo*)	289	−10 −30	5.9 (5.5) 21.1 (14.2)	Within CpG	0.0 (0.9)	77.2 (33.8)	N.A.
Narezkina et al. [64]	ASV	HeLa	Human (48 h *in vitro*)	226	±5	8.4 (4.3[d])	±1	3.1 (2.1[d])	39.8 (22.4[d])	N.A.
Mitchell et al. [54]	ASLV	293 TVA	Human (48 h *in vitro*)	469	±5	N.A.	±1	3.2 (1.9[h])	38.2 (33.2[h])	N.A.
Hu et al. [56]	ASLV	$CD34^+$	Rhesus monkey (PB at 4, 6, 12, and 18 months)	198	−10 -30	8.6 (5.5) 17.7 (14.2)	Within CpG	1.5 (0.9)	42.9 (33.8)	N.A.
Trobridge et al. [57]	FV	HT-1080	Human (8 days *in vitro*)	1008	±2.5	10.6[i]	±1	5.8 (2.6)	24.0 (34.1)	4.4[i] (3.2)
Trobridge et al. [57]	FV	$CD34^+$	Human (6 days *in vitro*)	1821	±2.5	10.6[i]	±1	12.1 (2.6)	31.1 (34.1)	4.4[i] (3.2)
Beard et al. [53]	FV	$CD34^+$	Canine (117–519 days *in vivo*)	263	±10 ±50	20.5 (14.2) 53.6 (44.7)	±5	3.4 (1.6)	30.0 (37.3)	2.7 (2.4)
Yant et al. [55]	SB	Huh-7	Human (14 days *in vitro*)	366	±5	8.5 (5.4)	±1 ±5	2.5 (1.9) 11.2 (8.3)	39.1 (33.2)	N.A.
Yant et al. [55]	SB	Liver	Mouse (2 days *in vivo*)	590	±5	10.5 (6.4)	±5	9.5 (6.4)	31.2 (26.0)	N.A.
Yant et al. [55]	SB	NIH 3T3	Mouse (14 days *in vitro*)	380	±5	9.2 (6.4)	±5	8.7 (6.4)	32.9 (26.0)	N.A.
Wilson et al. [63]	Piggy-Bac trans-poson	HeLa/HEK-293	Human (2–3 weeks *in vitro*)	575	±5	16.2 (5.4[j])	±1 ±5	3.8 (1.9[j]) 7.7 (8.3[j])	48.8 (33.2[j])	N.A.
Miller et al. [65]	AAV2	MHF2	Human (at least 14 days *in vitro*)	977	+1	3-Fold enrichment	Within CpG	4.03 (0.8)	38.8 (34.7)	N.A.

MLV, moloney leukemia virus; HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; ASV, avian sarcoma virus; ASLV, avian sarcoma leukosis virus; AAV, Adeno-associated virus; FV, Foamy virus; SB, Sleeping Beauty transposon; N.A., not available; PB, peripheral blood samples; Lin^-BM, lineage negative bone marrow cells; POG, proto-oncogenes; RTCGD, retroviral tagged cancer gene database; CIS, common integration site; VIS, vector integration site.

[a] Puromycin selection.

[b] Database contains integrations retrieved from both healthy and leukemic animals of long term bone marrow transplantation studies.

[c] Data are shown with respect to CIS, POG and self-renewal genes.

[d] Random calculations derived from Wu et al. [50].

[e] In silico generated random integrations.

[f] Healthy donor.

[g] Control sequences were randomly cloned by LM-PCR from $CD34^+$ DNA samples.

[h] Adjusted values from the work of Mitchell et al. [54] provided by Yant et al. [55].

[i] Combined data for both FV HT-1080 and $CD34^+$.

[j] Random calculations derived from Yant et al. [55].

to transform cells by an enhancer-mediated activation of the *Evi1* proto-oncogene [79]. In both backbones, replacing a viral enhancer-promoter with a more restricted cellular promoter greatly reduced the risk of insertional cell transformation, in most cases below background levels [79–81]. The inclusion of so-called insulator elements may also reduce the risk of clonal skewing or cell transformation by vector insertion [80–84], although here again a profound context-dependence can be noted in that insulators may function in a very promoter- (and potentially cell-) specific way [85].

Finally, when examining the integrome in large animals which showed normal hematopoiesis after transplantation of hematopoietic cells transduced with lentiviral or gammaretroviral vectors, both vectors showed a similar preference for integrating close to actively transcribed genes and proto-oncogenes, without obvious functional consequences [53]. Collectively, these studies show that not only vector origin and design, but also the context of cell selection and the molecular architecture of a given gene locus determine whether and to which degree the insertional differences impact on the genomic risk of integrating gene vectors.

Cell clones were isolated in which vector insertion sites were mapped to study the deregulation of adjacent cellular genes [71,86,87], circumventing a potential bias introduced by selection for transforming events. These studies revealed that the genetic components of the integrated transgene determine both the likelihood and magnitude of gene deregulation, with potential consequences for both the incidence of transforming events and the fitness of the transformed cells. The identification of promoter sequences that avoid long-distance enhancer interactions, genetic insulator sequences that inhibit the regulatory cross-talk of the transgene with surrounding cellular genes and the improvement of mRNA processing (to avoid the occurrence of fusion transcripts and reduce the need for strong enhancers) should thus have a great potential to reduce the risk of insertional mutagenesis.

Illustrating the importance of vector design, stable hematopoiesis without vector-induced clonal imbalance has been observed with lentiviral and gammaretroviral vectors lacking strong enhancer elements [88,89], in the latter case even when insertions occurred in the vicinity of proto-oncogenes such as *Lmo2* [89]. The distinction between induced and spontaneous clonal dominance can thus be difficult [4,74,89]. Furthermore, increasing evidence suggests that cell-intrinsic determinants greatly contribute to the risk of insertional mutagenesis. Thus, T cells and also hematopoietic progenitor cells lacking self-renewal capacity were shown to be more resistant than primitive hematopoietic stem cells to generate expanding cell clones with oncogenic lesions [71,74,90]. While cell-intrinsic features are unlikely to constitute suitable targets for prevention strategies, research addressing the nature of the systemic forces that contribute to the selection of cell mutants is likely to provide additional strategies for risk reduction. Such systemic forces may act via at least two distinct mechanisms: Avoiding "negative" milieu conditions in which insertional mutants have a growth advantage over normal competitors, and reducing the replicative stress, thus avoiding the acquisition of secondary mutations.

3. Tracking genetically modified clones

Identifying the insertion site profile of different vector types in unselected or selected cell populations represents an important aspect in comparative biosafety studies. Various polymerase chain reaction (PCR)-based methods have been developed for this purpose, including inverse PCR [91,92], vectorette- and splinkerette-PCR [93–95], ligation-mediated (LM) PCR [50,61,96–100], and linear amplification-mediated (LAM) PCR [69,101] or a related approach called EFRAM PCR [102]. The latter two methods (LAM PCR and EFRAM PCR) have a higher sensitivity than LM PCR for detecting polyclonal insertion patterns [69]. If the amount of DNA available for insertion site determination is limiting, genome amplification methods can be of additional help [103]. In situations of clonal dominance, the number of insertion sites obtained by LM PCR correlates well with Southern blot results, and recovery of insertion sites was in the range of about 80% when using a single restriction enzyme [46,66]. LM PCR thus generates reliable and reproducible results for the investigation of clonal samples *in vivo* and *in vitro* [60], and allows the identification of genes in the vicinity of vector integration with the potential to influence competitive fitness. A list of retroviral insertion sites associated with clonal dominance in case of normal or potentially preleukemic hematopoiesis or malignant transformation of murine hematopoietic cells was reported in the Insertional Dominance Database (IDDb). Remarkably, the IDDb shows substantial overlap not only with a database of insertion sites recovered from tumors induced by replication-competent gammaretroviruses (retrovirus-tagged cancer gene database, RTCGD [104]), but also with the transcriptome of hematopoietic stem/progenitor cells [60].

The potential bias introduced by the methods used for insertion site retrieval also deserves special attention. Methods that use restriction endonucleases to generate products of different sizes and rely on PCR to amplify the product for subsequent sequencing may not be suitable for the generation of complete integromes. More recently introduced methods exploit the large numbers of sequences that could be retrieved by pyrosequencing approaches with primer bar-coding [105] allowing the simultaneous analysis of hundreds if not thousands of samples in a single high throughput sequence run [70]. Furthermore, methods that do not rely on restriction digest for amplification of insertion site junctions [73,106] or the use of randomly inserting bacteriophage Mu transposons to generate fragments that can be sequenced [107] have been introduced.

To exactly quantify the kinetics of clonal dominance, selected clones can be studied by longitudinal locus-specific, semi-quantitative or quantitative PCR [7,9,100,108]. It would certainly be of advantage to the field if the retrieval rate of high throughput sequencing methods could serve as a measure of clonal abundance. However, spike-in gene expression measurements performed by Solexa sequencing showed that retrieval rates might not necessarily reflect the underlying amount of RNA and the protocol and amplification used seems to be critical [109]. Ideally, sequence read counts should be supplemented with locus-specific real-time PCR to determine clonal kinetics.

4. Bioinformatical approaches to integromics

The RTCGD [104] collects insertion sites retrieved from murine tumors that were induced by replicating retroviruses, and has proved a valuable reference for gene vector integration patterns [110–118]. The bulk of the RTCGD database contains gammaretroviral insertions, and a potential oncogenic function of the affected gene is assumed if common insertion sites (CIS) are found in the same gene region when analyzing different tumors. However, murine tumors described in the RTCGD contain up to 27 insertions per tumor (1–27, median 3 in 1643 tumors) many of which probably represent innocent "bystander" mutations. It is also worth noting that hotspots of retroviral insertion sites have been identified in freshly transduced, unselected hematopoietic cells [51]. Furthermore, although tumor tissue was analyzed, it could be expected that several of the insertions would originate from untransformed tumor-infiltrating leukocytes. As the majority of the potential proto-oncogenes described in the RTCGD are not functionally validated, this database may thus be "contaminated" with

gammaretroviral insertion hotspots that are not causally linked to tumor formation [4,119]. Since hotspots are relatively specific for the integration mechanism of a given type of retrovirus, it is not unexpected that gammaretroviral vector insertions observed in preclinical models and clinical trials typically show a larger degree of overlap with the RTCGD than those of other types of integrating vectors. This issue complicates the comparative genomic analyses of the transforming risk of different types of gene vectors [47,51,53,74,77].

With the advent of plate sequencing and high throughput pyrosequencing [18,70,76], due to the sheer number of integration sites retrieved, the list of CIS is also expected to increase. Originally, the definition of a CIS was made after a statistical analysis of 1200 randomly generated insertions, which predicted that a certain number of insertions within a certain area was significant [120]. Suitable approaches have been described to determine whether a given amount of insertions in a certain area is unexpectedly high [121]. Based on the RTCGD, a kernel permutation model was used to find statistically significant overrepresented sites [122]. In a follow-up paper, a number of clusters were identified with potential causal contribution to the tumor phenotype because they were shown to commonly occur in different tumors [123]; 86 clusters of insertions (out of a total of 5473 insertion in 1361 tumors) were filtered that indicated cooperation between genes targeted by the insertions, statistically weeding out the bystander mutations from the dataset.

Other commonly performed types of insertion site analysis can be divided into two categories: those relating to the chromosomal features and those relating to the genes that might be affected by the insertion. The first type of analysis includes the distance to the TSS, concurrence of integrations and histone modification marks [124], proximity of CpG island, DNAse hypersensitive sites, fragile sites, LINEs, SINEs and other repeat elements, gene density and frequency of in-gene insertions. These categories obviously reflect our current knowledge of functional genomic regions. Importantly, noncoding DNA sequences such as miRNAs and "intergenic" regions may be functionally more important than currently appreciated.

These generic genomic integration features have been described for a variety of retroviruses, including murine leukemia virus (MLV) [54], xenotropic murine leukemia virus-related virus (XMRV) [125], mouse mammary tumor virus (MMTV) [126], human endogenous retrovirus (HERV) [127], human immunodeficiency virus (HIV-1) [124,128], simian immunodeficiency virus (SIV) [129], porcine endogenous retrovirus (PERV) [130], Foamy virus [62], avian sarcoma leukosis virus (ASLV) [54], and human T-lymphotropic virus (HTLV) [131], and also DNA transposons [55,63]. For the latter, it is again important to avoid generalisation of findings made with the use of a given transposon, since integration properties are dependent on the nature of the transposase, which can be further engineered to increase activity and target specificity [132–136]. Preferential targeting of cellular sequences has been described for another plasmid-based integration vector based on the PhiC31 recombinase. However, in this case the anticipated site-specificity is relative rather than absolute [137], and two studies have revealed potential induction of chromosomal translocations by this system [138,139]. Similar genomic instabilities have not been associated with retrovirus-based integrases or transposases, but certainly require further investigation especially since they may escape conventional approaches of integrome analyses. Comprehensive integrome studies will also be required to assess the specificity of designer zinc-finger nucleases and meganucleases, which have been developed to mediate sequence-specific double-strand breaks to trigger loss-of-function by non-homologous end joining or to repair genes by homologous recombination with donor DNA [140,141].

A second type of analysis is usually associated with gene therapy experiments or functional oncogenomics, where the focus is less on generic insertion patterns but rather on the identity of the genes that are in the vicinity of insertions, potential gene disruption or deregulation and the consequences for the engraftment of a clone after transplantation, its survival, competitive fitness and malignant transformation. This usually involves annotation of gene sets using tools that are also commonly used in microarray analysis, such as pathway analysis (KEGG, PantherDB, Ingenuity) or cross-referencing with cancer gene lists (RTCGD, cancer gene database of human cancer genes [CGC]) or miRNA gene databases (miRbase).

However, the integration of a vector with strong enhancer elements close to a cellular gene does not necessarily trigger its upregulation. Even when using a gammaretroviral vector with strong enhancer sequences in the LTR for genetic modification of T cells, gene activation occurred only in a subset of analyzed clones (6% of the genes tested in 43 T cell clones isolated from peripheral blood of ADA-SCID patients harvested 10–30 months after transplantation of gene-modified hematopoietic stem/progenitor cells) [75]. In comparison, another study reported significant deregulation for 18% of gammaretrovirally targeted genes (8 of 44) analyzed after infusion of suicide gene-modified donor lymphocytes [71] whereas in a study addressing the impact of lentiviral vectors with an internal erythrocyte-specific enhancer-promoter on gene expression in primary erythroblasts the incidence of target gene deregulation was as high as 28% [86].

Moalic et al. [130] noticed that the relationship between MLV and PERV insertions, and the expression level of proximal genes disappears when stratified for CpG proximity, which leads to the conclusion that regulatory sequences show correlation with targeted genes. In turn, transcription factor binding sites present in the LTRs affect the distribution of vector insertions [49]. These variables contribute to the notion that the host cell type influences both the integration pattern and the potential functional consequences of gene deregulation.

The availability of large numbers of documented insertion sites has necessitated a more thorough look into the statistics concerned with background levels of certain features. For example, Ambrosi et al. [142] showed that when analyzing the promoter proximity of MLV integrations, one cannot simply assume the background to be a certain global occurrence rate, since randomly drawn insertion might already show a similar pattern of distribution to the one usually associated with MLV. A more robust test was introduced to better analyze the differences in TSS distribution between different vectors.

When analyzing large datasets, automation becomes a key issue. Manual alignment and evaluation of insertion sites may reveal important details [100,143], but is both cumbersome and prone to human error, while automated analysis allows uniform parameters to be set and applied in different analyses. Several online tools have been developed (Seqmap [144], Quickmap [145]). A particularly exhaustive analysis of the physical features of insertion sites is Quickmap available at GTSG.org [145], allowing large numbers of insertions to be rapidly annotated.

5. Perspectives

Rapid scientific and technological progress has shed a lot of light into the dark side of untargeted vector integration. New approaches to unbiased and highly sensitive determination of integromics have arrived just in time to provide a high-resolution description of stable polyclonality, clonal fluctuation or clonal imbalance following the infusion of gene-modified hematopoietic cells. Evidence has been obtained for a number of approaches to potentially reduce the genomic risk of gene therapy: targeting cells that lack sus-

tained replicative potential, using vectors with a more neutral or targeted integration spectrum, avoiding multiple vector copies in a single cell, and designing gene expression cassettes that avoid long-distance enhancer interactions or fusion transcripts are among the most prominent approaches. It thus appears likely that polyclonal gene-modified hematopoiesis can be established, resulting in a very long latency before clonal dominance and potential malignant progression will occur.

In future studies, more emphasis should be placed on the understanding of the systemic forces that contribute to clonal selection. In this context, important points to address include the potential suppression of clonal imbalance by a high number of fit competitor cells, and the impact of milieu stress that may result from co-morbidity. Progress is also expected in the development of gene targeting approaches or the establishment of episomally persisting vectors. The genomic risk of gene therapy can thus be prevented by the collective tackling of all contributing factors.

Conflict of interest

None.

Acknowledgements

The authors are grateful for the support of their work by the DFG (cluster of excellence REBIRTH, research priority program SPP1230), the BMBF (collaborative project iGene) and the EU (collaborative project PERSIST). The authors thank Tamaryin Godinho for critical reading of the manuscript.

References

[1] Donsante A, Miller DG, Li Y, Vogler C, Brunt EM, Russell DW, et al. AAV vector integration sites in mouse hepatocellular carcinoma. Science 2007;317(5837):477.
[2] Baum C, Dullmann J, Li Z, Fehse B, Meyer J, Williams DA, et al. Side effects of retroviral gene transfer into hematopoietic stem cells. Blood 2003;101:2099–114.
[3] Baum C, Kustikova O, Modlich U, Li Z, Fehse B. Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. Hum Gene Ther 2006;17(3):253–63.
[4] Fehse B, Roeder I. Insertional mutagenesis and clonal dominance: biological and statistical considerations. Gene Ther 2008;15(2):143–53.
[5] Dunbar CE. Stem cell gene transfer: insights into integration and hematopoiesis from primate genetic marking studies. Ann NY Acad Sci 2005;1044:178–82.
[6] Nienhuis AW, Dunbar CE, Sorrentino BP. Genotoxicity of retroviral integration in hematopoietic cells. Mol Ther 2006;13(6):1031–49.
[7] Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, 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(4):401–9.
[8] Li Z, Dullmann J, Schiedlmeier B, Schmidt M, von Kalle C, Meyer J, et al. Murine leukemia induced by retroviral gene marking. Science 2002;296:497.
[9] Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 2003;302:415–9.
[10] Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM. Gene therapy: therapeutic gene causing lymphoma. Nature 2006;440:1123.
[11] Zhang X, Beard BC, Trobridge GD, Wood BL, Sale GE, Sud R, et al. High incidence of leukemia in large animals after stem cell gene therapy with a HOXB4-expressing retroviral vector. J Clin Invest 2008;118(4):1502–10.
[12] Seggewiss R, Pittaluga S, Adler RL, Guenaga FJ, Ferguson C, Pilz IH, et al. Acute myeloid leukemia is associated with retroviral gene transfer to hematopoietic progenitor cells in a rhesus macaque. Blood 2006;107(10):3865–7.
[13] Kustikova OS, Fehse B, Modlich U, Yang M, Düllmann J, Kamino K, et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science 2005;308:1171–4.
[14] Calmels B, Ferguson C, Laukkanen MO, Adler R, Faulhaber M, Kim HJ, et al. Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. Blood 2005;106(7):2530–3.
[15] Deichmann A, Hacein-Bey-Abina S, Schmidt M, Garrigue A, Brugman MH, Hu J, et al. Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. J Clin Invest 2007;117(8):2225–32.
[16] Schwarzwaelder K, Howe SJ, Schmidt M, Brugman MH, Deichmann A, Glimm H, et al. Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. J Clin Invest 2007;117(8):2241–9.
[17] Themis M, Waddington SN, Schmidt M, von Kalle C, Wang Y, Al-Allaf F, et al. Oncogenesis following delivery of a nonprimate lentiviral gene therapy vector to fetal and neonatal mice. Mol Ther 2005;12:763–71.
[18] Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 2008;118(9):3132–42.
[19] Cartier N, Hacein-Bey-Abina S, Bartholomae CC, Veres G, Schmidt M, Kutschera I, et al. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. Science 2009;326(5954):818–23.
[20] Aiuti A, Cattaneo F, Galimberti S, Benninghoff U, Cassani B, Callegaro L, et al. Gene therapy for immunodeficiency due to adenosine deaminase deficiency. N Engl J Med 2009;360(5):447–58.
[21] Bainbridge JWB, Smith AJ, Barker SS, Robbie S, Henderson R, Balaggan K, et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. N Engl J Med 2008;358(21):2231–9.
[22] Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, et al. Cancer regression in patients after transfer of genetically engineered lymphocytes. Science 2006;314(5796):126–9.
[23] Björgvinsdóttir H, Ding C, Pech N, Gifford MA, Li LL, Dinauer MC. Retroviral-mediated gene transfer of gp91phox into bone marrow cells rescues defect in host defense against *Aspergillus fumigatus* in murine X-linked chronic granulomatous disease. Blood 1997;89(1):41–8.
[24] Dinauer MC, Li LL, Björgvinsdóttir H, Ding C, Pech N. Long-term correction of phagocyte NADPH oxidase activity by retroviral-mediated gene transfer in murine X-linked chronic granulomatous disease. Blood 1999;94(3):914–22.
[25] Sadat MA, Pech N, Saulnier S, Leroy BA, Hossle JP, Grez M, et al. Long-term high-level reconstitution of NADPH oxidase activity in murine X-linked chronic granulomatous disease using a bicistronic vector expressing gp91phox and a Delta LNGFR cell surface marker. Hum Gene Ther 2003;14(7):651–66.
[26] Sadat MA, Dirscherl S, Sastry L, Dantzer J, Pech N, Griffin S, et al. Retroviral vector integration in post-transplant hematopoiesis in mice conditioned with either submyeloablative or ablative irradiation. Gene Ther 2009;16(12):1452–64.
[27] Brenner S, Ryser MF, Choi U, Whiting-Theobald N, Kuhlisch E, Linton G, et al. Polyclonal long-term MFGS-gp91phox marking in rhesus macaques after nonmyeloablative transplantation with transduced autologous peripheral blood progenitor cells. Mol Ther 2006;14(2):202–11.
[28] Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science 2003;302(5644):415–9.
[29] Schmidt M, Hacein-Bey-Abina S, Wissler M, Carlier F, Lim A, Prinz C, 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;105:2699–706.
[30] Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J, et al. Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. Lancet 2004;364:2181–7.
[31] Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempski H, et al. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. J Clin Invest 2008;118(9):3143–50.
[32] 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. Science 2002;296:2410–3.
[33] Aiuti A, Cassani B, Andolfi G, Mirolo M, Biasco L, Recchia A, et al. Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. J Clin Invest 2007;117(8):2233–40.
[34] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57–70.
[35] Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, Kinner A, et al. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med 2010;16(2):198–204.
[36] Li Z, Beutel G, Rhein M, Meyer J, Koenecke C, Neumann T, et al. High-affinity neurotrophin receptors and ligands promote leukemogenesis. Blood 2009;113(9):2028–37.
[37] Modlich U, Schambach A, Brugman MH, Wicke DC, Knoess S, Li Z, et al. Leukemia induction after a single retroviral vector insertion in Evi1 or Prdm16. Leukemia 2008;22(8):1519–28.
[38] Kuramoto K, Follman D, Hematti P, Sellers S, Laukkanen MO, Seggewiss R, et al. The impact of low-dose busulfan on clonal dynamics in nonhuman primates. Blood 2004;104(5):1273–80.
[39] Davé UP, Akagi K, Tripathi R, Cleveland SM, Thompson MA, Yi M, et al. Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy. PLoS Genet 2009;5(5):e1000491.
[40] Pike-Overzet K, de Ridder D, Weerkamp F, Baert MRM, Verstegen MM, Brugman MH, et al. Gene therapy: is IL2RG oncogenic in T-cell development? Nature 2006;443(7109):E5.
[41] Thrasher AJ, Gaspar HB, Baum C, Modlich U, Schambach A, Candotti F, et al. Gene therapy: X-SCID transgene leukaemogenicity. Nature 2006;443(7109):E5–6, discussion E6–7.
[42] Scobie L, Hector RD, Grant L, Bell M, Nielsen AA, Meikle S, et al. A novel model of SCID-X1 reconstitution reveals predisposition to retrovirus-

induced lymphoma but no evidence of gammaC gene oncogenicity. Mol Ther 2009;17(6):1031–8.
[43] Ginn SL, Liao SHY, Dane AP, Hu M, Hyman J, Finnie JW, et al. Lymphomagenesis in SCID-X1 mice following lentivirus-mediated phenotype correction independent of insertional mutagenesis and gammac overexpression. Mol Ther 2010;18(5):965–76.
[44] Wicke DC, Meyer J, Buesche G, Heckl D, Kreipe H, Li Z, et al. Gene therapy of MPL deficiency: challenging balance between leukemia and pancytopenia. Mol Ther 2010;18(2):343–52.
[45] Quang CT, Wessely O, Pironin M, Beug H, Ghysdael J. Cooperation of Spi-1/PU.1 with an activated erythropoietin receptor inhibits apoptosis and Epo-dependent differentiation in primary erythroblasts and induces their Kit ligand-dependent proliferation. EMBO J 1997;16(18): 5639–53.
[46] Modlich U, Kustikova OS, Schmidt M, Rudolph C, Meyer J, Li Z, et al. Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. Blood 2005;105: 4235–46.
[47] Montini E, Cesana D, Schmidt M, Sanvito F, Bartholomae CC, Ranzani M, et al. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J Clin Invest 2009;119(4):964–75.
[48] Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, et al. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS ONE 2007;2(12):e1340.
[49] Felice B, Cattoglio C, Cittaro D, Testa A, Miccio A, Ferrari G, et al. Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS ONE 2009;4(2):e4571.
[50] Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. Science 2003;300:1749–51.
[51] Cattoglio C, Facchini G, Sartori D, Antonelli A, Miccio A, Cassani B, et al. Hot spots of retroviral integration in human CD34+ hematopoietic cells. Blood 2007;110(6):1770–8.
[52] Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, et al. Retroviral DNA integration: viral and cellular determinants of target-site selection. PLoS Pathog 2006;2(6):e60.
[53] Beard BC, Dickerson D, Beebe K, Gooch C, Fletcher J, Okbinoglu T, et al. Comparison of HIV-derived lentiviral and MLV-based gammaretroviral vector integration sites in primate repopulating cells. Mol Ther 2007;15(7):1356–65.
[54] Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2004;2:E234.
[55] Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA. High-resolution genome-wide mapping of transposon integration in mammals. Mol Cell Biol 2005;25(6):2085–94.
[56] Hu J, Renaud G, Gomes TJ, Golmes T, Ferris A, Hendrie PC, et al. Reduced genotoxicity of avian sarcoma leukosis virus vectors in rhesus long-term repopulating cells compared to standard murine retrovirus vectors. Mol Ther 2008;16(9):1617–23.
[57] Trobridge GD, Miller DG, Jacobs MA, Allen JM, Kiem HP, Kaul R, et al. Foamy virus vector integration sites in normal human cells. Proc Natl Acad Sci USA 2006;103(5):1498–503.
[58] Hematti P, Hong BK, Ferguson C, Adler R, Hanawa H, Sellers S, et al. Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. PLoS Biol 2004;2:e423.
[59] 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;10:874–81.
[60] Kustikova OS, Geiger H, Li Z, Brugman MH, Chambers SM, Shaw CA, et al. Retroviral vector insertion sites associated with dominant hematopoietic clones mark "stemness" pathways. Blood 2007;109(5):1897–907.
[61] Laufs S, Guenechea G, Gonzalez-Murillo A, Zsuzsanna Nagy K, Luz Lozano M, del Val C, et al. Lentiviral vector integration sites in human NOD/SCID repopulating cells. J Gene Med 2006;8(10):1197–207.
[62] Beard BC, Keyser KA, Trobridge GD, Peterson LJ, Miller DG, Jacobs M, et al. Unique integration profiles in a canine model of long-term repopulating cells transduced with gammaretrovirus, lentivirus, or foamy virus. Hum Gene Ther 2007;18(5):423–34.
[63] Wilson MH, Coates CJ, George AL. PiggyBac transposon-mediated gene transfer in human cells. Mol Ther 2007;15(1):139–45.
[64] Narezkina A, Taganov KD, Litwin S, Stoyanova R, Hayashi J, Seeger C, et al. Genome-wide analyses of avian sarcoma virus integration sites. J Virol 2004;78(21):11656–63.
[65] Miller DG, Trobridge GD, Petek LM, Jacobs MA, Kaul R, Russell DW. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J Virol 2005;79(17):11434–42.
[66] Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, et al. Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science 2005;308:1171–4.
[67] Calmels B, Ferguson C, Laukkanen MO, Adler R, Faulhaber M, Kim HJ, et al. Recurrent retroviral vector integration at the Mds1/Evi1 locus in nonhuman primate hematopoietic cells. Blood 2005;106:2530–3.
[68] 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;110:521–9.
[69] Schmidt M, Zickler P, Hoffmann G, Haas S, Wissler M, Muessig A, et al. Polyclonal long-term repopulating stem cell clones in a primate model. Blood 2002;100(8):2737–43.
[70] Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 2007;17(8):1186–94.
[71] Recchia A, Bonini C, Magnani Z, Urbinati F, Sartori D, Muraro S, 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 USA 2006;103:1457–62.
[72] Berry C, Hannenhalli S, Leipzig J, Bushman FD. Selection of target sites for mobile DNA integration in the human genome. PLoS Comput Biol 2006;2(11):e157.
[73] Gabriel R, Eckenberg R, Paruzynski A, Bartholomae CC, Nowrouzi A, Arens A, et al. Comprehensive genomic access to vector integration in clinical gene therapy. Nat Med 2009;15(12):1431–6.
[74] Kustikova OS, Schiedlmeier B, Brugman MH, Stahlhut M, Bartels S, Li Z, et al. Cell-intrinsic and vector-related properties cooperate to determine the incidence and consequences of insertional mutagenesis. Mol Ther 2009;17(9):1537–47.
[75] Cassani B, Montini E, Maruggi G, Ambrosi A, Mirolo M, Selleri S, et al. Integration of retroviral vectors induces minor changes in the transcriptional activity of T cells from ADA-SCID patients treated with gene therapy. Blood 2009;114(17):3546–56.
[76] Wang GP, Garrigue A, Ciuffi A, Ronen K, Leipzig J, Berry C, et al. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res 2008;36(9):e49.
[77] Montini E, Cesana D, Schmidt M, Sanvito F, Ponzoni M, Bartholomae C, et al. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol 2006;24:687–96.
[78] Bokhoven M, Stephen SL, Knight S, Gevers EF, Robinson IC, Takeuchi Y, et al. Insertional gene activation by lentiviral and gammaretroviral vectors. J Virol 2009;83(1):283–94.
[79] Modlich U, Navarro S, Zychlinski D, Maetzig T, Knoess S, Brugman MH, et al. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther 2009;17(11):1919–28.
[80] Zychlinski D, Schambach A, Modlich U, Maetzig T, Meyer J, Grassman E, et al. Physiological promoters reduce the genotoxic risk of integrating gene vectors. Mol Ther 2008;16(4):718–25.
[81] Arumugam PI, Higashimoto T, Urbinati F, Modlich U, Nestheide S, Xia P, et al. Genotoxic potential of lineage-specific lentivirus vectors carrying the beta-globin locus control region. Mol Ther 2009;17(11):1929–37.
[82] Evans-Galea MV, Wielgosz MM, Hanawa H, Srivastava DK, Nienhuis AW. Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. Mol Ther 2007;15(4):801–9.
[83] Ramezani A, Hawley TS, Hawley RG. Combinatorial incorporation of enhancer-blocking components of the chicken beta-globin 5′HS4 and human T-cell receptor alpha/delta BEAD-1 insulators in self-inactivating retroviral vectors reduces their genotoxic potential. Stem Cells 2008;26(12):3257–66.
[84] Li CL, Xiong D, Stamatoyannopoulos G, Emery DW. Genomic and functional assays demonstrate reduced gammaretroviral vector genotoxicity associated with use of the cHS4 chromatin insulator. Mol Ther 2009;17(4):716–24.
[85] Desprat R, Bouhassira EE. Gene specificity of suppression of transgene-mediated insertional transcriptional activation by the chicken HS4 insulator. PLoS ONE 2009;4(6):e5956.
[86] Hargrove PW, Kepes S, Hanawa H, Obenauer JC, Pei D, Cheng C, et al. Globin lentiviral vector insertions can perturb the expression of endogenous genes in beta-thalassemic hematopoietic cells. Mol Ther 2008;16(3):525–33.
[87] Maruggi G, Porcellini S, Facchini G, Perna SK, Cattoglio C, Sartori D, et al. Transcriptional enhancers induce insertional gene deregulation independently from the vector type and design. Mol Ther 2009;17(5):851–6.
[88] Gonzalez-Murillo A, Lozano ML, Montini E, Bueren JA, Guenechea G. Unaltered repopulation properties of mouse hematopoietic stem cells transduced with lentiviral vectors. Blood 2008;112(8):3138–47.
[89] Cornils K, Lange C, Schambach A, Brugman MH, Nowak R, Lioznov M, et al. Stem cell marking with promotor-deprived self-inactivating retroviral vectors does not lead to induced clonal imbalance. Mol Ther 2008;17(1):131–43.
[90] Newrzela S, Cornils K, Li Z, Baum C, Brugman MH, Hartmann M, et al. Resistance of mature T cells to oncogene transformation. Blood 2008;112(6):2278–86.
[91] Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. Genetics 1988;120(3):621–3.
[92] Dunbar CE, Young NS. Gene marking and gene therapy directed at primary hematopoietic cells. Curr Opin Hematol 1996;3(6):430–7.
[93] Devon RS, Porteous DJ, Brookes AJ. Splinkerettes—improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res 1995;23(9):1644–5.
[94] Eggert H, Bergemann K, Saumweber H. Molecular screening for P-element insertions in a large genomic region of *Drosophila melanogaster* using polymerase chain reaction mediated by the vectorette. Genetics 1998;149(3):1427–34.
[95] Tolar J, Osborn M, Bell S, McElmurry R, Xia L, Riddle M, et al. Real-time in vivo imaging of stem cells following transgenesis by transposition. Mol Ther 2005;12(1):42–8.
[96] Mueller PR, Wold B. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 1989;246(4931):780–6.

[97] Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD. Genomic sequencing and methylation analysis by ligation mediated PCR. Science 1989;246(4931):810–3.
[98] Izsvák Z, Ivics Z. Two-stage ligation-mediated PCR enhances the detection of integrated transgenic DNA. BioTechniques 1993;15(5):814–8.
[99] Schmidt M, Hoffmann G, Wissler M, Lemke N, Mussig A, Glimm H, et al. Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. Hum Gene Ther 2001;12:743–9.
[100] Kustikova OS, Modlich U, Fehse B. Retroviral insertion site analysis in dominant haematopoietic clones. Methods Mol Biol 2009;506:373–90.
[101] Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K, Ball C, Pilz I, et al. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods 2007;4(12):1051–7.
[102] Bozorgmehr F, Laufs S, Sellers SE, Roeder I, Zeller WJ, Zeller WJ, et al. No evidence of clonal dominance in primates up to 4 years following transplantation of multidrug resistance 1 retrovirally transduced long-term repopulating cells. Stem Cells 2007;25(10):2610–8.
[103] Bleier S, Maier P, Allgayer H, Wenz F, Zeller WJ, Fruehauf S, et al. Multiple displacement amplification enables large-scale clonal analysis following retroviral gene therapy. J Virol 2008;82(5):2448–55.
[104] Akagi K, Suzuki T, Stephens RM, Jenkins NA, Copeland NG. RTCGD: retroviral tagged cancer gene database. Nucleic Acids Res 2004;32:D523–7.
[105] Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods 2008;5(3):235–7.
[106] Pule MA, Rousseau A, Vera J, Heslop HE, Brenner MK, Vanin EF. Flanking-sequence exponential anchored-polymerase chain reaction amplification: a sensitive and highly specific method for detecting retroviral integrant–host-junction sequences. Cytotherapy 2008;10(5):526–39.
[107] Vilen H, Aalto J, Kassinen A, Paulin L, Savilahti H. A direct transposon insertion tool for modification and functional analysis of viral genomes. J Virol 2003;77(1):123–34.
[108] Nagy KZ, Laufs S, Gentner B, Naundorf S, Kuehlcke K, Topaly J, et al. Clonal analysis of individual marrow-repopulating cells after experimental peripheral blood progenitor cell transplantation. Stem Cells 2004;22(4):570–9.
[109] Willenbrock H, Salomon J, Søkilde R, Barken KB, Hansen TN, Nielsen FC, et al. Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. RNA 2009;15(11):2028–34.
[110] Justice MJ, Morse HC3, Jenkins NA, Copeland NG. Identification of Evi-3, a novel common site of retroviral integration in mouse AKXD B-cell lymphomas. J Virol 1994;68(3):1293–300.
[111] Joosten M, Vankan-Berkhoudt Y, Tas M, Lunghi M, Jenniskens Y, Parganas E, et al. Large-scale identification of novel potential disease loci in mouse leukemia applying an improved strategy for cloning common virus integration sites. Oncogene 2002;21:7247–55.
[112] Martín-Hernández J, Sørensen AB, Pedersen FS. Murine leukemia virus proviral insertions between the N-ras and unr genes in B-cell lymphoma DNA affect the expression of N-ras only. J Virol 2001;75(23):11907–12.
[113] Lund AH, Turner G, Trubetskoy A, Verhoeven E, Wientjens E, Hulsman D, et al. Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. Nat Genet 2002;32:160–5.
[114] Bijl J, Sauvageau M, Thompson A, Sauvageau G. High incidence of proviral integrations in the Hoxa locus in a new model of E2a-PBX1-induced B-cell leukemia. Genes Dev 2005;19(2):224–33.
[115] Hwang HC, Martins CP, Bronkhorst Y, Randel E, Berns A, Fero M, et al. Identification of oncogenes collaborating with p27Kip1 loss by insertional mutagenesis and high-throughput insertion site analysis. Proc Natl Acad Sci USA 2002;99:11293–8.
[116] Castilla LH, Perrat P, Martinez NJ, Landrette SF, Keys R, Oikemus S, et al. Identification of genes that synergize with Cbfb-MYH11 in the pathogenesis of acute myeloid leukemia. Proc Natl Acad Sci USA 2004;101(14):4924–9.
[117] Iwasaki M, Kuwata T, Yamazaki Y, Jenkins NA, Copeland NG, Osato M, et al. Identification of cooperative genes for NUP98-HOXA9 in myeloid leukemogenesis using a mouse model. Blood 2005;105(2):784–93.
[118] Uren AG, Mikkers H, Kool J, van der Weyden L, Lund AH, Wilson CH, et al. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. Nat Protoc 2009;4(5):789–98.
[119] Pike-Overzet K, de Ridder D, Weerkamp F, Baert MRM, Verstegen MMA, Brugman MH, et al. Ectopic retroviral expression of LMO2, but not IL2Rgamma, blocks human T-cell development from CD34+ cells: implications for leukemogenesis in gene therapy. Leukemia 2007;21(4):754–63.
[120] Suzuki T, Shen H, Akagi K, Morse HC, Malley JD, Naiman DQ, et al. New genes involved in cancer identified by retroviral tagging. Nat Genet 2002;32:166–74.
[121] Abel U, Deichmann A, Bartholomae C, Schwarzwaelder K, Glimm H, Howe S, et al. Real-time definition of non-randomness in the distribution of genomic events. PLoS ONE 2007;2(6):e570.
[122] de Ridder J, Uren A, Kool J, Reinders M, Wessels L. Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. PLoS Comput Biol 2006;2:e166.
[123] de Ridder J, Kool J, Uren A, Bot J, Wessels L, Reinders M. Co-occurrence analysis of insertional mutagenesis data reveals cooperating oncogenes. Bioinformatics 2007;23(13):i133–41.
[124] Wang GP, Levine BL, Binder GK, Berry CC, Malani N, McGarrity G, et al. Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol Ther 2009;17(5):844–50.
[125] Kim S, Kim N, Dong B, Boren D, Lee SA, Das Gupta J, et al. Integration site preference of xenotropic murine leukemia virus-related virus, a new human retrovirus associated with prostate cancer. J Virol 2008;82(20):9964–77.
[126] Faschinger A, Rouault F, Sollner J, Lukas A, Salmons B, Günzburg WH, et al. Mouse mammary tumor virus integration site selection in human and mouse genomes. J Virol 2008;82(3):1360–7.
[127] Brady T, Lee YN, Ronen K, Malani N, Berry CC, Bieniasz PD, et al. Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev 2009;23(5):633–42.
[128] Brady T, Agosto LM, Malani N, Berry CC, O'Doherty U, Bushman F. HIV integration site distributions in resting and activated CD4+ T cells infected in culture. AIDS 2009;23(12):1461–71.
[129] Monse H, Laufs S, Kuate S, Zeller WJ, Fruehauf S, Uberla K. Viral determinants of integration site preferences of simian immunodeficiency virus-based vectors. J Virol 2006;80(16):8145–50.
[130] Moalic Y, Félix H, Takeuchi Y, Jestin A, Blanchard Y. Genome areas with high gene density and CpG island neighborhood strongly attract porcine endogenous retrovirus for integration and favor the formation of hot spots. J Virol 2009;83(4):1920–9.
[131] Meekings KN, Leipzig J, Bushman FD, Taylor GP, Bangham CRM. HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. PLoS Pathog 2008;4(3):e1000027.
[132] Hackett PB, Largaespada DA, Cooper LJN. A transposon and transposase system for human application. Mol Ther 2010;18(4):674–83.
[133] Coates CJ, Kaminski JM, Summers JB, Segal DJ, Miller AD, Kolb AF. Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. Trends Biotechnol 2005;23(8):407–19.
[134] Yant SR, Huang Y, Akache B, Kay MA. Site-directed transposon integration in human cells. Nucleic Acids Res 2007;35(7):e50.
[135] Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z. Targeted Sleeping Beauty transposition in human cells. Mol Ther 2007;15(6):1137–44.
[136] Mátés L, Chuah MKL, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, et al. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet 2009;41(6):753–61.
[137] Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, et al. Integration specificity of phage phiC31 integrase in the human genome. J Mol Biol 2006;357(1):28–48.
[138] Liu J, Jeppesen I, Nielsen K, Jensen TG. Phi c31 integrase induces chromosomal aberrations in primary human fibroblasts. Gene Ther 2006;13(15):1188–90.
[139] Ehrhardt A, Yant SR, Giering JC, Xu H, Engler JA, Kay MA. Somatic integration from an adenoviral hybrid vector into a hot spot in mouse liver results in persistent transgene expression levels in vivo. Mol Ther 2007;15(1):146–56.
[140] Bohne J, Cathomen T. Genotoxicity in gene therapy: an account of vector integration and designer nucleases. Curr Opin Mol Ther 2008;10(3):214–23.
[141] Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, et al. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. Nucleic Acids Res 2009;37(16):5405–19.
[142] Ambrosi A, Cattoglio C, Di Serio C. Retroviral integration process in the human genome: is it really non-random? A new statistical approach. PLoS Comput Biol 2008;4(8):e1000144.
[143] Kustikova OS, Baum C, Fehse B. Retroviral integration site analysis in hematopoietic stem cells. Methods Mol Biol 2008;430:255–67.
[144] Peters B, Dirscherl S, Dantzer J, Nowacki J, Cross S, Li X, et al. Automated analysis of viral integration sites in gene therapy research using the SeqMap web resource. Gene Ther 2008;15(18):1294–8.
[145] Appelt J, Giordano FA, Ecker M, Roeder I, Grund N, Hotz-Wagenblatt A, et al. QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. Gene Ther 2009;16(7):885–93.