MDS1/EVI1 and *PRDM16* deregulation in hematopoiesis after experimental and clinical gene transfer

DISSERTATION

Submitted to

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the Degree of

Doctor of Natural Sciences

Diplom-Biologe Wei Wang

Born in Hebei, China

Heidelberg, 2012

Date of Oral examination

23.03.2012

First referee: **Prof. Dr. Lutz Gissmann**

Division of Genome Modifications and Carcinogenesis

German Cancer Research Center

Second referee: **Prof. Dr. Christof von Kalle**

Division of Translational Oncology

German Cancer Research Center and

National Center for Tumor Diseases, Heidelberg

All life is an experiment

- Ralph Waldo Emerson (1803-1882)

...dedicated to my family

Declarations

I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those expressly indicated.

I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbst verfasst und keine anderen als die von mir bezeichneten Quellen und Hilfsmittel verwendet habe.

Ich erkläre hiermit, dass diese Arbeit weder in dieser noch in einer anderen Form anderweitig als Dissertation oder Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt wurde.

Wei Wang

Heidelberg, 09.02.2012

Publications and conference abstracts

Publications:

Genome wide characterization of insertion profiles of integration deficient lentiviral vectors by two-sided non-restrictive and standard LAM-PCR. (*Manuscript in preparation*)

Wang, W.*, Bartholomae, C. C.*, Gabriel, R., Arens, A., Glimm, H., von Kalle, C., Schmidt, M.

Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation *in vivo.* (*Manuscript in revision*)

Friederike Herbst*, Claudia R. Ball*, Francesca Tuorto, Ali Nowrouzi**, Wei Wang**, Oksana Zavidij, Sebastian M. Dieter, Sylvia Fessler, Franciscus van der Hoeven, Ulrich Kloz, Frank Lyko, Manfred Schmidt, Christof von Kalle and Hanno Glimm

Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces transgene-specific immune tolerance with low genotoxic risk, ***Hepatology*, May;53(5):1696-707. (2011)**

Janka Mátrai*, Alessio Cantore*, Cynthia C. Bartholomae*, Andrea Annoni*, **Wei Wang**, Abel Acosta-Sanchez, Ermira Samara-Kuko, Liesbeth De Waele, Ling Ma, Pietro Genovese, Martina Damo, Anne Arens, Kevin Goudy, Timothy C. Nichols, Christof von Kalle, Marinee K.L. Chuah, Maria Grazia Roncarolo, Manfred Schmidt, Thierry VandenDriessche and Luigi Naldini.

Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nature Protocol* 5, 1379-1395. (2010)

Paruzynski, A*., Arens, A.*, Gabriel, R.*, Bartholomae, C.C.*, Scholz, S., **Wang, W.**, Wolf, S., Glimm, H., Schmidt, M., and von Kalle, C.

Comprehensive genomic access to vector integration in clinical gene therapy, *Nature Medicine* 15(12): 1431-6. (2009)

Gabriel, R.*, Eckenberg, R.*, Paruzynski, A.*, Bartholomae, C. C.*, Nowrouzi, A., Arens, A., Howe, S. J., Recchia, A., Cattoglio, C., **Wang, W.**, Faber, K., Schwarzwaelder, K., Kirsten, R., Deichmann, A., Ball, C. R., Balaggan, K. S., Yanez-Munoz, R. J., Ali, R. R., Gaspar, H. B., Biasco, L., Aiuti, A., Cesana, D., Montini, E., Naldini, L., Cohen-Haguenauer, O., Mavilio, F., Thrasher, A. J., Glimm, H., von Kalle, C., Saurin, W., Schmidt, M.

Conference abstracts:

Genome Wide Characterization of Insertion Profiles of Integration Deficient Lentiviral Vectors by Two-Sided non-Restrictive and Standard LAM-PCR

Wang, W., Bartholomae, C. C., Gabriel, R., Arens, A., Glimm, H., von Kalle, C., Schmidt, M., 13th Annual Meeting of the American Society of Gene and Cell Therapy, Washington DC, USA (2010)

Comprehensive Characterization of Integrase Deficient Lentiviral Vector Integration Profiles in Vitro

Wang, W., Bartholomae, C. C., Gabriel, R., Arens, A., Glimm, H., von Kalle, C., Schmidt, M. **,** 17th Annual Congress of the European Society of Gene and Cell Therapy, Hannover, Germany (2009)

Acknowledgements

First and foremost, I would like to express the deepest appreciation to Prof. Dr. Christof von Kalle, head of division Translational Oncology for offering me the opportunity to pursue this PhD study in his division, for his excellent guidance and constant support during the past four years, and for the critical reading of my dissertation.

The accomplishment of this thesis would not have been possible without the insightful guidance from Dr. Manfred Schmidt and Prof. Dr. Hanno Glimm. I will always be grateful for their support, wisdom, and kindness.

My sincere thanks also go to Prof. Dr. Lutz Gissmann for kindly being my first supervisor and to Prof. Hartmut Goldschmidt for acting as a member of my thesis advisory committee. I feel very grateful to both of them for participating actively in my progress reports. Furthermore it is an honor for me to have PD. Dr. Rainer König as chairman, and Prof. Dr. Valerie Bosch participate in my thesis defense.

Dr. Cynthia Bartholomä, Dr. Claudia Ball, Dr. Kerstin Schwarzwälder and Dr. Annette Deichmann provided a lot of guidance and assistance through this dissertation, I will never forget their kindness and enthusiasm. Dr. Richard Gabriel, Dr. Ali Nowrouzi, Dr. Anna Paruzynski, Dr. Friederike Herbst, Dr. Oksana Zavidij gave their best ideas in the experimental design, shared their profound expertise in cancer research and their fruitful discussion during the project proceeding. Anne Arens and Prof. Ulrich Abel wrote the bioinformatics programs and helped me a lot for analyzing the high-throughput sequencing data. With Christine Käppel, Eliana Ruggiero and Simone Scholz together, they provided a stimulating and fun environment in which to learn and grow. I would also like to thank Sylvia Fessler, Ina Kutschera, Christina Lulay and Christian Weber in our laboratory for their wonderful technical assistances.

Last but not least, I must thank my parents and my wife Yi Pan, for their generous supports and love. To them I dedicate this thesis.

Table of Contents

Figures List

Tables List

Abbreviations

µl	Microlitre
µM	Micromolar
3’	3 prime
5’	5 prime
AAV	Adeno-associated virus
ADA	Adenosine deaminase
APC	Antigen presenting cell
ASLV	Avian sarcoma-leukaemia virus
B	Biotin
Bio	Biotinylated
BLAT	BLAST like alignment tool
BM	Bone marrow
bp	Base pair
BSA	Bovine serum albumin
C	Concentration
CD	Cluster of differentiation
CDKN2A	Cyclin-dependent kinase 2A
cDNA	Complimentary DNA
CFCs	Colony-forming cells
CGD	Chronic granulomtous disease
Chr	Chromosome
CIS	Common integrationsite
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CMV	Cytomegalovirus
CpG	Cytosine and guanine separated by a phosphate
cPPT	Central polypurine tract
DC	Dendritic cell
ddH_2O	Distilled water
DKFZ	Deutsches Krebsforschungszentrum
DMEM	Dulbecco’s Modified Eagle Medium
DMF	Dimethylformamid
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
ds	Double-stranded DNA
EB	Ethidium Bromide
EDTA	Ethylenediaminetetraacetic acid
EF1α	Elongation Factor 1α
eGFP	Enhanced green fluorescent protein
EIAV	Equine Infectious Anaemia Virus
ELISA	Enzyme-linked immunosorbent assay
emPCR	emulsion-based PCR
Env	Envelope
Expo	Exponential
eYFP	Enhanced yellow fluorescent protein
FACs	Fluorescence-activated cell sorter
FBS	Fetal bovine serum

FCS	Fetal calf serum
FP	Flanking primer
FuP	Fusion primer
G	Granulocyte
Gag	Group specific antigens HBS HEPES buffered saline
GMP	Granulocyte/macrophage precursors
h	hour
HBSS	Hank's Buffered Salt Solution
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFV	Human foamy virus
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HPC	Haematopoietic progenitor cell
HPLC	High pressure liquid chromatography
HSC	Hematopoietic stem cell
HSV	Herpes simplex virus
IC	Internal control
ICLV	Integration-competent lentiviral vectors
IDLV	Integration-deficient lentiviral vectors
IL	Interleukin
IMDM	Iscove's Modified Dulbecco's Medium
IN	Integrase
IPLV	Integrating lentiviral vector
IS	Integration site
kb	Kilobases
kDa	Kilo Dalton
LAM-PCR	Linear amplification mediated polymerase chain reaction
LB	Luria-Bertani
LC	Linker cassette
LC480	LightCycler®480
LiCl	Lithium chloride
LT-HSC	Long-term HSC
LTR	Long terminal repeat
LV	Lentiviral vector
M	Molar
m	milli (10^{-3})
MC	Monocytes
MDS	Myelodysplasia
MEP	Megakaryocyte/erythroid precursor
MFI	Mean fluorescence intensity
Min	Minute
MkP	Megakaryocyte precursor
ml	Millilitre
MLV	Murine leukaemia virus
MOI	Multiplicity of infection
MoMLV	Moloney murine leukaemia virus
MPC	Magnetic particle concentrator
MPP	Multipotent progenitors
n	nano
NADPH	Nicotinamide dinucleotide phosphate

NCBI	National Center for Biotechnology Information
ng	Nanogram
NK	Natural killer cell
nrLAM-PCR	Non-restrictive LAM-PCR
OD	Optical density
ON	Overnight
P	Patient
p24	Gag HIV-1 p24 gag capsid protein
PB	Peripheral blood
PBCS	Peripheral blood stem cells
PBL	Peripheral blood leukocytes
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Phox	Phagocyte oxidase
PIC	Pre-integration complex
PID	Primary immunodeficiency
PPT	Polypurine tract
qPCR	Quantitative real time PCR
R	Redundant
RAG	Recombination activating genes
RCR	Replication competent retrovirus
RefSeq	Reference Sequence
RIS	Retroviral integration site
RNA	Ribonucleic acid
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute medium
RRE	Rev-response element
RSV	Rous sarcoma virus
RT	Reverse transcriptase
RV	Retroviral vector
S	Second
SA	Splice acceptor
SCF	Stem cell factor
SCID	Severe combined immunodeficiency
SCID-X1	X-linked severe combined immunodeficiency
SD	Splice donor
SFFV	Spleen focus forming virus
SFV	Semiliki-forest virus
SIN	Self-inactivating
ss	Single stranded
ST-HSC	Short-term HSC
T cell	Thymus derived lymphocyte
TAE	Tris-acetate-EDTA
T-ALL	Acute T lymphoblastic leukaemia
Taq	Thermus aquaticus
TBE	Tris-borate-EDTA
TCR	T cell receptor
Tpo	Thrombopoeitin
TSS	Transcription start site

U	Unit
U3	Unique 3
U5	Unique 5
UCSC	University of California, Santa Cruz
v/v	Volume per volume
VSV-G	Vesicular stomatitis virus glycoprotein
w/v	weight per volume
w/w	weight per weight
WAS	Wiskott - Aldrich syndrome
WPRE	Woodchuck hepatitis virus post-transcriptional regulatory element
ZFN	zinc finger nucleases
α	Alpha
β	Beta
γc	Gamma chain
μ	micro (10^{-6})
Ψ	Packaging signal

Summary

Comprehensive integration site analysis in clinical gene therapy has shown that therapeutic vector integrations can lead to clonal selection and even malignant transformation by transcriptional activation of cellular proto-oncogenes. Assessing the clonal dynamics of gene corrected cells revealed an expansion of hematopoiesis marked with unique retroviral insertion sites mainly in the loci of *MDS1/EVI1*, *PRDM16* and *SETBP1* in two patients treated for correction of chronic granulomatous disease (CGD) 5 months after gene therapy. This cell population remained stable for additional 1.5 years. At later time points, both individuals developed a myelodysplastic syndrome (MDS) driven by gene-corrected cell clones carrying integrations in the *MDS1/EVI1* locus.

In this thesis, large scale integration site mapping using non-restrictive and highly sensitive linear amplification mediated polymerase chain reaction (LAM-PCR) focused on the gamma-retroviral integration monitoring within the two gene loci *MDS1/EVI1* and *PRDM16* in Wiskott - Aldrich syndrome (WAS) clinical gene therapy. Both WAS patients experienced substantial improvement in their clinical conditions, showing a polyclonal repopulation of the hematopoietic system after reinfusion of autologous ex vivo transduced hematopoietic progenitor cells. Although *MDS1/EVI1* and *PRDM16* were scored as common integration sites (CIS), and *MDS1/EVI1* and *PRDM16* containing clones were mostly restricted to the myeloid compartment of reconstituted hematopoiesis, no signs of clonal expansion related to these two genes were observed in both patients at 2 years follow up.

We hypothesized that a concerted sustained or transient overexpression of *MDS1/EVI1* and *PRDM16* as pro-myelocytic key regulating transcription factors might be used to substantially influence proliferating and differentiation capacity of hematopoietic stem and progenitor cells. Therefore, murine *Mds1/Evi1* and *Prdm16* were cloned and expressed either by integration-proficient (LV) or integration-deficient lentiviral vectors (IDLV). Long-term *Mds1/Evi1* and *Prdm16* expression in hematopoietic progenitors driven by LV was achieved, as well transient expression by IDLV. However, no significant clonal expansion was observed after experimental overexpression of *Mds1/Evi1* and *Prdm16*, as transduced cells became apoptotic after expression.

To further assess the biosafety of IDLVs for clinical gene therapy we performed genome wide large-scale IDLV integration analysis *in vitro*, using conventional and non-restrictive

LAM-PCR, as well as newly established two-directional LAM-PCR. This first ever large scale IDLV integration analyses yielded more than 800 unique, mappable IDLV ISs *in vitro*. Our data revealed a genome-wide close to random integration pattern without any preference for gene coding regions. The results also provided direct molecular evidence that the background integration of D64V mutant IDLVs is not mediated by residual catalytic activity of the mutant integrase. The risk of insertional mutagenesis events mediated by IDLV is highly minimized compared to their integration proficient counterparts and it provides one of the potentially safest tools for efficient gene delivery for clinical applications. The utilization of IDLVs for transient overexpressing is versatile, but warrants careful evaluation of potential toxic effects, as *Mds1/EvI1* and *Prdm16* as chosen transgenes to expand hematopoietic progenitors have shown.

Zusammenfassung

Die umfassende Analyse von Integrationsstellen gentherapeutischer Vektoren in klinischen Studien zeigte, dass die Integration des therapeutischen Vektors zu einer klonalen Selektion bis hin zu einer malignen Transformation durch transkriptionelle Aktivierung von Protoonkogenen führen kann. Untersuchungen zur klonalen Dynamik genkorrigierter Zellen in zwei Patienten im Rahmen der klinischen Gentherapiestudie zur Behandlung septischer Granulomatose (CGD) zeigte 5 Monate nach Therapie eine Expansion der Hämatopoese, die hauptsächlich von Zellklonen mit Vektorintegrationsstellen innerhalb oder in der Nähe der Gene *MDS1/EVI1*, *PRDM16* und *SETBP1* angetrieben wurde. Diese Zellpopulation blieb für weitere 1.5 Jahre stabil. Im späteren Verlauf entwickelten beide Pateinten ein myelodysplastisches Syndrom (MDS), ausgelöst durch genkorrigierte Zellen, welche Integrationen in dem *MDS1-EVI1* Lokus aufwiesen.

Für die Charakterisierung retroviraler Vektorintegrationsstellen im Rahmen der klinischen Studie zur Behandlung des Wiskott-Aldrich-Syndroms (WAS) mit Fokus auf die beiden genomischen Bereiche *MDS1/EVI1* und *PRDM16* wurde in dieser Arbeit die nicht-restriktive und die hochsensitive lineare amplifikations-mediierte Polymerasekettenreaktion (LAM-PCR) verwendet. In beiden WAS Patienten führte die Behandlung zu einer grundlegenden Verbesserung der klinischen Situation. Es zeigte sich eine polyklonale Repopulation des hämatopoetischen Systems nach Reinfusion *ex vivo* transduzierter autologer hämatopoetischer Vorläuferzellen. Obgleich sich sowohl bei *MDS1/EVI1* als auch bei *PRDM16* eine auffällige Häufung der Vektorintegrationsstellen zeigte („common integration sites“, CIS) und diese vorwiegend auf das myeloide System beschränkt waren, wurden in beiden Patienten bis zwei Jahre nach Therapie keine Anzeichen einer klonalen Expansion in Verbindung mit diesen beiden Genloki beobachtet.

Wir stellten die Hypothese auf, dass eine gemeinsame fortwährende oder transiente Überexpression von *MDS1/EVI1* und *PRDM16*, die beide wichtige Transkriptionsfaktoren im myeloiden System darstellen, verwendet werden könnte, um die Proliferations- und Differenzierungskapazität von hämatopoetischen Stamm- und Vorläuferzellen zu beeinflussen. Daher wurden die murinen Genloki *MDS1/EVI1* und *PRDM16* kloniert und mithilfe von Integrations-profizienten (LV) und Integartions-defizienten lentiviralen Vektoren (IDLV) exprimiert. In hämatopoetischen Vorläuferzellen konnte eine Langzeitexpression von

MDS1/EVI1 und *PRDM16* durch LV sowie eine transiente Expression durch IDLV erreicht werden. Es konnte jedoch keine signifikante klonale Selektion beobachtet werden, da die experimentelle Überexpression in den transduzierten Zellen Apoptose auslöste.

Für die klinische Anwendung von IDLV wurde deren biologische Sicherheit weitergehend durch eine umfassende genomweite *in vitro* Integrationsstellenanalyse unter Verwendung der konventionellen und der nicht-restriktiven LAM-PCR, sowie einer neu entwickelten bidirektionalen LAM-PCR, untersucht. Diese im Hochdurchsatz angelegte Studie konnte erstmals mehr als 800 unterschiedliche, dem Genom eindeutig zuordenbare IDLV Integrate *in vitro* identifizieren und charakterisieren. Unsere Daten zeigten eine genomweite beinahe zufällige Verteilung von Integrationsstellen, ohne Präferenz für genkodierende Bereiche. Weiterhin erbrachten diese Daten erstmals den molekularen Beweis, dass die Hintergrund-Integration der D64V IDLV-Mutante nicht durch die residuale katalytische Aktivität der Integrase vermittelt ist. Das Risiko der insertionellen Mutagenese ausgelöst durch IDLV ist im Vergleich zu entsprechenden Integrations-profizienten Vektoren stark minimiert und somit könnte die Verwendung von IDLVs eines der potentiell sichersten Werkzeuge für einen effizienten Gentransfer in der klinische Anwendung darstellen.

Die Verwendung von IDLV zur transienten Überexpression ist vielfältig, bedarf aber einer genauen Abwägung des potentiell toxischen Effekts, wie die beiden zur Untersuchung der Expansion hämatopoetischer Progenitoren ausgewählten Transgene *MDS1/EVI1* und *PRDM16* zeigten.

1. Introduction

1.1 Hematopoietic stem cells and hematopoiesis

1.1.1 Human and mouse hematopoietic stem cells and hematopoiesis

Stem cells are single cells that are capable of both self-renewal and differentiation (Weissman, 2000). The essential meaning of self-renewal is that these cells can undergo division symmetrically or asymmetrically but remain undifferentiated, and they possess properties that allow large numbers of self-renewing cell divisions to occur in a regulated fashion over the life span of a host. The stem cells could be divided into three categories based on their differentiation potential: totipotent, pluripotent and multipotent. Totipotent stem cells like fertilized eggs and each cell of the embryo up to the four-cell stage can form an entire organism including all embryonic, extra-embryonic and adult cell types in a suitable environment. Pluripotent cells can self-renew and differentiate into any cell type of the body, but lack the capacity to independently form a placenta, or to organize the formation of proper primitive streak. Multipotent stem cells are single cells which are found in the tissues of adult body or in the cord blood, like the hematopoietic stem cells (HSCs), which have a restricted differentiation potential, can only self-renew and differentiate into more than one cell type in a particular tissue lineage.

The cellular biology of hematopoiesis encompasses three components: the hematopoietic stem cells and its progeny, the environment in which they reside, and the signal and cytokines that regulate their fate. In adults, hematopoiesis occurs in the bone marrow (BM), which supports the lifelong maintenance of stem cells and their regulated differentiation. The term hematopoiesis (from Greek ‘haima’ for blood and ‘poiein’, to make) refers to the formation and development of the blood system. Hematopoiesis is a dynamic and complex developmental process by which immature precursor cells develop into mature blood cells. Per day, hematopoiesis yields approximately 175 billion red cells, 70 billion granulocytes (neutrophils,

eosinophils, basophils), and 175 billion platelets. If needed, production can be increased 5-10 fold (Orkin, 1995, 1996).

In mice, it has been shown that an early intra-embryonic site of hematopoiesis is found in the paraaortic splanchnopleura and in a structure termed aorta-gonad-mesonephros region (AGM) (Nishikawa et al., 2001). The establishment of blood islands in the extra-embryonic yolk sac marks the onset of hematopoiesis and vasculogenesis in the developing embryo. Subsequently, hematopoietic cells are found within the embryo proper in the gonad-mesonephros region, from where HSCs migrate to the fetal liver and later to the BM. In humans, this process begins in the yolk sac in the first weeks of embryonic development. By the third month of gestation, stem cells migrate to the fetal liver and then to the spleen. Next, the BM becomes the major hematopoietic organ and hematopoiesis ceases in the liver and spleen.

HSCs are by far the most well-characterized adult stem cell population both in terms of markers for purification and assays to access functional potential. In 1950s, researchers demonstrated that just-lethal doses of whole body irradiation caused death by hematopoietic failure, and that hematopoietic failure could be averted either by shielding hematopoietic organs or by transplantation of unirradiated BM (Jacobson et al., 1950; Jacobson et al., 1951; Lorenz et al., 1951). The critical study that set the stage for the concept of HSCs was the demonstration that the injected BM reconstituted the hematopoietic system rather than provided radiation repair factors (Ford et al., 1956; Nowell et al., 1956). The seminal study in HSC field began with the observation that transplantation of limiting numbers of BM cells into lethally irradiated mice gave rise to colonies in the spleen that contained all elements of the myeloerythroid (but not lymphoid) lineages, and that the number of colonies was proportional to the number of BM cells injected (Till and Mc, 1961). Followed these observations, finding of a clonogenic precursor and lymphocytic potential of spleen colony-forming cells (CFCs) likely demonstrated that there existed in the BM single cells capable of self-renewal and full differentiation to all hematopoietic fates (Becker et al., 1963; Siminovitch et al., 1963; Wu et al., 1968).

Subsequently, it was demonstrated that at least two classes of spleen colony-forming cells existed: those that peaked at days 8-10 and were the progeny, mainly, of non-stem cell oligopotent progenitors; and those that peaked in size at days 12-14, which included much more primitive progenitors and stem cells (Magli et al., 1982; Visser and Van Bekkum, 1990). Recently, following the isolation of both stem and progenitor cells, it became clear that most of the day 8-10 spleen colonies are derived from committed myeloid progenitors, whereas the day 12-14 colonies are derived from HSCs, multipotent but self-renewing progenitors, and some late-peaking committed myeloid progenitors (Morrison et al., 1997; Morrison and Weissman, 1994; Na Nakorn et al., 2002; Spangrude et al., 1988).

The study of HSCs has been greatly facilitated in the last 20 years with the discovery of monoclonal antibodies and the development of high-speed multiparameter fluorescence-activated cell sorters. The eventual isolation of HSCs required not only these innovations, but also the development of *in vitro* and *in vivo* assays for the identification and quantification of clonogenic precursors of the all the hematopoietic cell lineages. The identification and purification of HSCs relies on the unique cell surface molecule expression found on these cells compared with the remainder of BM cells including closely related hematopoietic progenitor cell counterparts. Although there does not appear to be any single marker that segregates HSCs from other hematopoietic cell types, HSCs can be readily identified using multiparameter flow cytometry. However, several different antibody combination schemes have been developed by different laboratories to achieve this, although all have the ultimate goal of producing the highest yield of long-term, multilineage reconstituting HSCs. Almost all HSC purification strategies revolve around the cell surface phenotype of positive selection for the markers c-Kit and Sca-1 and negative selection for markers of mature hematopoietic cell lineages (typically B220, CD4, CD8, Gr-1, Mac-1, and Ter-119). Although this c-Kit^{+}Lin^{-}Sca-1^{+} (KLS) phenotype greatly enriches for hematopoietic reconstituting activity, this BM compartment contains progenitor cells in addition to long-term HSCs. In fact only ~ 10% of KLS cells are bona fide long-term HSCs, and as such the KLS compartment

should be regarded as merely enriched for HSCs. A variety of strategies have been used to further enrich BM for HSCs, with or without the KLS as a foundation. These strategies include identification of HSCs as KLS-CD34$^-$Flk-2$^-$, KLS-CD150$^+$CD48$^-$ cells, the Hoechst-effluxing side population (SP), and associated variations on that theme (e.g., CD45^{mid}Lin$^-$ Hoechst^{low}Rhodamine^{low} or SP-EPCR). A summary of murine cell surface phenotypes and the murine hematopoietic cell types they enrich for is presented in Table 1-1.

Marker Phenotype	Cell Types	Reference
KLS	Hematopoietic stem and progenitor cells	(Okada et al., 1992)
SP^{KLS}	LT-HSCs	(Adolfsson et al., 2001)
CD150$^+$CD48$^-$CD41$^-$KLS	LT-HSCs	(Goodell et al., 1996)
CD45midLin-RhodaminelowSP	LT-HSCs	(Dykstra et al., 2006)
Flk-2$^-$CD34$^-$KLS	LT-HSCs	(Yang et al., 2005)
Flk-2$^+$CD34$^+$KLS	ST-HSCs and MPPs	(Yang et al., 2005)
Lin$^-$Il7rα$^+$c-Kit$^+$Sca-1$^+$	Common lymphoid progenitors (CLP)	(Kondo et al., 1997)
Lin$^-$Il7rα$^-$c-Kit$^+$Sca-1$^-$	Myeloid progenitors	(Akashi et al., 2000)
Lin$^-$Il7rα$^-$c-Kit$^+$Sca-1$^-$ CD34$^+$CD16/32$^-$	Common myeloid progenitor (CMP)	(Akashi et al., 2000)
Lin$^-$Il7rα$^-$c-Kit$^+$Sca-1$^-$CD34$^-$ CD16/32$^-$	Megakaryocyte-erythrocyte (MEP)	(Akashi et al., 2000)
Lin$^-$Il7rα$^-$c-Kit$^+$Sca-1$^-$ CD34$^+$CD16/32$^+$	Granulocyte-macrophage progenitors (GMP)	(Akashi et al., 2000)

Table 1-1 Cell surface phenotypes of various murine hematopoietic stem and progenitor cell populations.

The advent of antibody technology, culture capability, transplantation, and finally genetic engineering of mice has enabled a detailed understanding of processes involved in the commitment and differentiation of HSCs into progenitor cells and ultimately mature blood cells. Combining antibody-based subselection of cells with transplantation has made possible the identification of progenitors and HSCs. A range of different methods for characterizing stem and progenitor cell populations are now available for use in hematopoietic cell research. In vivo, long-term, multilineage reconstitution is the gold standard for detection of HSCs. In a mouse model, test donor cells are retrospectively inferred to have contained one or more HSCs when their transplantation into lethally irradiated mice results in long-term multilineage reconstitution. Competitive repopulation is a modified version of

transplantation that ensures the survival of lethally irradiated recipient mice and permits the quantitative evaluation of repopulation levels. In experimental transplantation, test donor cells and competitor cells are distinguished by certain genetic markers. Common used are CD45.1 and CD45.2. CD45 is the leukocyte common antigen expressed on all leukocytes and their progenitor cells, including adult HSCs, but not on red blood cells, platelets, and their committed progenitors. When multilineage reconstitution takes place 4 months after transplantation, this is conceivably due to self-renewal and differentiation potentials in HSCs. The HSC compartment has been shown to be heterogeneous, comprising a hierarchy of HSCs that can be identified by their functional capacity. The most immature HSC in this hierarchy is capable of sustaining hematopoiesis throughout serial transplantation. In this assay, the source of HSCs is transplanted into sequential serial transplant recipients, and the ability of this population to sustain hematopoiesis by presumptive self-renewing division is determined.

In vitro and short term in vivo assays are also often used for measuring HSC and progenitor cell content but are generally a reflection of the more mature progenitor cells as opposed to the immature HSCs. Hematopoietic progenitor cells can grow in semisolid media such as agar and methylcellulose supplemented with growth factors, finally forming colonies. The majority of colony-forming cell (CFC) consist of lineage-restricted colonies: erythroid-restricted burst-forming units-erythroid (BFU-E), which are more immature than the colony-forming units-erythroid (CFU-E); megakaryocyte-restricted CFU-Mk; colony-forming units-granulocytes (CFU-G); and colony-forming-units-granulocytes/macrophages (CFU-GM). The most immature (multipotent) CFC measurable contains granulocytes, erythrocytes, macrophages, and often megakaryocytes (CFU-GEMM) and is usually measured at day 12 after culture initiation. In addition to colony-forming cell assays, short-term in vivo assays are also useful for analysis of HSCs and progenitor cells. As previous descript (Till and Mc, 1961), Colony-forming unit-spleen (CFU-S) cells are cells that, once injected into an irradiated recipient, home to the spleen and form macroscopic colonies that provide very short-term in vivo repopulation of the mouse. The CFU-S are therefore

early engrafting cells, providing radioprotection to the mouse and allowing it to survive more readily in the first 2-3 weeks post transplantation. These progenitors are more immature than CFCs but are more mature than HSCs.

1.1.2 *Hematopoietic stem cells self-renewal and differentiation*

HSCs have several unique features, among the core features are the ability to choose between self-renewal and differentiation. In addition, HSCs migrate in regulated fashion and are subject to regulation by apoptosis. The balance between these activities determines the number of stem cells that are present in the body.

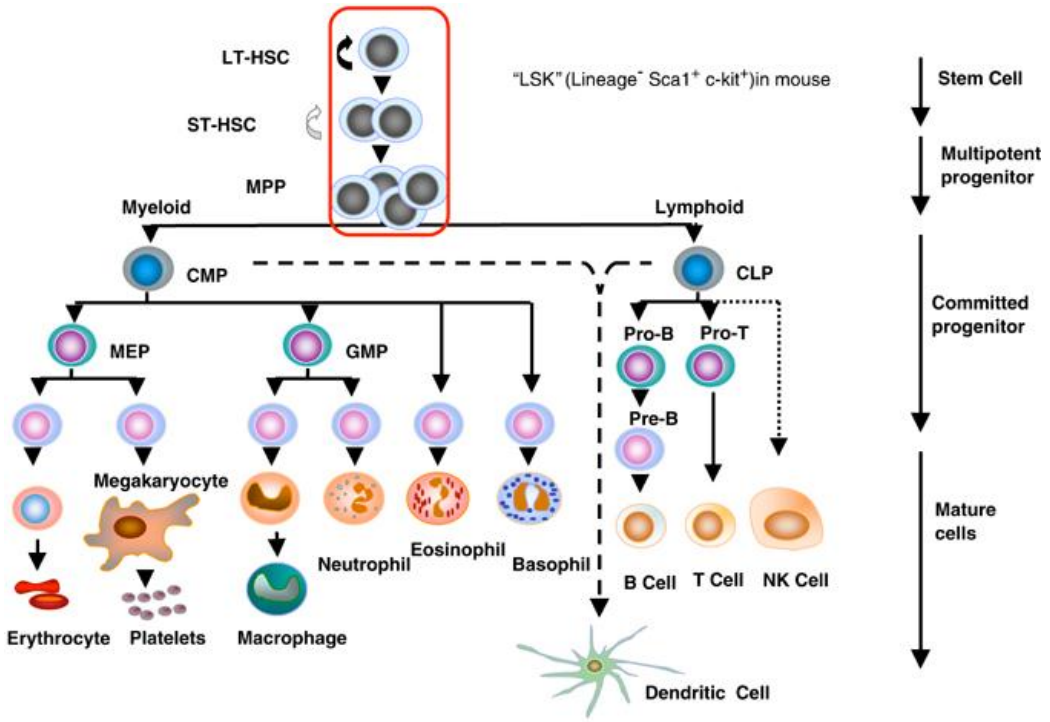

Figure 1-1 The hierarchy of hematopoietic cells. LT-HSC, long-term repopulating HSC; ST-HSC, short-term repopulating HSC; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte/erythroid progenitor; GMP, granulocyte–macrophage progenitor. The encircled pluripotent population, LT-HSC, ST-HSC and MPP are Lin-, Sca-1+, c-kit+ as shown (Larsson and Karlsson, 2005).

In the HSCs pool (Figure 1-1), multipotent long-term HSCs (LT-HSCs) reside in the BM and through a process of asymmetric cell division, can self-renew to sustain the stem cell pool or differentiate into short-term HSCs (ST-HSCs) or lineage-restricted progenitors that undergo extensive proliferation and differentiation to produce terminally differentiated, functional hematopoietic cells. ST-HSCs or multipotent progenitors (MPPs) are only able to sustain hematopoiesis in the short term,

whereas the LT-HSCs must persist for the lifespan of the organism to perpetually replenish the hematopoietic system. In the normal steady state, about 1.5% and up to 8% of LT-HSCs randomly enter the cell cycle each day, whereas higher percentages of ST-HSCs and MPPs are in the cell cycle at any time.

Like other stem cell types, one essential feature of HSCs is the ability to make copies with the same or very similar potential, so called self-renew. This is a single distinguished characteristic between LT-HSCs and the rest of the multipotent and oligopotent progenitors in the hematopoietic system. While it has not been established that adult HSCs can self-renew indefinitely, it is clear from serial transplantation experiments that they can produce enough cells to last several (at least four to five) lifetimes in mice. It is still unclear which key signals allow self-renewal, based on the research of embryonic stem (ES) cells, the genome of stem cells must be protected from DNA damage or mutation. Therefore DNA repair and maintenance of sufficient telomere length are necessary to support a number of self-renewal divisions in HSCs (Allsopp et al., 2003a; Allsopp et al., 2003b; Nijnik et al., 2007; Rossi et al., 2007). Expression of telomerase is associated with self-renewal activity. However, while absence of telomerase reduces the self-renewal capacity of mouse HSCs, forced expression is not sufficient to enable HSCs to be transplanted indefinitely; other barriers must exist.

It has proven surprisingly difficult to grow HSCs in culture despite their ability to self-renew, although expansion in culture is routine with many other stem cells, including ES cells and neural stem cells. The lack of this capacity for HSCs severely limits their application, because the number of HSCs that can be isolated from mobilized blood, umbilical cord blood, or BM restricts the full application of HSC transplantation in human. Engraftment periods of 50 days or more were standard when limited numbers of BM or umbilical cord blood cells were used in a transplant setting, reflecting the low level of HSCs found in these native tissues. Attempts to expand HSCs in tissue culture with known stem-cell stimulators, such as the cytokines stem cell factor (SCF), thrombopoietin (TPO), interleukins 1, 3, 6, 11 plus or minus the myeloerythroid cytokines GM-CSF, G-CSF, M-CSF, and erythropoietin have

never resulted in a significant expansion of HSCs (Reya et al., 2003). Rather, these compounds induce many HSCs into cell divisions that are always accompanied by cellular differentiation. Many experiments demonstrate that the transplantation of a single or few HSCs into an animal results in a 100,000-fold or greater expansion in the numbers of HSCs at the steady state while simultaneously generating daughter cells that permitted the regeneration of full blood-forming system (Iscove and Nawa, 1997; Morrison et al., 1997; Smith et al., 1991; Wagers et al., 2002). By investigating genes transcribed in purified mouse LT-HSCs, investigators have found that these cells contain expressed elements of the Wnt/fzd/beta-catenin signaling pathway, which enables mouse HSCs to undergo self-renewing cell divisions (Reya et al., 2003). Overexpression of several other proteins, including HoxB4 (Antonchuk et al., 2002; Sauvageau et al., 1995) and HoxA9 (Lawrence et al., 2005; Thorsteinsdottir et al., 2002) has also been reported to achieve this. Other signaling pathways are under investigation such as Notch and Sonic hedgehog (Reya, 2003). Among the intracellular proteins thought to be essential for maintaining the stemness state are Polycomb group genes, for instance Bmi-1 (Iwama et al., 2004). Other genes, such as c-Myc (Wilson et al., 2004) and JunB (Passegue et al., 2004) have also been shown to play a role in this process. The identification of bone morphogenetic protein (BMP) pathway (Zhang et al., 2003), parathyroid hormone (PTH) pathway (Calvi et al., 2003), as well as Angiopoietin-1/Tie2 pathway (Arai et al., 2004) in osteoblast cells, which offer the microenvironment as the niche for the HSCs, have been recently suggested to regulate HSCs numbers and quiescence. It is critical to discover which pathways operate in the expansion of human HSCs to take advantage of them to improve clinical hematopoietic transplantation.

Differentiation into progenitors and mature cells that fulfill the functions performed by the hematopoietic system is not a unique HSC property, but, together with the option to self-renew, defines the core functions of HSCs. Differentiation is driven and guided by an intricate network of growth factors and cytokines. Differentiation seems to be the default outcome for HSCs when stimulated by many of the factors to which they have been shown to respond. It appears that, once they

commit to differentiation, HSCs cannot revert to a self-renewing state. Thus, specific signals, provided by specific factors, seem to be needed to maintain HSCs. This strict regulation may reflect the proliferative potential present in HSCs, deregulation of which could easily result in malignant diseases such as leukemia or lymphoma.

1.1.3 *Hematopoietic stem cells expansion and transplantation*

Hematopoietic stem cell transplantation (HSCT) is a well-established and widely performed strategy as therapy for hematologic malignancies as well as many nonmalignant disorders. HSCs are routinely available in sufficient numbers from a variety of sources, including BM and mobilized peripheral blood. However, there are instances when the number of available HSCs is insufficient for engraftment of the host, in particular when cord blood is used as the source of HSCs for transplantation or in the case of gene therapy, where the frequency of genetic modified cells is very low. One potential solution to this problem of low stem cell numbers is *ex vivo* proliferation of the cells prior to transplantation. Extensive research has been done to define the optimal conditions for ex vivo expansion of HSCs, and various expansion techniques have been development for this purpose (Sauvageau et al., 2004; Sorrentino, 2004).

In murine studies, it has been shown that transplantation of single, limit dilution or small numbers of LT-HSCs results in both HSC expansion and lifelong self-renewal, through at least four or five serial transplant generations (Allsopp et al., 2001). Transplantation of downstream short-term, self-renewing ST-HSCs or MPPs results in robust but transient multilineage reconstitution after which residual HSCs recover (Christensen and Weissman, 2001). In preclinical studies, xenogeneic transplant models have been developed in which human hematopoietic cells can engraft in immunodeficient mice and then be serially passaged *in vivo* (Glimm et al., 2001; Holyoake et al., 1999). Growth factor and stroma-induced HSCs expansion experiments suggest that HSCs and HSC progeny retain the ability to engraft recipients after several transplantations, but may actually be more differentiated with less proliferative capacity, thus, cell with HSC properties may themselves be

heterogeneous, and their self-renewal capacity man depend in the source from which they are derived, with cord blood HSCs having greater self-renewal capacity then mobilized peripheral blood HSCs (Glimm and Eaves, 1999; Glimm et al., 2001; Glimm et al., 2002; von Kalle et al., 1998; Zavidij et al., 2010). Overall, preclinical studies of human HSCs expansion assayed based on cytokine-induced conditions and repopulating ability in xenogeneic transplant models suggest the potential for enhancement of ST- or LT- repopulating activity, only a few cell division at most have been observed. Thus, current efforts are now focused on identifying and exploiting other extrinsic and intrinsic regulators of stem cell fate.

Retroviral mediated overexpression of Hox transcription factors, in particular Hoxa9 and HoxB4, has led to extensive ex vivo HSC expansion in vitro and in vivo without loss of lymphomyeloid repopulating ability (Antonchuk et al., 2002; Sauvageau et al., 1995; Thorsteinsdottir et al., 2002). The self-renewal induced by HoxB4 has been explained by exploration of extrinsic regulators of cell fate involved in embryonic development, such as BMP-4, a member of the TGF-ß superfamily, and Sonic Hedgehog family of proteins, both have been implicated in early hematopoietic development. In ex vivo expansion of human cord blood, human BMP-4 has been shown to increase the survival of repopulating blood cells (Bhatia et al., 1999), while Sonic Hedgehog protein has been shown to induce an increase in repopulating cells via downstream BMP signals (Bhardwaj et al., 2001). Proteins of Wnt pathways, which are involved in the growth and differentiation of a variety of primitive tissues, have also been implicated in the regulation of hematopoiesis, possibly exerting their effects through stromal cells (Yamane et al., 2001), and have been shown to stimulate the proliferation of hematopoietic precursor cells (Reya et al., 2003). Interaction between various signaling pathways has also been reported for HSCs expansion regulation. For instance, Notch-1 receptors have been shown to be upregulated in response to Wnt signaling in HSCs (Reya et al., 2003); manipulating of Notch signaling pathway, along with cytokine-induced signaling pathways can enhance stem cell self-renewal ex vivo and thereby increase HSCs numbers for transplantation (Varnum-Finney et al., 2003; Varnum-Finney et al., 2000).

Clinically, ex vivo expansions studies of primitive cells from BM, peripheral blood cells and cord blood cells have demonstrated the feasibility of ex vivo expansion. But cytokine-induced effects on HSC self-renewal and expansion are limited, clinically significant expansion has not been achieved and optimal conditions for expanding HSCs remain undefined. Continued development of methods using novel stem cell regulators that are shown to expand HSCs in preclinical studies is necessary, and current trials are underway to assess the safety and efficacy of these approaches, which hold promise for the further of expanding HSCs ex vivo and efficient transplantation.

1.2 Somatic gene therapy

Genes are the basic physical and functional units of heredity. Genes are specific sequences of bases that encode instructions on how to make proteins. When genes are altered so that the encoded proteins are unable to carry out their normal functions, genetic disorders can result. The aim of gene therapy is to modify the genetic material of living cells for therapeutic purposes (Amado and Chen, 1999). Gene therapy involves the insertion of a functional gene or another molecular that contains information sequence into a cell to achieve a therapeutic effect. There are two types of gene therapy: somatic cell and germ line. In germ line gene therapy concept, a health gene is inserted into the fertilized egg of an animal; every cell that develops from this egg will have the corrected gene. Because of the serious social and ethical considerations with this type of gene therapy, somatic cell gene therapy is the only technique performed so far (Anderson, 1998). The purpose of the procedure is to eliminate the clinical consequences of a disease and the inserted gene is not passed on to the patient's offspring. Gene therapy provides modern medicine with new perspectives that were unthinkable two decades ago. Progress in molecular biology and, especially, molecular medicine is now changing the basics of clinical medicine.

1.2.1 *Gene delivery approaches*

Long-term correction of genetic diseases requires permanent integration of therapeutic genes into chromosomes of affected cells and continuous expression at therapeutic levels. There are various non-viral and viral methods to allow the insertion of the therapeutic gene in various cell types, tissues and organs by ex vivo and in vivo strategies. The introduction of a therapeutic gene by non-viral methods is known as transfection, or via a viral carrier as transduction.

Gene delivery using non-viral carriers can be divided into three broad categories: physical, chemical, and biological approaches. Ultrasound-mediated transfection (Taniyama et al., 2002), laser beam transduction technology (Zeira et al., 2003), DNA-coated gold particle bombardment technology (Yang et al., 1990), superparamagnetic nanoparticles, also called magnetofection (Kren et al., 2009; Scherer et al., 2002), direct microinjection (Capecchi, 1980) or electroporation (Baum et al., 1994; Bloquel et al., 2004; Potter et al., 1984) are among the physical approach. The chemical method includes Calciumphosphate Coprecipitation (Graham and van der Eb, 1973), Lipofection (Felgner et al., 1987), Polycations, Cationic Polymers such as Polyethylenimin, in which the negatively charged nucleic acid associates with the carrier. Biological transfection is carried out by DNA-protein conjugates (Kern et al., 1999; Wagner et al., 1990) which are taken up by the target cell via receptor-mediated endocytosis. Although, non-viral vectors are particularly suitable with respect to simplicity of use, ease of large-scale production, lack of toxicity, and lack of specific immune response, there are some drawbacks with each of these non-viral vectors, including their lower efficiency, compared with viral vectors, in gene transfer and their transient gene expression.

Among the viral-based vectors, manifold different recombinant replication-defective vector systems for the therapeutic gene delivery in clinical trials have been described and over the recent years. These viruses are capable of transducing their target cells and delivering their viral payload, but then fail to continue either the budding or lytic pathway. However, they all have their individual advantages but also their individual limitations and must be judged on a careful risk/benefit analysis. In vector systems derived from Adenovirus, Herpes simplex virus, Ebstein-Barr virus,

vaccinia virus, poliovirus, alphaviruses (*e.g.* Semiliki-Forest Virus (SFV), Sindbis virus (SIN)), hepatitis viruses, SV40 viruses, or minus-strand RNA viruse (*e.g.* influenza viruses, Ebola virus), the viral DNA does not integrate into the host genome and is not replicated during cell division since these vectors remain as episome in the nucleus of a target cell (Cooper et al., 1997; Ehrhardt et al., 2008). In contrast to the non-integrating, episomal vectors, recombinant adeno-associated vectors can persist as episomes or alternatively integrate into the cellular genome (Duan et al., 1998; Kay et al., 2001; Miao et al., 1998; Nakai et al., 1999). Integrating vector systems derived from the viral family *Retroviridae* (Gammaretro-, Lenti- or Spumaviruses) have the ability to deliver genetic material to a target cell with high efficiency enabling long-term expression of an encoded transgene (Ehrhardt et al., 2008; Kay and High, 1999; von Kalle et al., 1994). Although gene delivery with viral vectors is more efficient than non-viral vector systems, the main disadvantage of integrating vector systems is their potential risk of causing insertional mutagenesis. They currently represent the most effective means of gene delivery in clinical gene therapy trials

1.2.2 *Gene transfer into hematopoietic stem cells*

Over the past dozen years, it has been demonstrated that the transfer of genes into hematopoietic stem cells (HSCs) could be a tool for the treatment of genetic diseases as well as a number of acquired diseases, including cancer, AIDS and autoimmune and neurodegenerative disorders. With the clinical application of gene therapy progressing in parallel with a better understanding of stem cell biology, the number different target cells, tissues and organs used in gene therapy has increased dramatically. However, HSCs remain the main cellular target for genetic intervention in a number of clinical settings of primary medical and scientific relevance, especially, but not only, those that aim to correct or modulate the immune system.

Nearly all viral and synthetic vectors developed for different gene therapy models have been tested for their ability to transducer HSCs. The majority of clinical gene therapy trials and pre-clinical research have been focused on the viral vectors,

predominantly murine onco-retroviruses (Kay et al., 2001). These vectors permanently integrate the transferred gene into the genome of the host cell, which should maintain transgene expression during proliferation, differentiation and maturation in all the cell lineages. However, target cell division is required for gene transduction with these vectors, whereas immature HSCs are naturally quiescent. To overcome this formidable obstacle, many efforts were focused on inducing cell division while preserving the self-renewal ability of stem cells and their potential to expand and differentiate into all blood lineages. Strategies involved the use of different hematopoietic growth factors in various combinations, with or without the additional use of BM stromal cell layers (Emmons et al., 1997; Xu et al., 1995) and human recombinant fibronectin fragments (Moritz et al., 1996; Williams, 1999). Combinations of stem cell factor (SCF), interleukin 3 and 6 (IL-3, IL-6), Flt-3 ligand (Flt-3L) and thrombopoeitin (Tpo) have been used to have beneficial effects on HSCs survival, expansion and retroviral transduction efficiency (Ailles et al., 2002; Conneally et al., 1998; von Kalle et al., 1998). The abilities of these strategies to favour gene transfer efficiency while preserving the “stem cell potential” of the transduced population were explored. Significant progress in the development of animal models of HSC gene transfer, as well as a small number of clinical settings, including severe-combined immunodeficiencies (SCIDs) and several gene-marking studies in the context of BM transplantation were achieved in last few years.

The discovery and development of new vectors derived from HIV sharply accelerated progress in the field of HSC gene transfer. This family of lentiviral vectors can transduce dividing and nondividing cells from different lineages, tissues and organs, including HSCs, with great efficiency (Naldini et al., 1996a; Naldini et al., 1996b). Lentiviral vectors efficiently transduce nondividing cells after direct in vivo injection, thus making genetic engineering easier and more suitable for large-scale medical application.

In some cases in vivo gene transfer requires that expression of the transferred genes be confined to the lineage of interest. Targeted transduction could theoretically be achieved by manipulating the vector envelope, so that binding and

infection is confined to a specific cell lineage. Another approach, the insertion of regulatory sequences into the vectors to control transcription and confine expression of the transgene to a specific lineage after cell differentiation already showed promising results (Lois et al., 2002; Miller and Whelan, 1997). This strategy, called transcriptional targeting, and related approaches have been already used successfully in preclinical models to confine globin expression to the erythroid lineage after HSC transduction (May et al., 2000).

1.2.3 *Family of retroviridae and gene therapy vectors*

Retroviridae is a family of medium-sized (about 100nm diameter) enveloped viruses with an icosahedral capsid enclosing the single-strand RNA genome which is diploid, i.e. two copies per virion. There are three subfamilies: *Oncornavirinae*, which includes avian, bovine, feline and murine leukemia/sarcoma viruses; *Lentivirinae*, which includes maedi/visna virus of sheep, caprine arthritis encephalitis and equine infectious anemia viruses and human immunodeficiency virus 1 and 2; *Spumavirinae*, which includes nonpathogenic viruses of monkey, cattle and cats, recognized only in cell culture where they produce syncytia (bovine and feline syncytia-forming viruses) which have a vacuolated cytoplasm, hence also called foamy viruses. The virions of a retroviridae interact between a virally-encoded enveloped protein and a cellular receptor and then enter the host cells. The enzyme reverse transcriptase which is present in the virion transcribed viral RNA into DNA copy. The viral DNA copy merges and becomes a permanent part in the host genome. Provirus is referred to this merged DNA. The viral genes are being expressed using the host cell's transcriptional and translational machinery. To create new viral RNA, the provirus is being transcribed by the host RNA polymerase II. Other cellular processes then transported this new viral RNA out of the nucleus. Some of the new RNAs are being divided to enable the expression of some genes, and those undivided new RNAs are left as full-lengths RNA. The host cell's translational machinery combines the viral proteins. The viral proteins then gather and sprout from the host cell (Figure 1-2). All members of the retroviridae go through this reproduction cycle except for spumaviruses. The

reverse transcription is completed by the spumaviruses in the virus-producing cells, instead of the infected target cells.

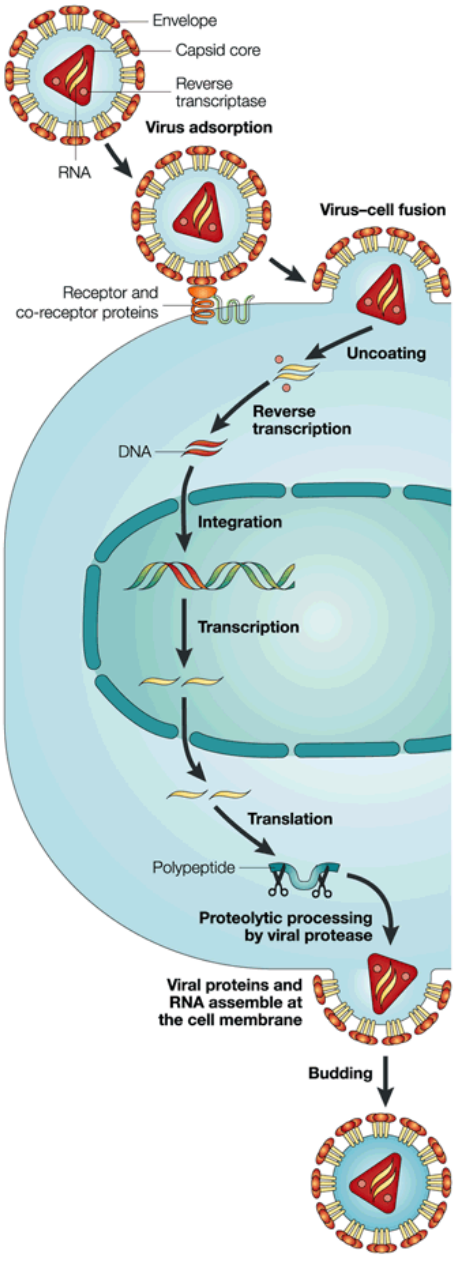

Figure 1-2 Retroviral Life Cycle. Entry of a retrovirus into a cell is initiated by interaction of its envelope proteins with cellular receptors followed by internalization through membrane fusion or phagocytosis. The vector core is released and the RNA serve as a template for reverse transcription and formation of the preintegration complex (PIC). The PIC gains access to chromatin during mitosis or, in the case of lentiviruses, by ingress through the nuclear membrane. Following integration into host cell DNA, the retroviral genome is expressed in RNA molecules which are transported to the cytoplasm to serve as a template for synthesis of new viral protein and, in the case of unspliced RNA species, as a substrate for formation of new viral particles. (De Clercq, 2002).

1.2.3.1 Retroviral vectors

Retroviruses are named for their ability to transcribe RNA into DNA. They have two identical single-stranded RNA molecules in the nucleocapsid. The two single stranded RNA molecules have a 5'cap and 3'polyA tail and from a weak dimer linkage structure at several sites along the strands. All retroviruses contain three coding domains: *gag*, *pol* and *env*. The *gag* domain encodes the matrix-, capsid- and nucleoproteins; the *pol* domain encodes the viral protease (pro) together with the reverse transcriptase (RT) and integrase, while the *env* domain encodes the envelope glycoprotein. When the virus particles infect the target cells, the viral RNA genome enters the cytoplasm as part of a nucleoprotein complex. The process of reverse transcription mediated by viral RT begins, which generates a linear DNA duplex via an intricate series of steps in the cytoplasm (Figure 1-2). This DNA is collinear with its RNA template, but it contains terminal duplication known as the long terminal repeats (LTRs) that are not present in viral RNA. LTR regions of provirus, which include promoter/enhancer regions & sequences, are required for integration and for production of virus RNA.

The basic principle behind vector development is to replace the coding region (*gag-pol-env*) with the transgene of interest. The *cis* elements required for the RNA encapsidation, reverse transcription and transcription need to be retained in the transfer vector. These include the packaging signal (Ψ), the viral LTR, primer binding site (PBS) and the polypurine tract (PPT). Additional proteins necessary for vector production are supplemented in *trans*, in packaging constructs. Retroviral vectors (RVs) are most frequently based upon the Moloney murine leukaemia virus (Mo-MLV), which is an amphotrophic virus, capable of infecting both mouse cells and human cells, enabling vector development in mouse models and human treatment. The viral genes (*gag, pol & env*) are replaced with the transgene of interest & expressed on plasmids in the packaging cell line. Transgene expression can either be driven by the promoter/enhancer region in the 5' LTR, or by alternative viral (e.g. cytomegalovirus, Rous sarcoma virus) or cellular (e.g. beta actin, tyrosine) promoters. The tropisms of RVs can be extended by replacing the *env* gene with that of another virus, in a technique called pseudotyping. Vesicular stomatitis virus G protein has

been introduced in Mo-MLV derived vectors (Burns et al., 1994), which are also more stable when purified by ultracentrifugation.

A requirement for retroviral integration & expression of viral genes is that the target cells should be dividing. This limits gene therapy to proliferating cells in vivo or ex vivo, whereby cells are removed from the body, treated to stimulate replication & then transduced with RV, before being returned to the patient. When treating cancers in vivo, tumor cells are preferentially targeted (Roth et al., 1996). However, ex vivo cells can be more efficiently transduced when exposure to higher virus titers & growth factors (Glimm et al., 1998; Glimm et al., 1997).

Though retroviral mediated transgene expression is usually adequate in vitro & initially in vivo, prolonged expression is difficult to obtain. Retroviruses are inactivated by c1 complement protein & an anti-alpha galactosyl epitope antibody, both present in human sera (Rollins et al., 1996; Rother et al., 1995). Transgene expression is also reduced by inflammatory interferons, specifically IFN-alpha & IFN-gamma acting on viral LTRs (Ghazizadeh et al., 1997). Complete transcriptional inactivation observed shortly after retroviral integration is the most dramatic type of silencing and is common in embryonic stem cells (Cherry et al., 2000). Some recent studies indicate that retroviral vectors are affected by silencing through multiple epigenetic pathways including DNA methylation and histone modifications (Ellis, 2005).

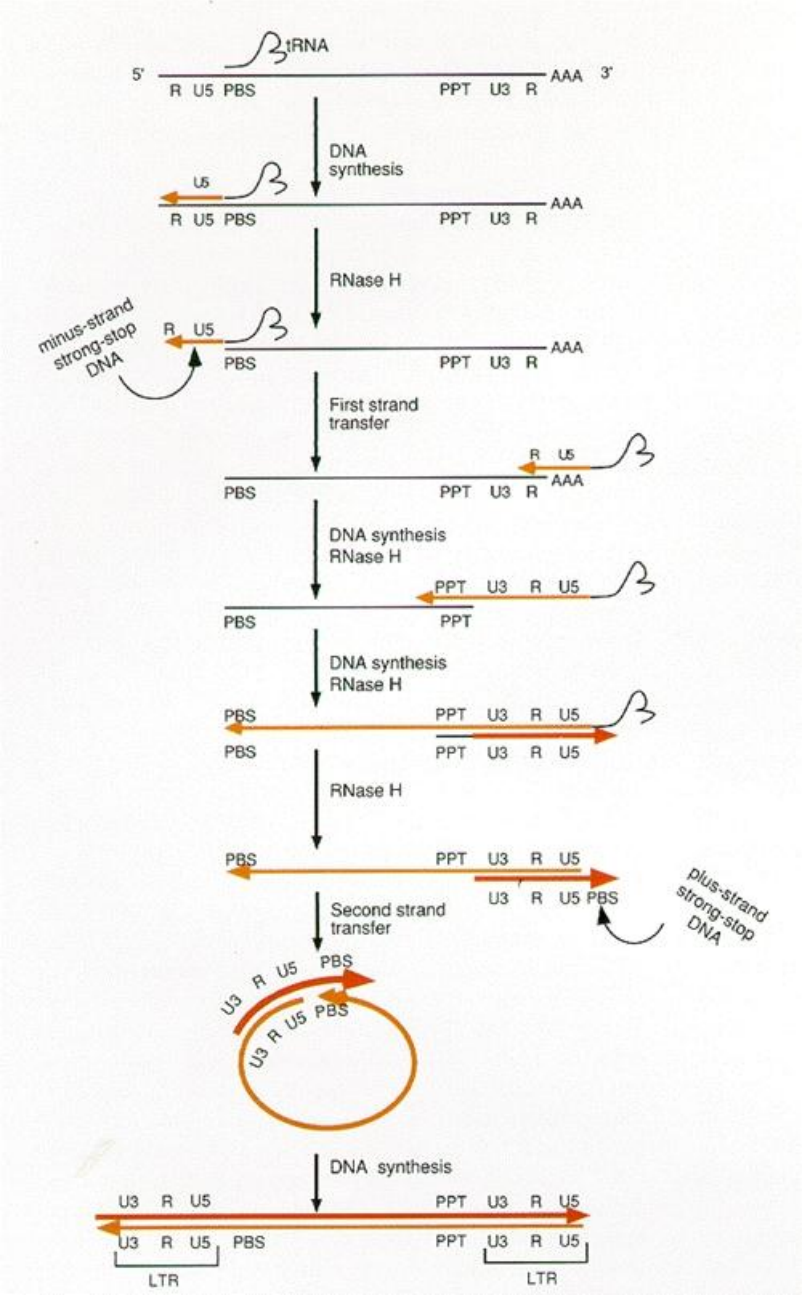

Figure 1-3 Process of reverse transcription of the retroviral genome. (Black line) RNA; (light color) minus-strand DNAs; (dark color) plus-strand DNA. Reverse transcription is initiated by the binding of specific tRNA(s) from the host cell to the virus RNA primer binding site (PBS) where DNA synthesis extends from the tRNA (acting as a primer) to the 5′ end of the virus RNA template, thus copying the U5 and R region at that end. This creates a short DNA minus-strand that still contains the tRNA at its 5′ end. This process is closely associated with degradation of the part of the template RNA that has just been copied (RT also has RNase H activity) except for the PBS region, which is protected by the tRNA binding. The R region of the newly synthesized DNA minus-strand is then free to align with the complementary 3′ R region of the partially degraded RNA template. This alignment may be aided by the affinity of the RT for the 3′ end of the RNA template as well as other factors. Synthesis of the DNA minus-strand then proceeds through the PBS region. As before, the reverse transcription is accompanied by degradation of the remaining RNA template, except for an RNase-resistant polypurine tract (PPT) that will serve as the primer for synthesis of the plus-strand DNA through its PBS region. The tRNA is cleaved from the minus-strand during the synthesis of the plus-strand PBS region. The next step involves the noncovalent circularization of the DNA minus-strand mediated by the complementary pairing of the PBS regions of both strands, which includes the loss of the PPT at the 5′ end of the plus-strand.This now allows DNA synthesis of both strands to go

to completion. Extension of the minus-strand from its 3' end displaces its own 5' region from the plus-strand, thus allowing this region of the minus-strand to become the complement for completion of synthesis of the plus-strand.The final result of this complex synthesis is a double-stranded DNA pre-integration complex with identical terminally redundant LTR elements (U3-R-U5) on both ends. This form of the virus has the capacity to insert into the cellular genomic DNA where it becomes a stably integrated provirus capable of producing virus RNA copies. The integrated provirus form has the structure: U3-R-U5-PBS-ψ-gag-pol-env-PPT-U3-R-U5 (Coffin, 1997).

1.2.3.2 Lentiviral vectors

Lentiviruses are a family of complex retroviruses inducing chronic and progressive disease typically associated with infection of human helper T cells and other macrophages. Replication incompetent vector particles derived from lentiviruses have been shown to mediate transfer and expression of transgene into a variety of cells. Dissimilar to other RVs, in particular those derived from gammaretroviruses, lentiviral vectors (LV) can mediate gene transfer into non-dividing cells, e.g. stem cells, lymphocytes, dendritic and nerve cells. Thus, in addition to use in ex vivo cell transduction LV could be useful for gene delivery in vivo. In addition, LV may allow for long-term transgene expression, as the transcript silencing observed with RVs is less frequent with LV and as such may provide the means for long-term in vivo clinical management of chronic diseases. However, in common with gammaretroviral vectors, LV suffer a number of drawbacks as gene transfer vectors, including limited insert size (8kb) of the transgene and regulatory sequences; difficulty in producing high titers of stable vector particles, and probability of activating or inactivating an endogenous DNA sequence through insertional mutagenesis. Additionally, lentiviral genomes are more complex than those of the gammaretroviruses making design of LV a greater challenge.

Currently, in comparison with gammaretroviral vectors, LVs appear to raise greater quality, efficacy and safety concerns, especially since one of the main foci for development of LV has been their derivation from HIV, a severe human pathogen. Major concerns regarding LV manufacture and clinical use are: the potential generation of replication competent lentiviruses (RCL) during LV production; in vivo recombination with lentiviral polynucleotide sequences and insertional addition of proviral DNA in or close to active genes, which may trigger tumor initiation or promotion. Overall, the biohazards associated with the contamination of the LV with

a RCL during production might be considered similar for all types of LV whilst the strategies for minimizing such contamination would be similar to those already in place for gammaretroviral vectors.

1.2.3.3 Integrase deficient lentiviral vectors

Recent gene therapy trials employing retroviral vector and lentiviral vector-mediated gene transfer showed improve treatment for various genetic, malignant, and infectious diseases, as well as benefit of clinical outcome. The efficient gene correction achieved by using integrating vector is contributed by their ability to integrate into the host genome and stable transgene expression. However, adverse events described in few gene therapy trials were imputed to vector insertional mutagenesis, for instance, insertion of transgenes in the vicinity of oncogenes or in the tumor suppressor gene, caused malignant transformation as well as leukemia (Hacein-Bey-Abina et al., 2003a; Hacein-Bey-Abina et al., 2003b).

The retroviral integrase (IN) protein encoded by the 3' end of the viral *pol* gene, along with the virus-specific sequences attachment sites (*att*), which located at each end of viral LTR, are the only viral factors known required essentially for integration of the retroviral genome into the host DNA. Retroviral integration is an essential process for viral replication in the viral life cycle. After retroviruses enter the host cells, reverse transcription of the viral genome is taken place, resulting in a DNA copy of the viral genome located within a viral nucleoprotein (preintegration) complex. The viral DNA in the preintegration complex were modified in cytoplasm by IN mediates cleavage of the 3' end of each strand immediately beyond a conserved subterminal CA dinucleotide, leaving a hydroxyl group at the 3' end of each strand of viral DNA. After the preintegration complex migrates to the nucleus, a concerted strand transfer mediated by IN through a nucleophilic attack involving the viral 3' hydroxyl residue and the phosphate backbone of the target DNA. The coordinated strand transfer occurs into the major groove of the target DNA, and the two viral ends attack the target DNA in a staggered fashion that determines the length of the virus specific direct repeats characteristically flanking each provirus (Figure 1-4).

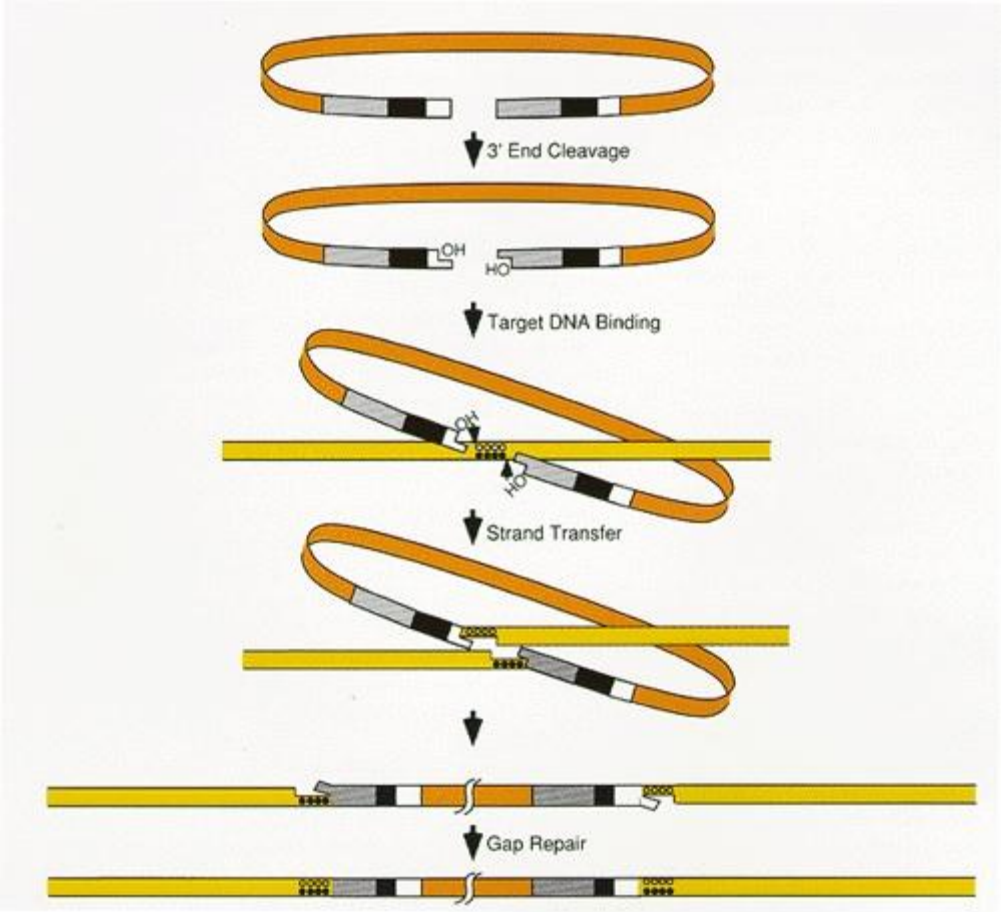

Figure 1-4 The principal steps in retroviral DNA integration. (Coffin, 1997)

In vitro assays using wild-type and mutant HIV-I IN have showed that classical I mutations in highly conserved D,D-35-E motif of the IN catalytic core domain have no apparent defect in 3' processing and strand transfer of viral life cycle (Engelman et al., 1995), but prevent the proviral integration (Gaur and Leavitt, 1998; Leavitt et al., 1996), resulting in enhanced levels of viral extrachromosomal DNA, which consist of late-linear viral DNA and circular forms (1-LTR and 2-LTR forms) in the nucleus of infected cells, and 10^3- to 10^4- fold reduction of integrated events in comparison to wild-type IN mediated transduction (Nightingale et al., 2006). Several studies have reported efficient gene expression in vitro and in vivo using integrase defective lentiviral vectors (IDLVs). IDLVs are also proficient vehicles for homologous recombination, site-specific integration and transposition (Lombardo et al., 2007; Vink et al., 2009), and can be converted into stable, replicating episomes (Vargas et al., 2004). These properties, combined with their highly reduced risk of causing

insertional mutagenesis, have led to a renewed interest on IDLVs for genetic analysis and gene therapy (Yanez-Munoz et al., 2006).

1.2.3.4 Novel development of clinical gene transfer vectors

The majority of clinical gene therapy trials for inherited genetic diseases and cancer therapy have been performed using murine onco-retrovirus as the gene delivery vector over the past dozen years. The earliest systems used were relatively inefficient in both the rates of transduction and expression of the transgene. Formidable obstacles, including low rates of transduction, low protein expression, silencing of expression from the integrated vector, risk from insertional mutagenesis and in the case of HSC the numerical disadvantage of the engrafting transduced cells, limited the efficacy of gene therapy for many target diseases. Developments in novel viral mediated gene transfer that are in progress have begun to overcome these obstacles.

To design a viral vector, it is crucial to separate *cis*-acting sequences, required for the transfer of the viral genome to target cells from the *trans*-acting sequences encoding the viral proteins. Thus the biosafety and the efficacy of vector primarily depend on the level to which segregation of *cis*- and *trans*-acting functions of the viral genome is obtained (Kay et al., 2001). Typically, these viral constructs are then introduced in to a producer cells that would generate infectious but replication deficient viral particles, as these particles can only encapsidate, which results in single round of infection, a process known as transduction. In the last several years, different variations in vectors design have been introduced, such as self-inactivating vectors and vectors which have additional safety features or designing the integrating vectors in such a way so that it can integrate into pre-existing transgenic loci within the genome. Furthermore, improvements in safety may be also gained by using tissue-specific promoters to drive transgene expression, as it would not only be more tightly regulated expression, but also it would minimize the potential of insertional oncogenesis that are often associated with over expression strategies (Brenner and Malech, 2003).

The replication deficient LVs have been demonstrated to mediate stable and long-term transgene expression in various non-dividing and other difficult to transduce cells such as stem cells, nerve cells lymphocytes, and dendritic cell. Long-term transgene expression potential is of particular important for *in vivo* clinical management of chronic disease. In addition, the LVs can be pseudotyped with specific viral envelopes that highly affect the vector tropism and enhanced biosafety and transduction efficiency. However though LVs are designed to be replication-defective, there are constant concerns about the possible generation of wild type or novel human pathogens, while using these vectors following clinical administration. Therefore significant modifications have been made in the LVs since the initial findings by Naldini and co-workers (Dull et al., 1998; Naldini et al., 1996b). Continual efforts are ongoing in recent years that have attempted to further delete wild-type HIV sequences in the lentiviral transfer (integrating) and packaging vector to enhance vector titer and transgene expression. The newer generation LVs technology has many in-built safety features that include separation of critical components of the viral genome onto distinct plasmids (split genome approach) that allows the safe production of one-round infectious viral particles. Besides, to delineate the mechanism by which a specific gene contributes to cell growth and viability, inducible and/or conditional systems were explored. Conditional gene expression is an absolute requirement in certain cases to have a tight regulation over transgene expression, for e.g. generation of a transgenic line may become problematic if the presence of the transgene leads to conditions which are toxic, lethal to cell growth or probably responsible for causing cell differentiations. Tetracycline-dependent regulatory systems (Tet based systems) have been successfully employed during enormous biological studies, including transgenic mouse modeling and Induced pluripotent stem cell (iPSc) generations (Brambrink et al., 2008; Stieger et al., 2009).

To minimize the possibilities of insertional mutagenesis, two major strategies are being used: first, directing LVs insertion into specific sites within the host genome (Du et al., 2009; Pfeifer et al., 2001); and secondly by exploring on integration

deficient LVs to install therapeutic transgenes. A variety of IDLVs by introducing point mutations into the catalytic site, chromosome binding site, and viral DNA binding site of the viral integrase (Leavitt et al., 1996) have been developed. Mutating or deleting the attachment (*att*) sites at the ends of viral genome, involved in recognition by IN and cut by it, has also been shown to hamper provirus formation (Masuda et al., 1998) but no further reduction of integration frequencies when used in conjunction with IN mutations in IDLVs (Apolonia et al., 2007; Nightingale et al., 2006) .

Modifications in the transfer plasmid include self-inactivation (SIN). SIN vectors are produced through the targeted elimination of the promoter sequences from the 3'-LTR U3 region, making transgene expression necessarily reliant on an internal promoter. SIN vectors have reduced risks of host gene transactivation, RCR generation, and mobilization upon superinfection with wild-type virus (Dull et al., 1998; Miyoshi et al., 1998; Zufferey et al., 1998). In addition, boundary elements such as the insulator from the chicken beta globin cluster, which has both enhancer blocking and chromatin boundary forming activity, can be added to the 3' LTR during vector assembly leading so that the transgene is flanked by insulator elements following vector integration (Evans-Galea et al., 2007).To further increase the safety, polyadenylation enhancers, internal promoters without enhancer capacity or regulatable and tissue specific promoters could be introduced into SIN vectors (Matrai et al., 2010).

1.2.4 *Clinical gene therapy*

1.2.4.1 Human primary immunodeficiency syndrome

The primary immunodeficiencies (PID) represent a class of disorders in which there is an intrinsic defect in the human immune systems (rather than immune disorders that are secondary to infection, chemotherapy, or some other external agent). They result from inherited mutations in genes required for the production, function or survival of specific leukocytes (e.g. T, B or NK lymphocytes, neutrophils and antigen-presenting cells). The bodies of people with primary immunodeficiency

can't get rid of germs or protect themselves from new germs as well as they should. The World Health Organization has identified more than 80 kinds of primary immunodeficiency diseases. But, the general category includes more than 100 diseases caused by an immune system that doesn't function properly.

Because these leukocytes are produced from the pluripotent HSCs in the BM, allogeneic BM transplantation (BMT) from a healthy donor into a patient with a PID can lead to immune restoration. BMT was first reported for the treatment of patients with severe combined immune deficiency (SCID) and Wiskott-Aldrich syndrome (WAS) in 1968 (Bach et al., 1968; Gatti et al., 1968). Since that time, hematopoietic stem cell transplantation (HSCT), including BM, mobilized peripheral blood and umbilical cord blood, has been performed for many of the life-threatening PID, including SCID, WAS, chronic granulomatous disease (CGD), leukocyte adhesion deficiency (LAD), and others.

While allogeneic HSCT may lead to essentially normal immune function and excellent quality of life for patients with severe PID, there are significant risks, both from the conditioning regimen, as well as from the immunologic consequences such as graft-vs-host disease (GVHD) wherein donor lymphocytes transplanted with the HSC source react against the recipient's tissues.

The concepts underlying gene therapy for PID is that gene correction and re-transplantation of the patient's own HSC could lead to the same clinical benefits as allogeneic HSCT, but without the immunologic complications. Many efforts have been taken and many stirring results have been achieved from the treatment of PIDs. New approaches using safer integrating vectors or direct correction of the defective gene underlying the PID are being developed and may lead to safer and effective gene therapy for PID.

1.2.4.1.1 X-linked Chronic Granulomatous Disease

X-linked chronic granulomatous disease (X-CGD) is a rare primary immunodeficiency, resulting from mutations in the gene encoding for $gp91^{phox}$ which, together with $p22^{phox}$, forms the membrane-bound flavocytochrome b558

complex, the redox center of the phagocyte nicotinamide adenine dinucleotide phosphate (NADPH)-oxidase complex (Ryser et al., 2007). NADPH oxidase catalyzes the formation of superoxide which is then dismutated to hydrogen peroxides (H_2O_2) and subsequently transformed by myeloperoxidase into reactive oxygen species (ROS). ROS can directly kill bacteria through their oxidizing activity. More recently it has been shown that production of superoxide anion triggers the influx of H^+ and K^+ in the phagocytic vacuole, activating neutral proteases that are responsible for destruction of phagocytosed microorganisms (Segal, 2005). Phagocytes from X-CGD patients are consequently unable to mount a characteristic respiratory burst that accompanies phagocytosis, resulting in inefficient bacterial and fungal killing, and affected individuals suffer from recurrent life-threatening infections (Babior, 2004).

Conventional treatment of X-CGD includes prophylactic drugs, such as antibiotics and antifungals, which are given regularly to patients. Recombinant IFN-γ therapy also appears beneficial in some patients. Until recently the only cure for X-CGD was allogeneic HSC transplantation; offered only to chronically ill patients due to the high morbidity and mortality associated with this procedure as a result of the myeloablative conditioning regimen and graft-versus-host disease (GvHD). However, improved outcomes following this procedure have been reported for patients free from infections at the time of transplant (Horwitz et al., 2001; Seger et al., 2002).

As only low fractions of the patients' hematopoietic cells have to be corrected to convert the phenotype to a healthy individual, and as even 1% of long-term fully corrected cells can be expected to be highly beneficial, CGD is a promising candidate disease for gene therapy. In vitro and in vivo preclinical studies using retroviral-vector mediated overexpression of one of the oxidase subunits results in high levels of oxidase production in transduced CD34+ cells (Weil et al., 1997), long-term high-level reconstitution of NADPH oxidase activity in rodent animal models (Dinauer et al., 1999; Mardiney et al., 1997; Sadat et al., 2003) as well as a moderate levels of gp91-phox stable expression about 8 months post-transplant in two rhesus macaques(Brenner et al., 2006).

In the mid-late 1990s, the first two gene therapy for CGD in trials performed at the National Institutes of Health and Indiana University School of Medicine. They used ex vivo gene transfer with gamma-retroviral vectors to peripheral blood stem cells (PBSC) mobilized by granulocyte-colony stimulating factor (G-CSF), which were reinfused without prior treatment with chemotherapeutic agents. Low frequencies of corrected granulocytes were seen in the peripheral blood for the first months in a few of the subjects, but no long-lasting effects were seen (Malech et al., 1997). Subsequently, a trial was performed in Germany in which a moderate dosage of busulfan (8 mg/kg) was given prior to reinfusion of the transduced PBSC (Ott et al., 2006). The two young men treated had severe infections that had persisted despite intensive medical therapy. These infections resolved after the gene therapy procedure, as the patients made neutrophils with functional oxidase activity. Relatively high levels of corrected leukocytes were seen in peripheral blood (~20%) in the first months after the gene therapy procedure and this rose to as high as 80% over the first year. The vector integration sites were studied in the CGD patients and revealed a highly restricted pattern, with the majority of vector integrants in the engrafted stem cells being near one of a few genes known to be involved in myeloid cell proliferation (*MDS1/EVI1, PRDM16 or SETBP1*). Both patients subsequently developed myelodysplasia with monosomy 7 and first one died 2.5 years after gene therapy as a result of a sepsis. The second patient has undergone successful allogeneic transplant (ESGT, 2006; Stein et al., 2010). The underlying mechanisms for the oligoclonal expansion and progression to myelodysplasia are not fully elucidated, but the retroviral vector used had an LTR with potent enhancer activity in myeloid progenitor cells and this may have led to trans-activation of genes that promote myeloid cell proliferation.

1.2.4.1.2 Wiskott-Aldrich Syndrome

Wiskott-Aldrich syndrome (WAS) is a rare X-linked hereditary disorder associated with combined immunodeficiency, thrombocytopenia, small platelets, eczema, and increased susceptibility to autoimmune disorders and cancers. (Thrasher and Kinnon, 2000). It caused by mutations in the gene for the Wiskott-Aldrich syndrome protein

(WASP), located at Xp11.22-p11.23 (Derry et al., 1994; Devriendt et al., 2001). WASP and several related proteins are involved in the reorganization of the actin cytoskeleton by activating the actin-related protein 2/3 complex that mediates actin polymerization in all cells of the hematopoietic system (Thrasher, 2002). As WASP is exclusively expressed in the hematopoietic cells, the complex biologic features of this disease result from multiple dysfunction in different subgroups of leukocytes, including defective function of T and B cells, disturbed formation of natural killer (NK) cell immunologic synapse, and impaired migratory responses in all leukocyte subgroups (Bosticardo et al., 2009; Thrasher and Burns, 2010). The patients bearing the *WAS* mutation suffering from recurrent pyogenic, viral and fungal infections as well as bleeding diathesis and an increased incidence of Epstein-Barr virus (EBV)-associated lymphoproliferative disease, severe WAS leads to an early death (Imai et al., 2004).

Currently, the only curative treatment approach for WAS consists of allogeneic hematopoietic stem cell transplantation (HSCT). However, HSCT is associated with significant mortality and morbidity resulting from infection, graft-versus-host disease, non-engraftment and post-transplant lymphoproliferative syndrome (Derry et al., 1994). In light of the less satisfactory results from HSCT fro WAS treatment, in particular regarding those patients for whom an HLA-matched donor is unavailable, several groups have embarked on the development of gene therapy as a more specific and potentially less toxic therapeutic option. Various in vitro studies as well as in vivo mouse model studies have demonstrated that retroviral gene transfer can reconstitute the cellular phenotype in WASP-deficient cells and mouse, the functional correction of gene modified cells restore the defective actin polymerization and immune protecting functions (Klein et al., 2003).

The first clinical trial of WASP gene therapy is being conducted at the Hannover Medical School in Germany, 9 out of 10 children were treated in this trial. The vector system used in this trial is a gammaonco-retroviral virus-based, LTR-driven construct encoding human *WASP* cDNA, which has been extensively assessed in preclinical experiments using mouse model and human CD34+ derived from WAS patients

(Devriendt et al., 2001; Dewey et al., 2006). This study shows that gene therapy for the WAS is feasible and no adverse treatment related events up to 3 years after gene therapy (Boztug et al., 2010).

1.2.4.2 Integration site preference of retroviral and lentiviral vectors

Unbiased, high-throughput studies are important to identify integration hot spots and provide insights into mechanisms governing integration site selection of viral vectors used in clinical settings. Historically, the integration events of retroviruses were believed to be random, and the chance of accidentally disrupting or activating a gene was considered remote. Early retroviral integration studies proposed that condensed chromatin from inactive DNA regions disfavored retroviral integration, thereby concentrating the integration in more open active chromatin areas of the host genome (Panet and Cedar, 1977; Rohdewohld et al., 1987; Vijaya et al., 1986). The availability of the human genome sequence and the development of methodology to define DNA sequence at the site of retroviral integration have shown that integration is far from random. The murine gamma-retroviral vectors used in gene therapy trails to integrate favor the transcription start sites of actively transcribed genes in the target cell population (Wu et al., 2003), whereas lentiviral vectors favor active genes with more uniform distribution within transcriptional units (Schroder et al., 2002). This is suggests that lentiviral vectors are less prone to transactivate a downstream oncogene by transcriptional reading through. Integration into or near genes involved in cell-cycle progression or cell survival has the potential to alter the biologic properties of cells which, in the context of engraftment, can result in preferential engraftment and expansion and even clonal dominance. In the first case, a mouse leukemia developed as a consequence of the aberrant expression of the Evi1 proto-oncogene caused by a nearby integrated proviral genome (Li et al., 2002). Side by side comparisons in valid models of oncogenic transformation have shown the impact of integration site selection between different vector designs. Study with gene trap vectors performed in primary hematopoietic cells confirmed a stronger bias of MLV vectors to insert in promoter-proximal windows, as compared with similarly designed HIV-derived vectors (De

Palma et al., 2005). Moreover, in mice engrafted with cdkn2a-/- BM, a gamma-retroviral vector was found to accelerate tumorigenesis, whereas a lentiviral vector, present at higher copy number, did not have such a detrimental effect (Montini et al., 2006). Studies on transduced BM and CD34+ cells have further revealed that vector integration events are undeniably non-random, frequently clustered and potentially able to induce immortalization *in vitro*, conversion of immortalized cells to growth factor independence, clonal dominance *in vivo*, or even contribution to leukemogenesis *in vivo* (Du et al., 2005; Kustikova et al., 2005; Modlich et al., 2006; Modlich et al., 2009).

1.2.4.3 Risk and side effects of retroviral gene transfer in clinical trials

Integration of the retroviral provirus in the host genome can lead to stable transgene expression, one of the outstanding features of retroviral vectors. Unfortunately, it is also their demerit, because integration can cause insertional mutagenesis and cellular transformation, whose undesired effects may lead to cancer in a worst-case scenario. In a clinical gene therapy trial for X-linked severe combined immunodeficiency using retroviral vectors expressing common γ chain of the interleukin-2 receptor (γc), integration of the retroviral vector in the LMO2 proto-oncogene resulted in insertional mutagenesis and development of T-cell leukemia in 4 children of Paris trial and 1 child of London trial (Hacein-Bey-Abina et al., 2008; Hacein-Bey-Abina et al., 2003b; Howe et al., 2008), this theoretical risk has unfortunately materialized.

The discovery of common integration sites (CISs), namely the nonrandom insertional clustering of integration sites into the same genetic locus in 2 or more cells in a population, is thought to reflect the biases toward integration into actively transcribed genes in the target cells. Retroviral integration analysis has been reported in detail for X-SCID and the Adenosine deaminase (ADA)-SCID trial in Italy (Aiuti et al., 2007; Deichmann et al., 2007; Howe et al., 2008; Schwarzwaelder et al., 2007). The distribution of integration sites favoring transcription start sites and CpG islands and the predilection for actively transcribed genes were similar to that

reported for cell lines and animal models. CISs were defined in patients in the Paris trial for X-SCID with 5 insertions within LMO2, 8 within ZNF217, and 9 within or near CCND2, a gene involved in cell-cycle progression. CCND2 and LMO2 also emerged as CISs in the patients in the Italian ADA trail as well as in the London X-SCID trail. Vector insertion induced upregulation of LMO2 has triggered abnormal T cell proliferation, potentially enhanced by the constitutive expression of the Interleukin-2 receptor gamma chain transgene transduced in X-SCID trials. Further observation showed that insertional side effects are not restricted to lymphopoiesis. Large-scale mapping of retroviral integration sites in the first successful gene therapy trial of chronic granulomatous disease (CGD) revealed that activating insertions in the zinc finger transcription factor homologs MDS1/EVI1, PRDM16, or in SETBP1 have expanded gene-corrected long term myelopoiesis 3- to 5- fold in both patients, providing direct evidence that these genes influence regulation of normal hematopoiesis (Ott et al., 2006). In both patients treated in CGD trial, progressive decline in blood counts has been reported, including platelets, red blood cells and neutrophils, at 15 and 28 months after gene therapy, respectively. BM examination showed myelodysplasia, a preleukemic clonal marrow failure syndrome characterized by ineffective and disordered myeloid maturation. Despite the persistent high frequency of vector-corrected neutrophils, expression of NADPH oxidase dropped precipitously in both subjects over time, occurred through progressive CpG methylation of the promoter contained in the LTR of the vector. Cytopenias and loss of oxidase function in one of the patients resulted in a series of infections that led to his death. The second patient was referred for unrelated donor stem cell transplantation while still infection free (Stein et al., 2010).

Despite the negative finding described above, the majority of the patients benefited from gene therapy trials. For example, in 17 out of 20 X-SCID patients, T-cell counts reached normal levels and became functionally competent, as demonstrated by normal response to mitogens and specific antigens. Ten years of follow-up available for patients who were first to be treated have provided evidence for still active thymopoiesis, with broadly diversified TCR repertoire (Hacein-Bey-

Abina et al., 2008). Gene therapy continues to have enormous potential for the treatment of disorders of the hematopoietic system, cancer and other diseases. The occurrence of these severe adverse events led to the development of a new area of research focused on virus-mediated oncogenesis and significantly contributed to improving safety standards for gene therapy vectors.

2. Projects aims

Human gene therapy trials using insertional vector systems have achieved unprecedented levels of efficiency, but unfortunately also side effects. It has been recently shown that insertional mutagenesis in preclinical and clinical gene therapy may frequently lead to gene activation in hematopoiesis, producing insertion-activated expansion of cell clones.

Thereby, intense attention was focused on the *MDS1/EVI1* and *PRDM16* gene complexes as recurrent targets for gamma-retroviral vector integration. In the first clinical gene therapy trial for chronic granulomatous disease, both patients showed significant clinical improvement and clearance of chronic infections. Surprisingly, ~3-4 months after gene therapy, the frequency of genetically modified CD15+ neutrophils began to rise, peaking at around 60% in both patients, resulting in the formation of a myelodysplastic syndrome ~2 years later. Analysis of vector integration of this two patients revealed that the expansion of marked cells accounted for vector-containing clones with insertions in three gene complexes: *MDS1/EVI1, PRDM16* and *SETBP1*. Malignant transformation was related to cell clones harboring *MDS1/EVI1* insertions.

Similar to the findings in the human CGD trial, vector insertions in the *MDS1/EVI1* and *PRDM16* loci were highly overrepresented in *in vitro* and in preclinical *in vivo* animal models, leading to over-expression of Evi1 and Prdm16, and myelomonocyctic leukemia in mice.

In contrast to full LTR driven gamma-retroviral vectors, self-inactivating (SIN) LTR containing lentiviral promise immensely for sustainable and long-term mediated clinical gene therapy. Even integrase deficient lentiviral vectors (IDLV) demonstrated a high applicability for stable, though transient gene delivery, rendering them interesting for therapeutic applications. However, little is known about the existing residual integration events of IDLV, and comprehensive characterization of IDLV integration profiles in mammalian genomes is still missing.

Towards a precise characterization of the role of MDS1/EVI1 and PRDM16 in human hematopoiesis, and to determine the biological properties of IDLVs as well as to develop potentially innovative utilization in combination with transient expression of MDS1/EVI1 and/or PRDM16 in hematopoietic stem cells, we specifically aim:

1. To analyze long-term effects of *MDS1/EVI1* and *PRDM16* loci as gamma-retrovirus integration sites in hematopoietic stem cells after WAS gene therapy.

2. To develop a lentiviral vector expression system for investigating the impact of Mds1/Evi1, Prdm16 and their truncated forms on more defined hematopoietic stem cells.

3. To test whether transient expression of Mds1/Evi1 or Prdm16 by integrase deficient lentiviral vectors (IDLVs) is sufficient to influence long-term repopulating cells, and to study integration properties and safety aspects of IDLVs.

3. Materials and Methods

3.1 Materials

3.1.1 Mice

C57BL/6J (CD45.2) and B6.SJL-*Ptprc*^{a} *Pep3*^{b}/BoyJ (CD45.1) mice (6-12 weeks age) were purchased from Charles River Laboratories and were bred at the animal facility of German Cancer Research Center. All animals were maintained under specific-pathogen-free conditions. All animal experiments were conducted in accordance with local and institutional guidelines and were approved by the Regierungspräsidium Karlsruhe.

3.1.2 Cell lines

Name	Organism	Description	Source
293T	*Homo sapiens*	Derivative of human embryonic kidney 293 cells, contains the large T-antigen of SV40	Luigi Naldini's laboratory in Milan, Italy
HeLa	*Homo sapiens*	Human cervical carcinoma cell line	Laboratory stock
C1498	*Mus musculus*	Mouse acute myeloid leukemia cell line	ATCC TIB-49
SC-1	*Mus musculus*	Mouse embryonic fibroblast cell line	Adrian Thrasher's laboratory in London, England

3.1.3 Cell culture media and additives

Name	Manufacturer
Dulbecco's Modified Eagle Medium (DMEM)	Gibco, Invitrogen
Iscove's Modified Dulbecco's Medium (IMDM)	Gibco, Invitrogen
StemSpan Serum-Free Medium (SFM)	StemCell Technologies
Opti-MEM® I Reduced Serum Medium (MEM)	Gibco, Invitrogen
Fetal Bovine Serum (FBS)	Biosera
Fetal Calf Serum (FCS)	StemCell Technologies
L-Glutamine (Glu)	Gibco, Invitrogen
Penicillin/Streptomycin (Pen/Strep)	Gibco, Invitrogen

3.1.4 General buffers chemicals and kits for cell culture

Name	Manufacturer
Dulbecco's Phosphate-Buffered Saline (PBS)	Gibco, Invitrogen
Typsin-EDTA 0.05%	Gibco, Invitrogen
Cell Dissociation Buffer, enzyme free, Hanks'-based	Gibco, Invitrogen
20% FBS in PBS	Laboratory stock
Türk's solution	Merck Chemicals
Typan-blue solution	Gibco, Invitrogen
70% Alcohol	Laboratory stock
Freezing medium (15% DMSO, 30% FBS in cell culture medium)	Laboratory stock
Thawing medium (50% FBS in cell culture medium)	Laboratory stock
Dimethyl sulfoxide (DMSO)	Karl-Roth
2x HEPES balanced saline (281 mM NaCl, 1.5 mM Na_2HPO_4, 100 mM HEPES)	Laboratory stock
0.1x TE/H_2O (1 mM Tris-HCl, 0.1mM EDTA)	Laboratory stock
Calcium Chloride (2.5 M)	Laboratory stock
Polybrene (hexadimethrine bromide, 8 mg/ml)	Sigma-Aldrich
Polyethyleneimine (PEI)	Sigma-Aldrich
Magnetofection ViroMag L	OZ Biosciences

3.1.5 *Medium and kits for primary cell culture*

Name	Manufacturer
Mouse hematopoietic progenitor enrichment kit	StemCell Technologies
MethoCult methylcellulose-based media M3434	StemCell Technologies
MethoCult methylcellulose-based media M3630	StemCell Technologies

3.1.6 *Cytokines for primary cell culture*

Name	Manufacturer
Recombinant mouse-Interleukin 6 (rm-IL6)	R&D Systems
Recombinant mouse-Interleukin 3 (rm-IL3)	R&D Systems
Recombinant mouse-Stem Cell Factor (rm-SCF)	R&D Systems
Recombinant mouse-FLT-3 Ligand (rm-FLT-3L)	R&D Systems
Recombinant mouse-Thrombopoietin (rm-TPO)	R&D Systems

3.1.7 *General buffers and antibodies for flow cytometry*

Name	Manufacturer
Hanks Balanced Salt Solution (HBSS)	Sigma-Aldrich
2% FBS in HBSS (HF)	Laboratory stock
Red Blood Cells Lysis Buffer (RBC)	Laboratory stock
Prooidium Iodide (PI)	Sigma-Aldrich
7-Amino-actinomycin D (7-AAD)	Sigma-Aldrich
APC-H7 Rat anti-Mouse CD117 (c-kit)	BD Pharmingen
PE-Cy7 Rat anti-Mouse Ly-6A/E (Sca-1)	BD Pharmingen
PE Rat anti-Mouse CD115 (Flk-2/Flt3)	BD Pharmingen
APC Mouse Lineage Antibody Cocktail	BD Pharmingen

PE-Cy7 Rat anti-Mouse CD11b (Mac-1)	BD Pharmingen
PE Rat anti-Mouse Ly-6G/C (Gr-1)	BD Pharmingen
APC Rat anti-Mouse Ly-76 (Ter-119)	BD Pharmingen
PE Rat anti-Mouse CD117	BD Pharmingen
APC Rat anti-Mouse CD19	BD Pharmingen
PE Rat anti-Mouse CD3	BD Pharmingen
FITC Rat anti-Mouse CD34	eBiosciences
Purified Rat anti-Mouse CD16/CD32	BD Pharmingen
APC Rat anti-Mouse CD45	BD Pharmingen
PE Rat anti-Mouse CD45	BD Pharmingen
PE-Cy7 Rat anti-Mouse CD45	BD Pharmingen
FITC Rat anti-Mouse CD45	BD Pharmingen
PE-Cy7 Rat IgG2a, κ Isotype Control	BD Pharmingen
PE-Cy7 Rat IgG2b, κ Isotype Control	BD Pharmingen
PE Rat IgG2a, κ Isotype Control	BD Pharmingen
PE Rat IgG2b, κ Isotype Control	BD Pharmingen
APC Rat IgG2a, κ Isotype Control	BD Pharmingen
APC Rat IgG2b, κ Isotype Control	BD Pharmingen
APC-H7 Rat IgG2b, κ Isotype Control	BD Pharmingen
FITC Rat IgG2b, κ Isotype Control	eBiosciences

3.1.8 *Bacteria*

Name	Genotype	Manufacturer
TOP10 Chemically Competent *E. coli*	*F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG*	Invitrogen
NEB 5-alpha Competent *E. coli* (High Efficiency)	fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England BioLabs
dam$^-$/dcm$^-$ Competent *E. coli*	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)* Tet^S *endA1 rspL136 (*Str^R*) dam13::Tn9 (*Cam^R*) xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	New England BioLabs

3.1.9 *Antibiotics for bacteria culture*

Name	Stock Concentration	Working Concentration	Manufacturer
Ampicillin	100 mg/ml	100 µg/ml	Sigma-Aldrich
Kanamycin	100 mg/ml	50 µg/ml	Sigma-Aldrich
Chloramphenicol	25 mg/ml	25 µl/ml	Sigma-Aldrich

3.1.10 *General media for bacteria culture*

Name	Formulation	Final concentration
Luria-Broth (LB)	SELECT Peptone 140	10 g (1%)

Media	SELECT Yeast Extract Sodium Chloride Add $Aqua_{Bidest}$ to 1L pH 7.5 Autoclaved	10 g (1%) 5 g (0.5%)
LB-Agar	SELECT Peptone 140 SELECT Yeast Extract Sodium Chloride Bacto-agar Add $Aqua_{Bidest}$ to 1L pH 7.5 Autoclaved	10 g (1%) 10 g (1%) 5 g (0.5%) 5 g (0.5%)
S.O.C Media	Bacto-tryptone Bacto-yeast extract Sodium Chloride Potassium Chloride Magnesium Chloride Magnesium Sulfate Glucose Add $Aqua_{Bidest}$ to 1L pH 7.5 Autoclaved	20 g (2%) 5 g (0.5%) 0.584 g (10mM) 0.186 g (2.5mM) 0.952 g (10mM) 1.204 g (10mM) 3.603 g (20mM)
Glycerol Solution	100% Glycerol filtered with 0.22µm filter	50%

3.1.11 *Plasmids*

3.1.11.1 Lentiviral vectors

Name	**Description**
pMDL*gag/pol* RRE	This 3rd generation LV packaging plasmid encodes the HIV-1 Gag and Pol proteins driven by CMV promoter, and a Rev protein regulatory element (RRE) in pMDL backbone.
pMDL*gag/pol* RREintD64V	Similar to pMDL*gag/pol* RRE, but contains a single amino acid D64V mutation in HIV-1 integrase sequences , which results defective provirus integration
pRSV-REV	Part of 3rd generation LV packaging system, which encodes the HIV-1 Rev protein under the transcriptional control of RSV U3 promoter
pMD2-VSVG	Part of 3rd generation LV packaging system, which encodes the VSV G envelope protein
pRRL pptCMV-GFP-PRE	3rd generation LV transfer vector contains a Tat-independent chimeric 3'LTR in which the HIV promoter is replaced by RSV promoter and enhancer sequences, central polypurine tract (cppt) as well as posttranscriptional regulatory element (PRE) in its backbone. It encodes enhanced green fluoresces protein (GFP) driven by internal CMV promoter upon provirus integration.
pCCL pptPGK-GFP-PRE	Similar to the vector above, which Tat-independent

	chimeric 5'LTR in which the HIV promoter is replaced by CMVpromoter and enhancer sequences, and its GFP expression is driven by human PGK promoter
pCCL pptPGK-IRES-GFP-PRE	Similar to the vector above, which contains a IRES-GFP expression cassette, allows for translation initiation in the middle of a mRNA sequence as part of the protein synthesis.
pCCL pptSFFV-GFP-PRE	Similar to the pCCL pptPGK-GFP-PRE, which GFP expression is driven by SFFV promoter
pCCL pptSFFV-IRES-GFP-PRE	Similar to the vector above, contains IRES-eGFP expression cassette
pCCL pptEF1a-GFP-PRE	Similar to the pCCL pptPGK-GFP-PRE, which GFP expression is driven by human EF1a promoter

3.1.11.2 Cloning vectors

Name	Manufacturer
pCMV-Sport6	Invitrogen, 4396bp, CMV promoter, Ampicillin resistance
pYX-Asc	RZPD, 1691bp, Ampicillin resistance
pBluescriptR	Stratagene, 2998bp, Ampicillin resistance
pPCR-Script	Stratagene, 2961bp, Ampicillin resistance
pCR2.1-TOPO	Invitrogen, 3931bp, Ampicillin and Kanamycin resistance

3.1.12 *Restriction enzymes*

Restriction enzymes were purchased from New England Biolabs if there are no specific descriptions. The names of restriction enzymes were not list here. All the enzymes are stored and worked under the conditions by manufacturer's recommendation.

3.1.13 *Regents, chemicals and buffers for nucleic acid analysis*

3.1.13.1 Oligonucleotides

HPLC (High Pressure Liquid Chromatography) purified and lyophilized oligonucleotides were obtained from MWG Biotech Germany. Oligonucleotides used for analysis were partly labeled at the 5'end with biotin (B). In case of the single-stranded Linker used in the non-restrictive LAM-PCR (LC1rev), this oligonucleotide is phosphorylated at the 5'end and tagged with a di-desoxynucleotid at the 3' end. Sequences of all the oligonucleotides are listed below.

Standard PCR	**Primer Name**	**Sequence (5‘-3‘)**
GAPDH-PCR	GAPDH-Hs-Fwd	5‘-GAAGGTGAAGGTCGGAGT-3‘
	GAPDH-Hs-Rev	5‘-GAAGATGGTGATGGGATTTC-3‘
GAPDH-PCR	GAPDH-Mm-Fwd	5‘-ATATTGGAAAATGTGGAAGTGAAAGG-3‘
	GAPDH-Mm-Rev	5‘-TCGCCAGAGCCATCTTTTG-3‘
HPRT-PCR	HPRT down	5‘-GCTGGTGAAAAGGACCTCT-3‘
	HPRT middle rev	5‘-CCAACAACAAACTTGTCTGG-3‘
GFP-PCR	GFP-Fwd	5‘-TGAGCAAGGGCGAGGAGCTGTT-3‘
	GFP-Rev	5‘-GCCGGTGGTGCAGATGAACT-3‘
WPRE-PCR	WPRE-Fwd	5‘-GAGGAGTTGTGGCCCGTTGT-3‘
	WPRE-Rev	5‘-TGACAGGTGGTGGCAATGCC-3‘
Mds1/Evi1-PCR	ME-Mm-PR-Fwd	5‘-CAAGGAAACTGGCCACAAGT-3‘
	ME-Mm-PR-Rev	5‘-AGGAGAGCATGGCTCTTGAA-3‘
Evi1-PCR	cME-Mm-Fwd	5‘-ACCAGCCCCTGGATCTAAGT-3‘
	cME-Mm-Rev	5‘-AGCTTCAAGCGGGTCAGTTA-3‘
Prdm16-PCR	cPrdm16-Fwd	5‘-GCGGCCGTCCCCACTTCTGTT-3‘
	cPrdm16-Rev	5‘-GCGCCCTCGTGTTCGTGCTTCT-3‘
sPrdm16-PCR	Prdm16-PR-Fwd-3	5‘-TGGGAGATGCTGACGGATAC-3‘
	Prdm16-PR-Rev-3	5‘-TCACAGGAACACGCTACAC-3‘
Cloning PCR	**Name**	**Sequence (5‘-3‘)**
Prdm16/sPrdm16	Prdm16-388-C	5‘-GCCGACTTTGGATGGGAGATGCTGACGGATACAGAG-3‘
	E1S1Prdm16-A	5‘-GGAATTCCTGCAGGACCATGCGATCCAAGGCGAGG-3‘
	H3K1Prdm16-741D	5‘-CCCAAGCTTATTAGGCGCCTCAGGTCCAG-3‘
	Prdm16-388-B	5‘-CTCTGTATCCGTCAGCATCTCCCATCCAAAGTCGGC-3‘
	EIAI-Prdm16-C-F	5‘-GGAATTCCGACCGGTCATTCAGCATCTC-3‘
	XIAI-Prdm16-C-R	5‘-GCTCTAGAGCGATCGCTCAGAGGTGGTTGATGGG-3‘
	EISI-Prdm16S-N-F	5‘-GGAATTCCTGCAGGACCATGTGTCAGATCAACG-3‘
Setbp1	EISI-Setbp1-N-F	5‘-GGAATTCCTGCAGGACCATGGAGCCAGAGG-3‘
	XIKI-Setbp1-N-R	5‘-GCTCTAGAGCGGTACCACACTTCCCAAG-3‘
	EIKI-Setbp1-C-F	5‘-GGAATTCCACATGGCTCGGGAGG-3‘
	XIAI-Setbp1-C-R	5‘-GCTCTAGAGCGATCGCCTAGGGAAGGACATCACTCTC-3‘
Mds1/Evi1	EIBI-Evi1-N-F	5‘-GGAATTCCGGGATCCATGGCGCCTGACATCCA-3‘
	MINI-Evi1-N-R	5‘-CGACGCGTCTGCAATCAGCATG-3‘
LTR PCR	**Name**	**Sequence (5‘-3‘)**
LTR-Test (LV)	SK LTR 1 bio	(B) 5‘-GAGCTCTCTGGCTAACTAGG-3‘
	SK LTR 9 rev	5‘-GCTAGAGATTTTCCACACTG-3‘
	SK LTR 2 bio	(B) 5‘-GAACCCACTGCTTAAGCCTCA-3‘
	SK LTR 8 rev	5‘-CTAAAAGGGTCTGAGGGATC-3‘
LAM-PCR	**Name**	**Sequence (5‘-3‘)**
3’-LAM; LV	SK LTR 1bio	(B) 5‘-GAGCTCTCTGGCTAACTAGG-3‘
	SK LTR 2bio	(B) 5‘-GAACCCACTGCTTAAGCCTCA-3‘
	SK LTR 3bio	(B) 5‘-AGCTTGCCTTGAGTGCTTCA-3‘
	SK LTR 4bio	(B) 5‘-AGTAGTGTGTGCCCGTCTGT-3‘
	SK LTR 5	5‘-GTGTGACTCTGGTAACTAGAG-3‘
	SK LTR 5 1/2	5‘-GATCCCTCAGACCCTTTTAGTC-3‘
5’-LAM; LV	U3_Lenti_Bio_4	(B) 5‘-AGGCTTAAGCAGTGGGTTCC-3‘
	U3_Lenti_Bio_5	(B) 5‘-TTAGCCAGAGAGCTCCCAGG-3‘
	U3_Lenti_Bio_6	(B) 5‘-GATCTGGTCTAACCAGAGAG-3‘
	U3_Lenti_7	5‘-CCCAGTACAAGCAAAAAGCAG-3‘

5'-LAM; MLV RV	LTRa bio	(B) 5'-TGCTTACCACAGATATCCTG-3'
	LTRb bio	(B) 5'-ATCCTGTTTGGCCCATATTC-3'
	LTR II bio	(B) 5'-GACCTTGATCTGAACTTCTC-3'
	LTR III	5'-TTCCATGCCTTGCAAAATGGC-3'
3'-LAM; MLV RV	U5Ivbio	(B) 5'-TCCGATTGACTGAGTCGC-3'
	U5Vbio	(B) 5'-GGTACCCGTGTATCCAATA-3'
	U5VI bio	(B) 5'-TCTTGCAGTTGCATCCGACT-3'
	LTR25	5'-GTGGTCTCGCTGTTCCTT-3'
5'-internal LAM	Lenti 1 bio	(B) 5'-GATAGTAGGAGGCTTGGTAG-3'
	Lenti 2 bio	(B) 5'-AGTGAATAGAGTTAGGCAGG-3'
	Lenti 4 bio	(B) 5'-CACCATTATCGTTTCAGACCC-3'
	Lenti 5	5'-GGCCCGAAGGAATAGAAGAA-3'
	GFP1-Bio	(B) 5'-AAACGGCCACAAGTTCAGCG-3'
	GFP2-Bio	(B) 5'-GATGCCACCTACGGCAAGCT-3'
	GFP3-Bio	(B) 5'-TACGGCGTGCAGTGCTTCAG-3'
	GFP4-Bio	(B) 5'-AAGGGCATCGACTTCAAGGA-3'
	GFP6-Bio	(B) 5'-ACTACCTGAGCACCCAGTCC-3'
	GFP7-Bio	(B) 5'-GCCCTGAGCAAAGACCCCAA-3'
	GFP8-Bio	(B) 5'-GATCACATGGTCCTGCTGGA-3'
	GFP9-Bio	(B) 5'-TGACCGCCGCCGGGATCACT-3'
3'-internal LAM	Lenti-Rev1-Bio	(B) 5'-TTCTATTCCTTCGGGCCTGT-3'
	Lenti-Rev2-Bio	(B) 5'-GGAGGTGGGTCTGAAACGAT-3'
	Lenti-Rev3-Bio	(B) 5'-TGGTGAATATCCCTGCCTAA-3'
	Lenti-Rev4	5'-TTCTTAAACCTACCAAGCCTCCT-3'
	GFP-Rev1-Bio	(B) 5'-GAACTCCAGCAGGACCATGT-3'
	GFP-Rev2-Bio	(B) 5'-GCTTCTCGTTGGGGTCTTTG-3'
	GFP-Rev3-Bio	(B) 5'-ACTGGGTGCTCAGGTAGTGG-3'
	GFP-Rev4-Bio	(B) 5'-ATGGGGGTGTTCTGCTGGTA-3'
	GFP-Rev5-Bio	(B) 5'-AAGTCGATGCCCTTCAGCTC-3'
	GFP-Rev6-Bio	(B) 5'-GGTCTTGTAGTTGCCGTCGT-3'
	GFP-Rev7-Bio	(B) 5'-AGTCGTGCTGCTTCATGTGG-3'
	GFP-Rev8-Bio	(B) 5'-GAACTTCAGGGTCAGCTTGC-3'
Expo-PCR	LCI	5'-GACCCGGGAGATCTGAATTC-3'
	LCII	5'-GATCTGAATTCAGTGGCACAG-3'

Linker Cassette	Name	Sequence (5'-3')
Linker cassette for LAM-PCR	LC1	5'-ACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG-3'
	LC3 (AATT)	5'-AATTCCTAACTGCTGTGCCACTGAATTCAGATC-3'
	LC3 (TA)	5'-TACCTAACTGCTGTGCCACTGAAATCAGATC-3'
	LC3 (CATG)	5'-CATGCCTAACTGCTGTGCCACTGAATTCAGATC-3'
	LC3 (CG)	5'-CGCCTAACTGCTGTGCCACTGAATTCAGAT-3'
Linker cassette for nrLAM-PCR	LC1rev	(P)5´-CCTAACTGCTGTGCCACTGAATTCAGATCTCCCGGGTC-3´

3.1.13.2 Regents and kits molecular biological experiments

Reagents and Chemicals	**Manufacturer**
100bp DNA ladder	Invitrogen
1Kb plus DNA ladder	Invitrogen
Absolute ethanol	VWR

Acetic acid	Applichem
Agarose	Sigma-Aldrich
Agencourt AMPure XP PCR Purification Beads	Beckman Coulter
Aqua ad iniectabilia	Braun
Bovine Serum Albumine (BSA)	Sigma-Aldrich
Bromphenolblue	Sigma-Aldrich
Calf intestinal alkaline phosphatase (CIAP)	Invitrogen
Dimethylformamid (DMF)	Sigma-Aldrich
dNTPs Sets	Genaxxon
Dynabeads M-280 Streptavidin	Dynal, Invitrogen
EndoFree Plasmid Maxi Kit	Qiagen
Ethidium bromide	Applichem
Ethylenediaminetetraacetic acid (EDTA)	Applichem
Fast-Link DNA Ligation Kit	Epicentre Biotechologies
FastStart High Fidelity PCR System	Roche
Glycerol	Sigma-Aldrich
Guanidin hydrochloride	Applichem
Hexamin-$CoCl_2$	Fluka
Hexanucleotide mixture	Roche
High Pure PCR Product Purification Kit	Roche
High Pure PCR Template Preparation Kit	Roche
Human Genomic DNA	Roche
Klenow polymerase	Roche
Licium chloride	Sigma-Aldrich
LightCycler 480 SYBR Green I Master	Roche
LightCyler Uracil-DNA Glycosylase	Roche
Loading Buffer 5x	Elchrom Scientific
Magnesium chloride	Sigma-Aldrich
Microcon-30, -50	Millipore
Mid-Range PCR System	PeqLab
MS2 RNA (0.8µg)	Roche
Nonidet P-40 10%	Sigma-Aldrich
PCR SuperMix with *Taq* DNA polymerase	Invitrogen
PEG 8000	Genaxxon
Phosphate Buffered Saline (PBS, pH 7.5)	Gibco, Invitrogen
Platinum *Pfx* DNA polymerase	Invitrogen
ProofStart PCR System	Qiagen
Proteinase K	Roche
QIAGEN LongRange 2Step RT PCR System	Qiagen
QIAGEN Plasmid High-Speed Maxi Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel extraction Kit	Qiagen
QIAquick PCR Purification Kit	Qiagen
QIAshredder	Qiagen
Quickchange II XL site-directed mutagenesis kit	Statagene
RNAlater RNA stabilization reagent	Qiagen
Rnase*Zap*	Ambion
Rneasy Mini Kit	Qiagen
Sodium Acetate	Sigma-Aldrich
Sodium Dodecyl Sulfate (SDS)	Applichem

Sodium hydroxide	Fluka
Spreadex EL1200 gel mini/midi	Elchrom Scientific
SuperScript III First-Strand Synthesis SuperMix	Invitrogen
SuperScript III First-Strand Synthesis System for RT-PCR	Invitrogen
SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase	Invitrogen
T4 DNA-Ligase and 10x reaction buffer	New England Biolabs
T4 RNA-Ligase and 10x reaction buffer	New England Biolabs
Taq DNA polymerase and 10x PCR buffer	Genaxxon/ Qiagen
TOPO-TA cloning kit with pCR2.1 TOPO vector	Invitrogen
Tris-Acetat-EDTA (TAE) Buffer 40x	Elchrom Scientific
Tris-Borat-EDTA (TBE) Buffer 10x	Amresco
Tris-HCl (pH 7.5)	Applichem
TURBO DNA-*free*	Ambion
Tween-20	Sigma-Aldrich
Venor GeM Mycoplasma Detection Kit	Minerva Biolabs
X-gal	Applichem
β-Mercaptoethanol	Sigma-Aldrich

3.1.13.3 Buffers for LAM-PCR and PCR

Name	Chemicals	Final concentration
5x Blue run loading buffer	Tris-HCl (pH 7)	25mM
	EDTA (pH 8)	150mM
	Bromphenol-Blue	0.5% (v/v)
	Glycerol	25% (v/v)
Direct lysis buffer	10x PCR buffer	1x
	Tween-20	0.5%
	Nonidet P-40	0.5%
	Proteinase K	0.91mg/ml
Alkaline denaturation solution	NaOH (1M)	0.1M
Magnetic bead washing solution	PBS (pH 7,4)	
	BSA	0.1%(v/v)
Magnetic bead 3M LiCl washing and binding buffer	Tris-HCl (pH 7,5)	10mM
	EDTA	1mM
	LiCl	3M
Magnetic bead 6M LiCl washing and binding buffer	Tris-HCl (pH 7,5)	10mM
	EDTA	1mM
	LiCl	6M

3.1.13.4 Buffers and kits for southern blotting

Name	Manufacturer
PCR DIG Probe Synthesis Kit	Roche
Expand High Fidelity Plus PCR System	Roche
DNA Molecular Weight Marker VII, DIG-labeled	Roche
High Pure PCR Product Purification Kit	Roche
Positively charged Nylon Membranes	Roche
DIG Wash and Block Buffer Set	Roche

Anti-Digoxigenin-AP	Roche
CSPD, ready-to-use	Roche
DIG Easy Hyb	Roche
DIG Nucleic Acid Detection Kit	Roche
DNA depurination buffer (250mM HCl)	Laboratory stock
DNA denaturation solution (0.5M NaOH, 1.5M NaCl)	Laboratory stock
DNA neutralization solution (0.5M Tris-HCl, pH 7,5, 1.5M NaCl)	Laboratory stock
20x SSC buffer	Roche
2x SSC buffer	Laboratory stock
Salmon Sperm DNA (10mg/ml)	Roche
Low stringency buffer (2xSSC, 0.1%SDS)	Laboratory stock
High stringency buffer (0.5xSSC, 0,1%SDS)	Laboratory stock

3.1.14 *Regents, chemicals and buffers for protein analysis*

3.1.14.1 Buffers reagents and kits for western blotting

Name	Manufacturer
RIPA buffer (50mM Tris-HCl pH7.4, 0.1% NP-40, 0.5% Na-deoxycholate, 150mM NaCl, 0.1%SDS)	Laboratory stock
Protease inhibitor cocktail tablet	Roche
Ponceau S staining Solution	Sigma-Aldrich
Blocking Buffer (3% fat-free milk powder in wash buffer A)	Laboratory stock
Wash buffer A (0.9% NaCl, 10mM Tris-HCl, 0.05% Tween-20, pH 7.5)	Laboratory stock
Primary antibody dilution buffer (same to blocking buffer)	Laboratory stock
Secondary antibody dilution buffer (5% fat-free milk powder or BSA in wash buffer A)	Laboratory stock
Blot stripping buffer (0.2M Glycine, pH 2.5)	Invitrogen
Novex Tris-Glycine SDS Sample Buffer 2x	Invitrogen
Novex Tris-Glycine SDS Running Buffer (10X)	Invitrogen
Novex Tris-Glycine Transfer Buffer (25X)	Invitrogen
Novex Tris-Glycine Gel, 10 wells, 15 wells	Invitrogen
PageRuler unstained protein ladder	Fermentas
PageRuler prestained protein ladder plus	Fermentas
Novex sharp protein standard	Invitrogen
BCA Protein Assay kit	Thermo Fisher
SuperSigna WestPico chemiluminescent substrate	Thermo Fisher
Whatman 3 mm paper	Whatman GmbH
Immobilon-P Membrane (PVDF)	Millipore

3.1.14.2 Antibodies for western blotting

Name	Manufacturer
Mouse anti Mds1/Evi1 (N-terminus)	Abgent
Rabbit anti Evi1 (C-terminus)	Santa cruz biotechology
Rabbit anti GFP	Invitrogen
Rabbit anti GAPDH	Imgenex

Rabbit anti Prdm16 (N-terminus, 1446, 1447)	Biogenes GmbH
Rabbit anti Prdm16 (middle, 1448, 1449))	Biogenes GmbH
Rabbit anti Setbp1 (N-terminus, 1450, 1451)	Biogenes GmbH
Rabbit anti mouse IgG-HRP	Abcam
Rabbit anti goat IgG-HRP	Abcam
Goat anti rabbit IgG-HRP	Abcam

3.1.14.3 ELISA assay for HIV-p24 quantification

Name	Manufacturer
HIV-1 p24 ELISA kit	Perkin Elmer

3.1.15 *Laboratory instruments and disposables*

Name	Manufacturer
Balances	Sartorius
BD FACSAria cell sorter	Becton Dickinson
BD LSRII flow cytometer	Becton Dickinson
Cell culture dishes (3.5cm/6cm/10 cm/15cm)	Greiner-Bio-One
Cell culture flasks(25cm^2/75cm^2/150cm^2)	Nunc
Cell culture hood	Heraeus
Cell culture incubator	Heraeus
Cell culture plates (6wells/12wells/24wells/48wells/96wells)	Greiner-Bio-One
Cell scrapers	Corning
Centrifuge Tube 6ml	Becton Dickinson
Cryogenic Vials	Nalgene
Electrophoresis Power Supply	Elchrom Scientific
Eppendorf tubes 0.5ml, 1.5ml and 2ml	Eppendorf
Falcon 15ml and 50ml	Becton Dickson
Filter System 0.2 µm (250 ml/500 ml)	Millipore
Fluoresces microscope	Zeiss
Freezer -20°C	Liebherr
Freezer -80°C	Sanyo
Fridge 4°C	Liebherr
Gel documentation system	Peqlab
Gel electrophoresis apparatus	Peqlab
Gel electrophoresis comb (20, 24)	Peqlab
Gel electrophoresis Power Pac Supply 300	BioRad
Glasses and coverslips	Neolab
Heating block Thermo-mixer Comfort	Eppendorf
Heraeus Benchtop Multifuge 3SR Plus	Thermo Scientific
Horizontal shaker KS 260 basic	IKA
Ice machine	Ziegra
Incubator 37°C	Heraeus
LAM 96-well plate	Greiner-Bio-One
LAM 96-well plate sealing foil	ABgene
Latex Exam Gloves	Kimberly-Clark
LightCycler 96-well plate	Roche
LightCycler 96-well plate sealing foil	Roche

LightCycler480 system	Roche
Magnetic particle concentrator MPC 96	Dynal, Invitrogen
Magnetic particle concentrator MPC-E1	Dynal, Invitrogen
Microscope	Zeiss
Microwave	Siemens
Nanodrop ND-1000 spectrophotometer	Thermo Fisher
Needle and syringe	Becton Dickinson
Neubauer Hemacytometer	Hausser Scientific
Nitrogen tank	Thaylor-Wharton
PCR soft tubes 0.2ml	Biozym
pH-Meter	Knick
Picofuge centrifuge	Neolab
Pipetboy	Integra Bioscience
Pipettes (P2, P10, P100, P200, P1000)	Eppendorf
Polystyrene Round Bottom	Nalgene
Printer paper VM 65 H	Mitsubishi
Scales	Sartorius
Scalpels	Feather
Serological Pipette 2, 5, 10, 25, 50 ml	Corning
Submerged gel electrophoresis apparatus SEA 2000	Elchrom Scientific
Syringe Filter 0.2 µm and 0.45 µm	Millipore
Tabletop centrifuge (Microcentrifuge)	Eppendorf
Thermo cycler programmed with desired protocols	Biometra
TipOne Filter tips (10µl, 20µl, 100µl, 200µl, 1000µl)	Starlab
TipOne RPT tips (10µl, 20µl, 100µl, 200µl)	Starlab
Ultracentrifuge	Beckman Coulter
Video printer	Mitsubishi
Vortexer (MS1)	IKA
Water-bath	Thermo Fisher
Microplate Reader SpectraMax plus 384	Molecular Devices

3.1.16 *Computer Programs and Data Bases*

Name	**Manufacturer**
Adobe Acrobat 9 Professional	Adobe
Adobe Illustrator CS4	Adobe
Adobe Photoshop CS4	Adobe
BLAST Search	www.ncbi.nlm.nih.gov/blast
BLAT Search	genome.ucsc.edu/
Ingenuity Pathway Analysis Software	Ingenuity Systems
Lasergene	DNA Star
LightCycler LC480 Software (Version 1.5)	Roche
Office 2007 (World, Excel, PowerPoint)	Microsoft
Primer3	frodo.wi.mit.edu/primer3
R 2.7.1	cran.r-project.org
SoftMax Pro Software	Molecular Devices
Vector NTI	Invitrogen
SigmaPlot	Systac Software

3.2 Methods

3.2.1 *Cell biology methods*

3.2.1.1 Cell lines propagation

The following cell lines were used in this work: 293T, HeLa, SC-1 and C1498. 293T cells are a highly transfectable cell line derived from the 293 human embryonic kidney cell line that express the SV40 large T antigen. HeLa is an immortal cell line which derived from cervical cancer cells taken from Henrietta Lacks. SC-1 is a murine embryonic fibroblast cell line which has a normal karotype. C1498 is an atypical myeloid leukemia that originated in a C57BL/6 mouse and has been used as a model for acute myelogenous leukemia.

HeLa, SC-1 and C1498 cell lines are maintained in Dulbecco's modified eagle medium (DMEM) supplemented with fetal bovine serum (10% v/v), L-glutamine (4mM), penicillin (100µg/ml) and streptomycin (100µg/ml); 293T cells are maintained in Iscove's modified Dulbecco's medium (IMDM) with the same supplements. Cells were cultured in tissue culture treated cell flasks at 37°C, 5% CO_2 and 95% humidity atmosphere. Cells were passaged when confluence reached 90-95%. The adherent cell lines were first washed with Dulbecco's phosphate buffered saline (PBS) and then incubated for 2-5min at 37°C with 0.05% trypsin/EDTA, the cells were subsequently diluted 1:10 in fresh complete culture media and transferred to new tissue culture flasks. For the suspension cell line, cell suspension was centrifuged first for 5min at 1200rpm and RT, the cell pellets were then resuspended with fresh complete culture media, diluted 1:10 and transferred to new tissue culture flasks.

3.2.1.2 Cryo-conservation of cells

Cells from a 90% confluent monolayer in a 75cm^2 tissue culture flask were pelleted by centrifugation for 5min at 1200rpm and RT in a tabletop centrifuge, resuspended in 900µl complete culture media, transferred to a 2ml cryovial which

contains 900µl freezing medium (30% FBS, 15% DMSO in complete culture media) and mixed well. Cells were frozen slowly overnight to -80°C in an isopropanol freezing box and then transferred to liquid nitrogen.

To revive frozen cells, aliquots were thawed rapidly in a 37°C water bath and added drop wisely to 50ml thawing medium (50% FBS in complete culture media). The cells were pelleted at 1000rpm by centrifugation for 10min to remove the DMSO and then resuspended in 10ml complete culture media and transferred to a 75cm^2 tissue culture flask.

3.2.1.3 Determination of the number of cells

To determine the number of living cells, the cells were treated with trypsin-EDTA as described. An aliquot of cells was stained with Typan-Blue solution and counted in a hemocytometer under the microscope. The total number of cells was determined from the average number of cells in 4 large squares (each with 16 small squares), the factor 10^4, the dilution factor and the total suspension volume in ml.

3.2.1.4 Lentiviral vectors production

Integration proficient and deficient lentiviral vectors (LV) were produced by transient co-transfection of packaging plasmid pMDLgag/polRRE (for integration proficient LV) or pMDLgag/polintD64VRRE (for integration deficient LV), REV expressing plasmid pRSV-REV, envelop plasmid pMD2-VSVG and transgene expression vectors on 293T cells. Transfection was mediated by calcium phosphate ($CaPO_4$) or polyethylenimine (PEI). The procedure of LV production was performed under the biosafety level 2 (S2) conditions.

3.2.1.5 Calcium phosphate ($CaPO_4$) transfection

1x10^7 293T cells were seeded in 15cm cell culture disk one day before transfection. 2 hours prior to transfection, 22.5ml fresh growth medium were replaced. The DNA premix was prepared by mixing 9µg envelop plasmid, 12.5µg packaging plasmid, 6.25µg REV expressing plasmid and 28µg vector plasmid (32µg if

the length of transgene greater than 2.8kb) with 1125µl 0.1 x TE/dH_2O in a 15ml Falcon tube. 125µl 2.5M $CaCl_2$ were added to the DNA premix gently and mixed well. After 5min incubation at RT, 1250µl pre-tested 2xHBS with preferred pH value were added to DNA premix drop wisely while vortex. The precipitate was added to 293T cells immediately. The transfected cells were incubated in 37°C, 5% CO_2 and 95% humidity conditions. The transfection media were removed and replaced with 16ml fresh growth media 12-14 hours post transfection. 24 hours after media change, the supernatant were harvested and the cell debris were removed by filtration through a 0.22µm filter, and then concentrated by ultracentrifugation at 20,000rpm, 20°C for 2 hours. The virus particles were resuspended by adding 40µl Opti-MEM medium and incubated on RT for 20min. 10µl-20µl aliquots were then distributed and stored rapidly at -80°C.

3.2.1.6 Polyethylenimine (PEI) transfection

The DNA premix of PEI mediated transfection were made by adding the same amount of vector DNAs of $CaPO_4$ transfection and 440µl Opti-MEM medium. 180µl PEI (PEI:DNA=3:1) were mixed with 320µl Opti-MEM and added to DNA premix, the DNA/PEI complexes were incubated at RT for 30min, then added to 293T cells which seeded and prepared same to $CaPO_4$ transfection. The following steps remained the same to $CaPO_4$ transfection.

3.2.1.7 Virus titration

$5x10^4$ HeLa cells were seeded in 12-well cell culture plate in IMDM medium supplemented with 10% FBS, 4mM L-glutamine, 100µg/ml Pen/Strep one day before. A 10-fold dilution series of 2x viral vector stocks were generated at the day of titration. Briefly, for concentrated vector stock, 2µl of thawed virus was added to 1ml IMDM medium (10^{-3} dilution), further dilutions were achieved by adding 100µl of first 10^{-3} dilution with 900µl IMDM medium, until 10^{-7} dilution. HeLa cells were washed once with PBS and added with 0.5ml IMDM medium with 2x polybrene (16µg/ml). After adding 500ml of series dilution of virus stocks to each represent wells, HeLa cells were incubated at 37°C, 5% CO_2 and 95% humidity for 48 hours (for

integration deficient LV) of 72 hours (for integration proficient LV). FACS analyses were performed for functional titer using standard protocol for eGFP expression, p24 ELISA assays were performed for the total particles present in the virus stock. The titer of virus stock was calculated using followed formula:

GFP titer (TU/ml) = (number of cells at the time of vector addition) x (% GFP positive cells) x (dilution factor)

P24 titer (Virus particle/ml) = (pg of P24/0.01) x (dilution factor)

3.2.1.8 p24 antigen ELISA assay

The quantity of p24 antigen in lentiviral supernatants was measured using a Perkin Elmer HIV-1 p24 antigen assay as per the manufacture's instruction. Supernatants dilutions from 10^{-5} to 10^{-7} were added to p24 antibody-coated wells together with diluted calibration standards. Samples were lysed, washed and then incubated with biotinylated anti-HIV-1. The samples were washed again and subsequently incubated with streptavidin-HRPO followed by the addition of tetramethybenzidine and stop reagent (4N H_2SO4). The absorbance of each plate was then read at 490nm and the quantity of p24 reagent calculated from a standard curve generated from the diluted calibration standards.

3.2.1.9 Isolation of hematopoietic progenitors

Bone marrow was harvested from femurs and tibias of 8-12 week-old C57Bl/6J mice sacrificed with cervical dislocation. Hematopoietic progenitors were isolated using EasySep Mouse Hematopoietic Progenitor Cell Enrichment Kit of StemCell Technologies, according to manufacturer's recommendation. Briefly, hematopoietic stem and progenitor cells from single cell suspensions of bone marrow were isolated by negative selection. Unwanted cells are targeted for removal with biotinylated antibodies directed against non-hematopoietic stem and progenitor cells (CD5, CD11b, CD19, CD45R, 7-4, Ly-6G/C (Gr-1), TER119) and Tetrameric Antibody Complexes recognizing biotin and dextran-coated magnetic particles. Labeled cells

are separated using an EasySep magnet without the use of columns. Desired cells are poured off into a new tube.

3.2.1.10 Lentiviral transduction in cell lines and primary hematopoietic progenitors

For transduction of cell lines, $1x10^5$ cells were plated in a well of 12-well plate with 1ml complete growth medium supplemented with 8µg/ml Polybrene one day before. Different MOIs of LV were added into each well the day of transduction. Transduction efficiency was monitored by FACS after 72 hours for integration-proficient lentiviral vectors, and after 48 hours for integration-deficient Lentiviral vectors.

For primary hematopoietic progenitors, $1x10^5$ cells/ml fresh isolated lin^- cells was seeded in a 96 well-plate with round bottom. The transduction was carried out in IMDM medium supplemented with 15% FBS, 100ng/ml rmSCF, 100ng/ml rmFlt3L, 50ng/ml rmTPO and 4µg/ml Polybrene. MOI of 50 to 100 viruses were loaded into each well, and incubated in 37°C, 5% CO_2, 95% humidity for 16 to 20 hours.

3.2.1.11 Hematopoietic progenitors proliferation assay

After lentiviral transduction, BM cells were then expanded as mass cultures in IMDM supplemented with 10% FBS, 100IU/ml penicillin, 100µg/ml streptomycin, 2mM L-glutamine and 100ng/ml rmSCF, 6ng/ml rmIL-3, 10ng/ml rmIL-6 cytokine cocktail. During this time, cell density was adjusted to $5x10^5$ cells/ml every 3 days. eGFP positive cells were monitored every week by FACS, and cells number was counted after every medium change.

3.2.1.12 Colony forming cell assay

For the detection and quantification of mouse hematopoietic progenitors and mouse pre-B progenitor cells in BM, 200 and 600 sorted eGFP positive lin^- cells after lentiviral transduction were plated on 35-mm dishes in a 1.1 ml methylcellulose-based medium (M3434 for mouse hematopoietic progenitors and M3630 for mouse

pre-B progenitors, StemCell Technologies), according to manufacturer's instruction. The plates were then incubated at 37°C, 5% CO_2 and 95% humidity for 2 weeks. Colonies were counted under inverted light microscope after 14 days.

3.2.2 *Microbiology methods*

3.2.2.1 Growth and maintenance of E.coli

Escherichia coli (E.coli) were grown in liquid 1 x LB media at 37°C with agitation at 250rpm or streaked out on solid LB agar plates containing 1.5% bacto agar. *E.coli* transformed with plasmids was grown on the same media supplemented with appropriate antibiotics. For long-term storage, bacterial cultures were stored in 25% volume for volume (v/v) of glycerol in LB media at -80°C.

3.2.2.2 Transformation of chemical-competent E.coli

To transform chemical competent bacteria, chemical competent bacteria were slowly thawed on ice and 100pg to 100ng Plasmid DNA were added to each aliquot. This mixture was left on ice for thirty minutes. Bacteria were heat shocked for 30sec at 42°C and immediately put back on ice for 2min. 250µl of S.O.C medium was added and bacteria recovered in a shaking incubator at 37°C for 1 hour. Bacteria were then spread on LB agar plates containing the appropriate antibiotics and incubated overnight at 37°C.

3.2.3 *Molecular biology methods*

3.2.3.1 Plasmid DNA preparation

Plasmid DNA samples were prepared using Qiagen Plasmid Purification Kits. Small-scale plasmid DNA was extracted from a 4mL bacteria culture grown overnight at 37°C using the Qiagen QIAprep Spin Miniprep Kit as per manufacture's instruction. Large-scale plasmid DNA was obtained from 400ml bacteria cultures using a Qiagen plasmid high-speed maxi kit or EndoFree plasmid maxi kit, respectively, following the manufacture's instruction. The procedure is based on alkaline lysis of bacterial cells followed by binding of plasmid DNA to the Qiagen Anion-Exchange Resin under low-

salt and pH conditions. RNA, proteins, dyes, and low-molecular-weight impurities were removed by a medium-salt wash. Plasmid DNA was eluted in a high-salt buffer and then concentrated and desalted by isopropanol precipitation. The resulting plasmid DNA was dissolved in 30µl (Mini prep) or 500µl ddH_2O (Maxi prep).

3.2.3.2 Genomic DNA isolation from cultured cells

Genomic DNA was isolated from ~ $1x10^7$ cells and purified using High Pure PCR Template Preparation Kit (Roche) following the manufacture's instruction. Cell pellets were resuspended in 200µl 1x PBS and lysed with 40µl Proteinase K in 200µl Binding Buffer during an incubation period of 10 min at 70°C to directly inactivate all nucleases. Following the addition of 100µl isopropanol, the cellular DNA was bound to glass fibres pre-packed in the High Pure Purification Filter tube and purified in a series of 'wash and spin' steps by pipetting either 500µl Inhibitor Removal Buffer and 500µl Washing Buffer to the upper reservoir of the Filter tube and centrifuge the tube for 1min at 8,000 x g to get rid of contaminating cellular components. Finally, the purified DNA was recovered by applying three times 50µl ddH_2O to the column and incubating the ddH_2O on the column for 1min at room temperature before eluting the DNA by centrifugation for 1 min at 8,000 x g.

3.2.3.3 RNA preparation

Total RNA samples were prepared from various cell lines using TRIzol reagent following the manufacturer's instructions. One ml TRIzol was added to $1x10^7$ cells to lyse and homogenize the cells. 200µl chloroform were added to the homogenate and centrifuged, the upper aqueous phase was transferred to a new tube and total RNA was precipitated with 500µl isopropyl alcohol, following one wash with 75% ethanol. The RNA pellet was dissolved in Nuclease-free water. RNA was stored at -80°C until use.

3.2.3.4 Spectrophotometric determination of DNA and RNA concentration

The concentration and quality of DNA and RNA samples were calculated by measuring the absorbance of light with a wavelength of 260nm using a NanoDrop

ND-1000 spectrophotometer with a 0.2mm path length. Absorbency (optical density) of 1.0 of DNA equals a solution containing 50µg of double strand DNA per ml, or 40ng of RNA, respectively. The ratio of A260/A230 should be arranged between 1.7-1.9 for DNA samples and 1.9-2.1 for RNA samples.

3.2.3.5 DNase Treatment of RNA samples

After RNA preparation, 1/10 volume of the reaction buffer and 0.1U/µl TURBO DNase were added to the RNA solution, incubated at 37°C for 30min. Then, 1/10 volume of DNase remove reagent was pipetted into the reaction mixture, incubated for 2min at RT, centrifuged, and the upper aqueous phase which contains pure RNA samples was transferred into a new RNase-free tube.

3.2.3.6 Cloning

3.2.3.6.1 Restriction enzyme digestion

Plasmid DNA (typically 1-2µg) was digested in a final volume of 50µl of 1x buffer (supplied by the manufacturer) and bovine serum albumin (0.1mg/ml). The amount of enzyme used varied depending of the concentration of the enzyme stock and the amount of DNA, but never exceeded 10% (v/v) of the total reaction volume. The endonuclease reaction was carried out at the manufacturer's recommended temperature for 1 hour and DNA digestion was verified by agarose gel electrophoresis. Double or triple digestions were performed either in compatible buffers or sequentially, after clean-up of DNA by ethanol precipitation using the QIAquick PCR purification kit by following the manufacturer's protocol.

3.2.3.6.2 Polymerase Chain Reaction (PCR)

Invented in 1983 by Mullis and co-workers, PCR is a standard technique for *in vitro* enzymatic amplification of a specific segment of DNA. PCR reaction mixtures were set up using disposable tips containing hydrophobic filter with the aim to minimize cross-contamination. Per 50µl PCR reaction, template DNA (≤1µg/reaction) was pipetted with a master mix consisting of 5µl of 10x PCR buffer (Qiagen), 1µl of deoxyribonucleoside triphosphates (dNTPs, 10mM each), 1µl of each forward and

reverse primer (10pmol each), and variable amount of ddH_2O to give a final volume of 50µl mixture. A negative control without template DNA was included. The amplification was performed in a PCR thermal cycler using relatively cycling program optimized for each target and primer pair. In most cases, the amplification reaction was started by an initial denaturation at 95°C for 2min, followed by 25 to 35 cycles of denaturation at 95°C for 45sec, annealing at 56°C to 60°C for 45sec and extension at 72°C for 1min, with a final extension step at 72°C for 4min. PCR products were separated by electrophoresis on 0.8-2% agarose gels stained with ethidium bromide.

3.2.3.6.3 DNA agarose gel electrophoresis

DNA fragments were resolved by electrophoresis through 0.8-2% agarose gels in 1 x TAE buffer or TBE buffer. To prepare the gels, agarose was dissolved in 1xTAE buffer by boiling in a microwave. After cooling, ethidium bromide was added (to obtain a final concentration of 0.5g/ml) for visualization of DNA. DNA samples were mixed with DNA loading buffer before loading onto agarose gels. 5µl 1kb plus DNA ladder or 100bp DNA ladder was included in each gel to enable size determination of DNA fragments. Gels were electrophoresed using a voltage of 100-130V (up to 150mA) and the separated DNA fragments subsequently visualized by exposure to ultra-violet light using PeqLab gel documentation system.

3.2.3.6.4 Extraction of DNA fragments

Following electrophoresis, DNA fragments were excised from agarose gels using a clean scalpel blade under ultra-violet light. The DNA was then extracted from the agarose using a QIAquick gel extraction kit as the manufacturer's instructions.

3.2.3.6.5 Ligation of DNA fragments

When the digested vector DNA ends had compatible termini with the insert DNA, the digested plasmid vector DNA was treated with calf intestinal alkaline phosphatase (CIAP) to dephosphorylate the DNA ends prior to ligation. Dephosphorylation reactions were performed directly in restriction endonuclease buffers, immediately following digestion by adding 10 units of CIAP enzyme to the reaction mixture and incubating at 37°C for one extra hour. Ligation of DNA

fragments were performed using vector to insert ratios of 1:3 (v/v) in a final volume of 10µl of 1 x T4 DNA Ligase buffer containing 1 unit of T4 DNA Ligase. The vector DNA concentration used was typically 20-50ng of DNA. Ligation reactions were incubated overnight at room temperature or 16°C. The ligation reactions were either immediately transformed into *E.coli* or stored at -20°C for later use.

3.2.3.7 Reverse transcription PCR (RT-PCR)

3.2.3.7.1 First strand cDNA synthesis for RT-PCR

First strand cDNA was synthesized using SuperScript III First-Strand Synthesis System for RT-PCR. For each reaction, 0.5-1.0µg total RNA were mixed with 1µl of 100µM oligo(dT)18 primers, 1µl of 10mM dNTP and distilled water to 12µl, denatured at 65°C for 5min and then chilled on ice. Then 4µl of 5x First-Strand Buffer, 2µl of 0.1M DTT, and 1µl of RNase Block were added and mixed well. The mixtures were incubated at 42°C for 2min, 1µl of SuperScript III reverse transcriptase was added and mixed by pippetting. Mix contents were incubated at 42°C for 50min, then enzymes were inactivated at 70°C for 15min and the mixtures were cooled down to 4°C. Synthesized cDNAs were stored at -20°C.

3.2.3.7.2 Two-step RT-PCR

To verify transgene expression mediated by lentiviral vectors in transduced cell lines, 2µl first strand cDNA was used as template for transgene expression detection. The samples which synthesized without adding the reverse transcriptase for detecting the cross over genomic DNA contamination (here referred to non-RT control NRT) and DNase and RNase free H_2O were used as negative control (NC), as well as the GAPDH primer pairs were used in positive control samples. PCR amplification was performed as described in 50µl reaction volume containing primer pairs for gene of interested for 25 cycles.

3.2.3.7.3 One-step RT-PCR

To avoid the cross over contamination in the two-step RT-PCR, both cDNA synthesis and PCR amplification were performed in a single tube using SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase. cDNA synthesis was

performed by 1 cycle at 50°C for 30min. Amplification was carried out by denaturation at 94°C for 15sec (2 min in the first cycle), annealing at 56°C for 30sec, and extension at 68°C for 1min (5min in the last cycle) for 25 to 30 cycles.

3.2.3.8 Site directed mutagenesis

To repair the mismatched nucleotides in plasmid DNA, Site directed mutagenesis was performed using the Quickchange II XL site-directed mutagenesis kit from stratagene. The primer design was performed according to the manufacturer directions. The PCR reaction was performed in a total volume of 50µl, containing a total of 100ng of template DNA, buffer (1x), 10pmol of primer forward and primer reverse, 1µl of dNTP mix, 3µL of quick solution reagent and 2.5U of *Pfu Turbo* DNA polymerase. All reactions followed the thermal cycling parameters: 1min at 95°C followed by 16 cycles of 30sec at 95°C, 1min at 55°C and 11min at 68°C. After the PCR reaction, the parental DNA was digested with *Dpn* I enzyme for 3 hours. The resulting DNA was then transformed in chemical competent bacteria. DNA was extracted from single clones and sequenced to confirm correct mutagenesis.

3.2.4 *Integration site analysis by linear amplification mediated PCR*

To characterize the physiology of the gene-modified hematopoietic repopulation and the vector-related influences on clonal contribution, the highly sensitive linear amplification-mediated polymerase chain reaction (LAM-PCR) was performed as previously. Using this method it is possible to amplify vector-genome junctions. The initial linear PCR with 5’ biotinylated primers hybridizing to the U3- and/ or U5 region of the vector LTR facilitates the pre-amplification of the vector genome junctions. Magnetic capture of the amplified biotinylated target DNA on streptavidin-coupled paramagnetic beads allows enrichment of the target DNA. The following steps are carried out on the semisolid streptavidin phase. After a hexanucleotide primed double strand synthesis using the Klenow-Polymerase and a restriction digest, a restriction site complementary linker cassette of known sequence is ligated on the genomic end of the fragments. Subsequently, the non-biotinylated strands are

separated from the biotinylated strands in an alkaline denaturation step and amplified in two nested PCR reactions with linker- and vector-specific primers.

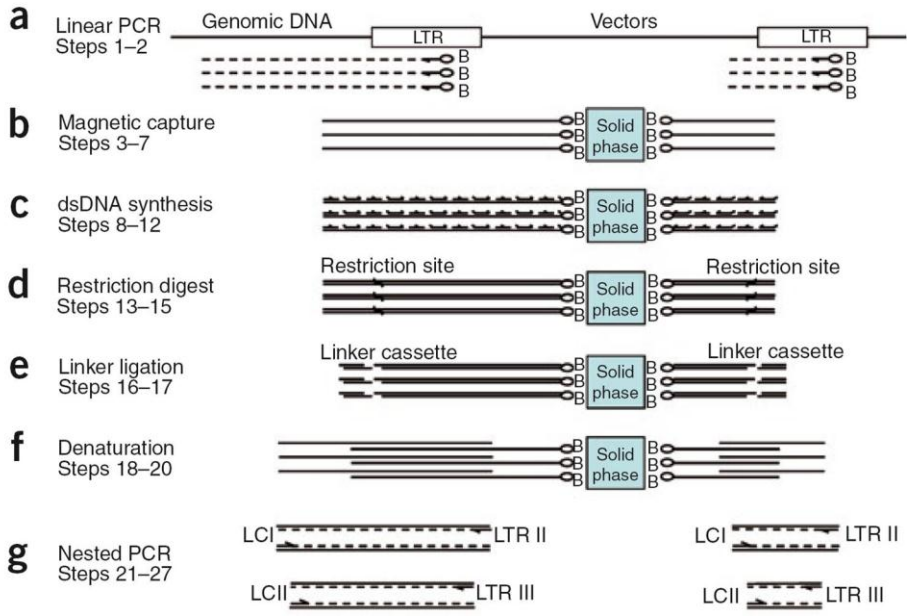

Figure 3-1 Schematic outline of LAM-PCR to amplify 5'-LTR retroviral vector-genomic fusion sequences (Schmidt et al., 2007).

3.2.4.1 Linear PCR

For detection of lentiviral integration sites, DNA was extracted and pre-amplified by repeated primer extension using vector-specific 5' biotinylated primers. Depending on the transduction efficiency and the available amount of template DNA, between 100ng and 500ng DNA sample were applied in 2 rounds of a 50-cycle linear PCR reaction using 5' biotinylated primers. 1.67nM of a 5' biotinylated primer is applied to the linear amplification reaction, if one vector-specific LTR primer is used, whereas 0.83nM of a 5' biotinylated primer is added in case of two vector-specific LTR primers. The amplification was performed using one initial denaturation cycle of 95°C for 2min, followed by 50 cycles of 95°C for 45sec, 60°C for 45sec and 72°C for 1min. After the first 50 cycles of the PCR, 0.5µl (2.5U) *Taq* polymerase were added and a second 50-cycle amplification was carried out using the same PCR program.

3.2.4.2 Magnetic Capture

200µg of magnetic beads per reaction were exposed for 60sec on a magnetic particle concentrator (MPC) and the supernatant was discarded. The beads were resuspended in 40µl PBS containing 0.1% BSA, exposed to the MPC and the supernatant discarded. This step was repeated once. Following a washing step with 20µl 3M LiCl buffer, the beads were resuspended in 50µl of 6M LiCl buffer. The magnetic bead solution was transferred with a ratio of volumes 1:1 (50µl:50µl) to the reaction tube containing the linear PCR product. The samples were incubated ON at RT on a horizontal shaker at 300rpm to link the biotinylated PCR products to streptavidin-coupled beads, which termed as DNA-bead complexes in the following chapters.

3.2.4.3 Hexanucleotide Priming

After overnight incubation, the DNA-bead complexes were collected on an MPC for 60sec, the supernatant discarded and once washed in 100µl ddH_2O. The DNA-bead complexes then were resuspended in 20µl hexanucleotide priming mixture containing 1 x hexanucleotide mixture, 200µM dNTPs and 2U klenow enzyme. The reaction was incubated at 37°C for exact 1 hour in a thermo cycler to generate double strand DNA.

3.2.4.4 Restriction Digest

80µl of ddH_2O was added to the hexanucleotide priming mixture and once washed in 100µl ddH_2O. The DNA-bead complexes were resuspended in 20µl restriction digest mixture containing 1 x restriction enzyme buffer, 1U restriction enzyme and 1 x BSA if needed. The restriction enzyme was chosen in a way that no restriction site was located within the known sequence of interest/the amplified part of the vector. The restriction digest mixture was incubated at the temperature recommended by the manufacturer to achieve maximum enzyme activity for 1 hour in a thermo cycler.

3.2.4.5 Linker Cassette Ligation

To ligate a known oligonucleotide sequence to the unknown flanking sequence, the DNA-bead complexes were washed and resuspended in 100µl ddH_2O, and then 100pmol of a double-stranded asymmetric linker cassette and Fast-link ligase (2U) were applied to the beads and incubate at room temperature for 5min.

3.2.4.6 Generation of a Linker Cassette

To generate a unidirectional double-stranded linker cassette, oligonucleotides were used which form overhangs after their hybridization complementary to the emerging overhangs of the restriction enzyme used in the LAM-PCR. 40µl 100pmol/µl primer LC1, 40µl 100pmol/µl enzyme depended LC3 primer, 110µl 250mM Tris-HCl (pH7.5) and 10µl 100mM $MgCl_2$ incubated at 95°C for 5min in a heating block and followed by cooling down the reaction mixture overnight in the heating block. The next day, 300µl ddH_2O were added to the linker cassette formation reaction and transferred on a Microcon-30 column. The sample was concentrated by centrifugation at 12,600rpm for 12min. To elute the linker cassette, 10µl ddH_2O was applied to the column and the reversed column placed onto a clean tube was centrifuged at 3,600rpm for 3min. The volume of the concentrated sample was adjusted with ddH_2O to a final volume of 80µl and frozen in small aliquots at -20°C to thaw the linker cassette only once.

3.2.4.7 Alkaline Denaturation

After 90µl ddH_2O was added to the ligation reaction, the DNA-bead complexes were exposed to an MPC and the supernatant was discarded. Following a washing step with 100µl of ddH_2O, the beads were resuspended in 5µl of fresh 0.1 N NaOH and the denaturation reaction was incubated for 10 min at room temperature on a horizontal shaker at 300rpm. The DNA-bead complexes were collected using an MPC and the supernatant that contained the non-biotinylated ssDNA was transferred to a fresh 0.5ml reaction tube. 2µl of the supernatant (ssDNA/ NaOH fraction) were applied as template for the 1st exponential PCR reaction.

3.2.4.8 Exponential PCRs

Both exponential PCRs were accomplished with nested vector- and linker cassette (LC)-specific primers. In the first exponential amplification of the vector-genome junctions, a 5'/3' biotinylated vector-specific LTR primer was used in conjunction with the linker cassette primer LCI. After an additional magnetic capture purification step, 2% of the 1st PCR product served as template for a 2nd nested exponential PCR with the linker cassette primer LCII and a 5'/3' vector-specific LTR primer.

3.2.4.9 Magnetic Capture after 1st Exponential PCR

200µg of magnetic beads per reaction were exposed for 60sec on an MPC and the supernatant was discarded. The beads were washed twice with 40µl of PBS supplemented with 0.1% BSA. Following a washing step with 20µl 3M LiCl buffer, the beads were resuspended in 20µl of 6M LiCl buffer. 20µl of the 1st expo PCR product was then incubated with 20µl of the magnetic bead solution for 1h at RT on a horizontal shaker at 300rpm. 60µl ddH_2O was added to the reaction, the DNA-bead complexes were collected using an MPC and the supernatant discarded. Following two washing step of the DNA-bead complexes with 100µl of ddH_2O, the beads were resuspended in 10µl of fresh 0.1N NaOH and the denaturation reaction was incubated for 10min at RT on a horizontal shaker at 300rpm. The DNA-bead complexes were exposed to an MPC and the supernatant containing the non-biotinylated ssDNA was transferred to a fresh 0.5ml reaction tube. 1µl of the supernatant (ssDNA/ NaOH fraction) was applied as template for the 2nd exponential PCR reaction.

3.2.5 *Integration Site Analysis by non-restrictive LAM- PCR*

All currently used integration site analysis methods rely on the use of restriction enzymes. Dependent on the distance between the vector integration site and recognition sequence of the enzyme used, each restriction motif introduces an integration site detection bias. Therefore, we developed the non-restrictive (nr) LAM-PCR in our group which is capable of covering all accessible integrant locations without the use of restriction enzymes. The linear PCR using biotinylated LTR primers

allows the initial pre-amplification of the vector-genome junctions. After removal of non-target DNA via magnetic selection, an ssDNA linker is ligated to the unknown genomic part of the ssPCR product, allowing its subsequent exponential amplification by nested PCRs.

3.2.5.1 Linear PCR (nrLAM-PCR)

200ng-1µg DNA was amplified using LTR-specific biotinylated primers that are identical to those used in the LAM-PCR approach. After the first 50-cycles of linear PCR, 0.5µl (2.5U) *Taq* polymerase were added per sample and a second 50-cycle amplification using the same program was carried out.

3.2.5.2 Magnetic Capture (nrLAM-PCR)

PCR products were subsequently purified with Microcon YM-50 to remove DNA fragments smaller than 125 nucleotides. The volume of each PCR sample was adjusted with ddH_2O to a final volume of 500µl, transferred on an YM-50 column and centrifuged at 12,600rpm for 12min. To elute the concentrated PCR product samples, 10µl ddH_2O was applied to the column, the reversed column placed onto a clean tube was centrifuged at 3,600rpm for 3min and ddH_2O was added to a final volume of 50µl. Selection of the biotinylated extension products was performed with 200µg of magnetic beads as described. The DNA-beads complexes were transferred into a 96-well plate and incubated at room temperature on a horizontal shaker at 300rpm for 2 hours.

3.2.5.3 Ligation of the Single-Stranded Linker Cassette (nrLAM-PCR)

The DNA-beads complexes were exposed to an MPC, the supernatant discarded and washed once with 100µl ddH_2O. Following resuspension of the beads in 10µl T4-RNA-ligase mixture (20U T4 RNA-ligase, 100pmol LC1-rev, 5µl 50% PEG 8000 and 1mM Hexamin-$CoCl_2$), the reaction mixture was incubated ON at RT on a horizontal shaker to allow direct ligation of a single-stranded linker DNA to the unknown genomic end of the amplicon.

3.2.5.4 Exponential PCRs (nrLAM-PCR)

The following day, direct amplification of the linker ligated ssDNA was performed. First, 90µl ddH_2O was added to the DNA-bead complexes and washed once with 100µl ddH_2O. Following resuspension of the DNA-bead complexes in 10µl ddH_2O, 2µl of this suspension were directly applied to the 1st exponential PCR. After additional magnetic capture purification step as described above, 4% (2µl) of the 1st exponential PCR product served as template for a 2nd nested exponential PCR. In contrast to the 1st exponential PCR, the 2nd exponential PCR was completed with a final elongation step at 72°C for 2min. According to standard LAM-PCR, the same primers were used to perform the two exponential PCRs

3.2.6 *Purification and Concentration of PCR Products*

The LAM-PCR and nrLAM-PCR amplicons were purified with QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instruction to sequence the vector genome junctions by either Sanger or 454 sequencing technology or both. Therefore, 5 volumes of Buffer PB were added to 1 volume of the PCR sample (e.g. add 200µl PB buffer to 40µl PCR product), mixed thoroughly and transferred to the QIAquick column to bind DNA. Following centrifugation at 13,000rpm for 1min, the flow-through was discarded and the samples were washed twice with 750µl of a 35% guanidine hydrochloride solution (v/v) and once with 750µl Buffer PE. After an additional centrifugation step for 1min at 13,000rpm to remove residual ethanol from the Buffer PE, the DNA was eluted into a clean 1.5ml tube by applying 30µl ddH_2O to the center of the QIAquick membrane, letting the column stand for 1min before centrifugation for 1min at 13000rpm.

PCR products were purified with High Pure PCR Product Purification Kit (Roche) following the manufacturer's recommendation to further create the standard curve for the LightCycler480 system. After PCR amplification the total volume for each PCR sample is adjusted with ddH_2O to 100µl. Following addition of 500µl Binding Buffer, the sample is transferred to the upper reservoir of the Filter Tube, centrifuged at 13,000rpm for 1min and the flow-through discarded. The samples were washed

twice with 500µl and 200µl Wash Buffer to ensure optimal purity before eluting the DNA into a clean 1.5 ml tube with 30µl Elution Buffer.

To purify and concentrate PCR products, Microcon YM-50 was used. The volume of the PCR products was adjusted with ddH_2O to a final volume of 500µl, transferred on an YM-50 column and centrifuged at 12,600rpm for 12min. To elute the concentrated PCR product samples, 10µl ddH_2O was applied to the column and the reversed column placed onto a clean tube was centrifuged at 3,600rpm for 3min. ddH_2O was added to a final volume of 15.5µl.

3.2.7 *Sequencing*

Incipiently, LAM-PCR /nr LAM-PCR products were purified, shotgun cloned into TOPO TA vector, transformed into competent cells, colonies picked and sequenced using the Sanger method. Nowadays, the LAM-PCR amplicons were purified and prepared for the 454 pyrosequencing technology which is performed either at GATC (Konstanz, Germany) or at the Genomics and Proteomics Core Facility (DKFZ, Heidelberg, Germany).

3.2.7.1 TOPO TA Cloning - Preparation for Sanger Sequencing Method

Cloning of the purified LAM-PCR/ nrLAM-PCR products was performed using TOPO TA Cloning Kit. 4µl of the purified PCR products was added to a mixture containing 1µl salt solution, 0.7µl TOPO vector and 0.3µl ddH_2O and incubated at RT for 4 ½min before stop on ice. For each transformation, one vial (50µl) of One Shot TOP10 chemically competent cells were thawed on ice, added to 3µl of the ligation reaction and incubated on ice for 30 min. Following heat-shock for 30 s at 42°C, the reaction was stopped on ice; 250µl S.O.C medium was added and shaked at 37°C for 1h at 300rpm. 75µl from each transformation was spread on pre-warmed LB-agar plates containing 50µg/µl Kanamycin and 40µl X-gal (40mg/ml) selective LB-plate and incubated overnight at 37°C. The next day, white colonies of the Blue/White selection system were picked into 96-well plates (one colony per well) which contained LB soft-agar and sent for sequencing to GATC (Konstanz, Germany).

3.2.7.2 Pyrosequencing Using the 454 GS Flx Platform (Roche)

DNA-sample preparation of the final LAM-PCR and/or nrLAM-PCR amplicons were performed as suggested by the manufacturer, consisting of an additional PCR amplification reaction with special Fusion Primers. Primer design was done as recommended by 454 Life Science (Roche). In brief, the fusion primers (FuP) consist of a 20-25bp linker or LTR specific sequence. In addition to a 19bp fixed sequence at the 5' end, the vector-specific LTR primer contains an individual recognition sequence (barcode) of 2 to 6 bases at the 3'end. The individual barcodes differ in at least 2bp to exclude misinterpretation of the data due to sequencing mistakes. Primer FuP-A was fused to a vector-specific LTR primer and primer FuP-B was fused to a linker-specific primer. The unique barcodes in the LTR-specific FuP-A allow the parallel analysis of various samples in a single sequencing run. Having purified the LAM-PCR/ nrLAM-PCR amplicons with the PCR purification Kit according to the manufacturer's description (Qiagen Kit) including a guanidine hydrochloride step, 40ng DNA were amplified in 12-PCR cycles with the special Fusion primers (linker and LTR, 10pmol/µl each). After a second PCR purification, DNA concentration was measured photometrical by Nanodrop and then subjected to pyrosequencing as implemented by 454 Life Sciences. 454 pyrosequencing was performed either at GATC (Konstanz, Germany) or at the Genomics and Proteomics Core Facility (DKFZ, Heidelberg, Germany).

3.2.7.3 Sequence Analysis

Originally, the sequencing results were trimmed manually by eliminating sequences derived from the linker cassette, the viral vector and/ or the plasmid used for sequencing (Lasergene software). After the alignment of the trimmed sequences (Lasergene software), the genomic consensus sequences were aligned to the mouse genome (assembly July 2007, NCBI37/mm9) using University of California, Santa Cruz (UCSC) BLAT genome browser

For a faster and more efficient analysis of huge data volumes, our group established a bioinformatics program. This program allows the analysis of

untrimmed and trimmed integration site sequences and also allows alignment of the genomic sequences of the vector-genome junction to the human genome using BLAT. Information as the sequence count (retrieval frequency; frequency of mappable integration sites), genomic length, span, identity, chromosome, integration locus, vector orientation in the genome, distance to the next RefSeq gene or in the gene to the TSS, intron/exon, gene length, RefSeq ID and missing LTR bases is analyzed and exported to an excel sheet for further analysis. Integration sites were judged to be authentic if the sequences began within 3bp of vector LTR ends, had either a >95% sequence match or no more than one base mismatch if the read length was <50bp, and had a unique best hit when aligned to the human genome using BLAT.

3.2.8 *Quantitative real time PCR*

Absolute quantitative real-time PCR was performed to detect and quantify the copy number of IDLV and LV in transduced samples. 2µg plasmid DNA of pCCL pptSFFV-IRES-GFP-PRE vector was linearized by EcoRV digestion and purified using High Pure PCR product purification kit. The concentration of purified linearized plasmid DNA was determined by Nanodrop spectrophotometer. A series of 10-fold dilutions in 20ng/µl MS2 RNA, starting at 2 x 10^{10} down to 2 x 10^{-1} copies/ 5µl were made for standard curve. PCR product was amplified by the complementary primers to WPRE sequence. For negative controls, 5µl H20 and 5µl genomic DNA from untransduced cells were used. The amplification was performed using one cycle of 95°C for 10min, followed by 50 cycles of 95°C for 10sec, 56°C for 5sec and 72°C for 25sec in Roche Light Cycler 480 with SYBR Green I Master Mix. The average of integrated viral vectors number in analyzed was estimated by interpolation the WPRE copies form each DNA sample in the standard curve.

4. Results

4.1 MDS1/EVI1 and PRDM16 are preferred gamma-retroviral integration loci in Wiskott-Aldrich syndrome clinical gene therapy

In the first gamma-retroviral clinical gene therapy trial for WAS, which was conducted and accomplished in Hannover medical school, leading by Prof. Dr. Christoph Klein and colleagues, 9 of 10 young patients were successfully treated with HSC gene therapy. Unfortunately, 3 patients developed T-cell acute lymphoblastic leukemia related to *LMO2* integrations. Samples from the first 2 treated patients were analyzed and reported in this dissertation.

Autologous CD34+ HSCs were collected by leukapheresis and then transduced with WASP-expressing gamma-retroviral vectors derived from MFG Moloney murine leukemia virus (CMMP) and pseudotyped with gibbon ape leukemia virus (GALV) envelope protein. The transduced cells were then reinfused 4 days later after myeloablation with busulfan conditioning. After gene therapy, sustained expression of WAS protein in HSC, lymphoid and myeloid cells, as well as platelets were found. T and B cells, nature killer (NK) cells, and monocytes were functionally corrected. After treatment, the patients' clinical condition markedly improved, with resolution of hemorrhagic diathesis, eczema, autoimmunity, and predisposition to severe infection (Boztug et al., 2010).

4.1.1 *Clonality analysis of gene-modified cells*

Almost 2.5 years after gene therapy, there were no morphologic or cytogenetic aberrations observed in patients' BM aspirates. Large-scale analysis of retroviral vector insertions sites (RISs) with LAM PCR combined with 454 pyrosequencing was performed to monitor the clonal distribution and fate of gene-corrected cells in vivo. In both patients, a sustained highly polyclonal reconstitution pattern was demonstrated in BM cells, primary blood leukocytes, sorted lymphoid CD3+ T cells,

as well as myeloid CD15+ granulocytes (Figure 4-1, A and B). 5709 unique RISs (in Patient 1) and 9538 unique RISs (in Patient 2) were retrieved and could be mapped to a single position in the human genome. Vector integration occurred preferentially in the vicinity of transcription start sites and into gene coding regions of the human genome, reflecting a typical gammaretroviral target-site distribution (Boztug et al., 2010).

As the LAM-PCR approach relies on the use of restriction enzymes and thus only allows the identification of a certain fraction of genomic integration sites, in this thesis the patient samples were additionally analyzed by non-restrictive (nr) LAM-PCR, which was newly developed in our group, for less sensitive but more unbiased retrieval of insertion sites. The nrLAM-PCR allows us to retrieve and identify the retroviral integration sites genome-widely in a single reaction, circumventing the detection bias accompanied by methods dependent on restriction enzymes (Gabriel et al., 2009; Paruzynski et al., 2010). Standard LAM-PCR produces amplicons of a defined length for each vector-genome junction, whereas nrLAM-PCR results in a variety of amplicon lengths for each insertion, appearing as a 'smear' on electrophoresis (Figure 4-1, C).

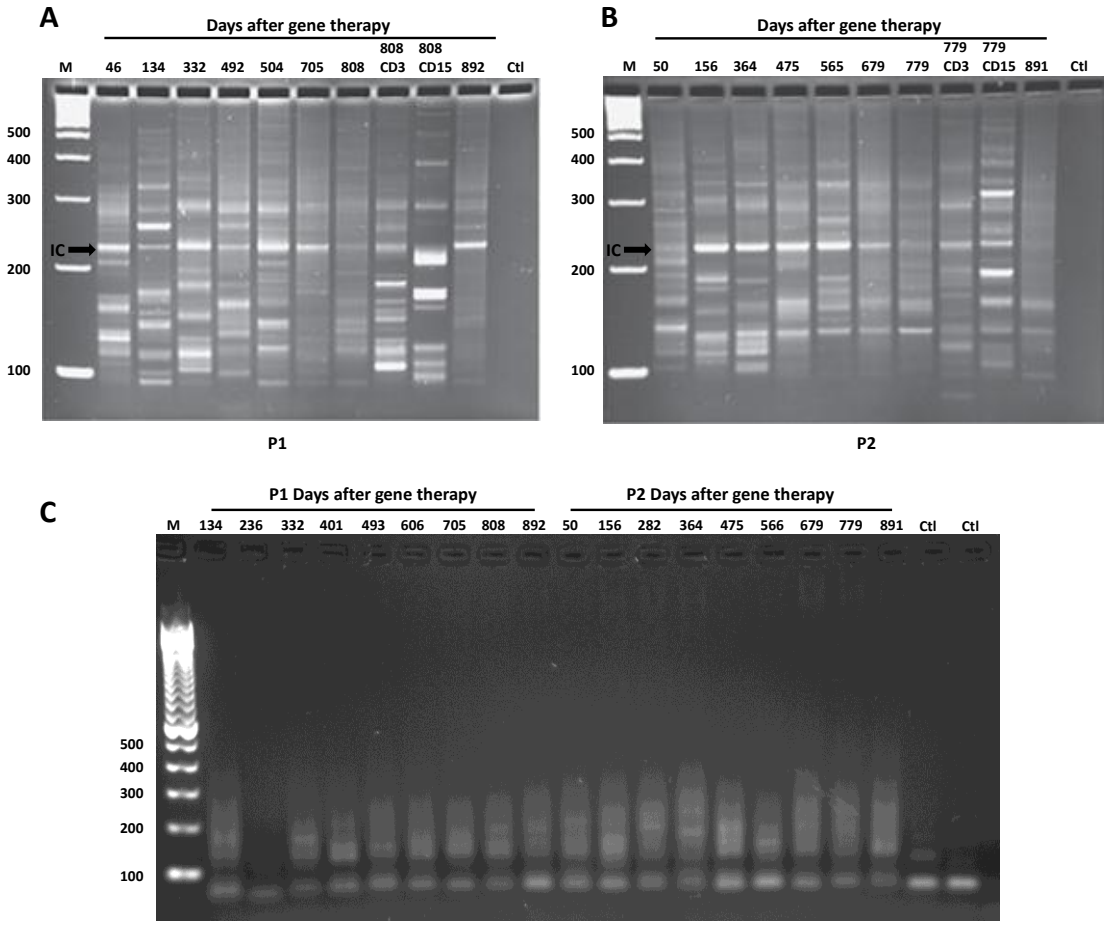

Figure 4-1(nr)LAM-PCR analysis of vector integrants. RISs were detected in tranduced hematopoietic cells at various days after gene therapy from BM cells and primary blood leukocytes obtained from Patient 1 (A) and 2 (B) using conventional LAM-PCR (Boztug et al., 2010) and non-restrictive LAM-PCR (C). CD3 denotes sorted T cells, CD15 sorted granulocytes, Ctl: healthy control subject, IC: internal vector control and M 100-bp marker.

Multiplex Barcoding of LAM-PCR samples together with downstream bioinformatical analysis of up to 1 million reads generated in one deep sequencing run enables analyzing over hundred thousand RISs in multiple samples at the same time (Paruzynski et al., 2010). This technique allows us to semi-quantitatively analyze the insertion sites, which represent unique molecular identities of particular marked clones in a certain sample, and to monitor clonal progeny and dynamics in hematopoiesis. In the WAS study, the abundance of individual gene-corrected cells and their clonal progeny in the hematopoietic system was estimated by the retrieval frequency of individual RIS sequences generated by deep sequencing. The presence of individual gene corrected cells is given as the percentage of all LAM-PCR amplicon

sequence reads identified in a particular sample. The 10 most frequently sequenced RISs were ranked from 1 to 10 according to retrieval frequency. In both patients, several clones appeared repeatedly within the 10 most abundant clones in each analyzed time point, including *PRDM16, MDS1/EVI1, CCND2,* and *LMO2* containing clones, indicating that clonal skewing had emerged among active hematopoietic gene-corrected clones (Boztug et al., 2010).

4.1.2 *Distribution of integration events in MDS1/EVI1 and PRDM16 loci in patients over time*

To analyze the long-term impact of *MDS1/EVI1* and *PRDM16* retroviral integration sites in hematopoietic stem cell lineages, *MDS1/EVI1* and *PRDM16* loci as vector genome adjunction are monitored and analyzed by LAM-PCR prospectively more than two years after gene therapy. For this, the specific *MDS1/EVI1* and *PRDM16* IS identified by LAM-PCR (Boztug et al., 2010) were further analyzed with the nrLAM-PCR data generated in this thesis. In both patients, the same multiple vector targeted genes that were previously found to expand clones in gene-corrected myelopoiesis (*MDS1/EVI1, PRDM16,* and *SETBP1*) and lymphopoiesis (*LMO2*, *CCND2*, and *BMI1*) were noted. The most active cell clone in Patient 1 showed a vector integrant upstream of *CCND2*, and in Patient 2, the site of insertion was *MDS1/EVI1***.** Although *MDS1/EVI1* and *PRDM16* loci were evaluated as common insertion sites (CIS) in this study, gene-modified clones containing insertions in *MDS1/EVI1* and *PRDM16* exhibited and remained constantly lower than 7% of the whole population (Figure 4-2, A and B)(Boztug et al., 2010). Similar results were observed in this thesis with nrLAM-PCR analyzed samples: although *MDS1/EVI1* and *PRDM16* containing clones were overrepresented in those 2 patients over time, there was no obvious clonal expansion accounted for these clones. The proportion of those *MDS1/EVI1* and *PRDM16* containing clones in whole BM and PBC samples remained constantly below than 8%. These results were different compared to the recent X-CGD study, in which patient clones showed *in vivo* clonal expansion after gene therapy. This expansion of vector marked cells was completely accounted for vector-containing clones with insertions in three gene complexes: *MDS1/EVI1,*

PRDM16 and *SETBP1*. No sign of hematopoietic clone expansion related to *MDS1/EVI1* and *PRDM16* containing clones was observed so far in the WAS study.

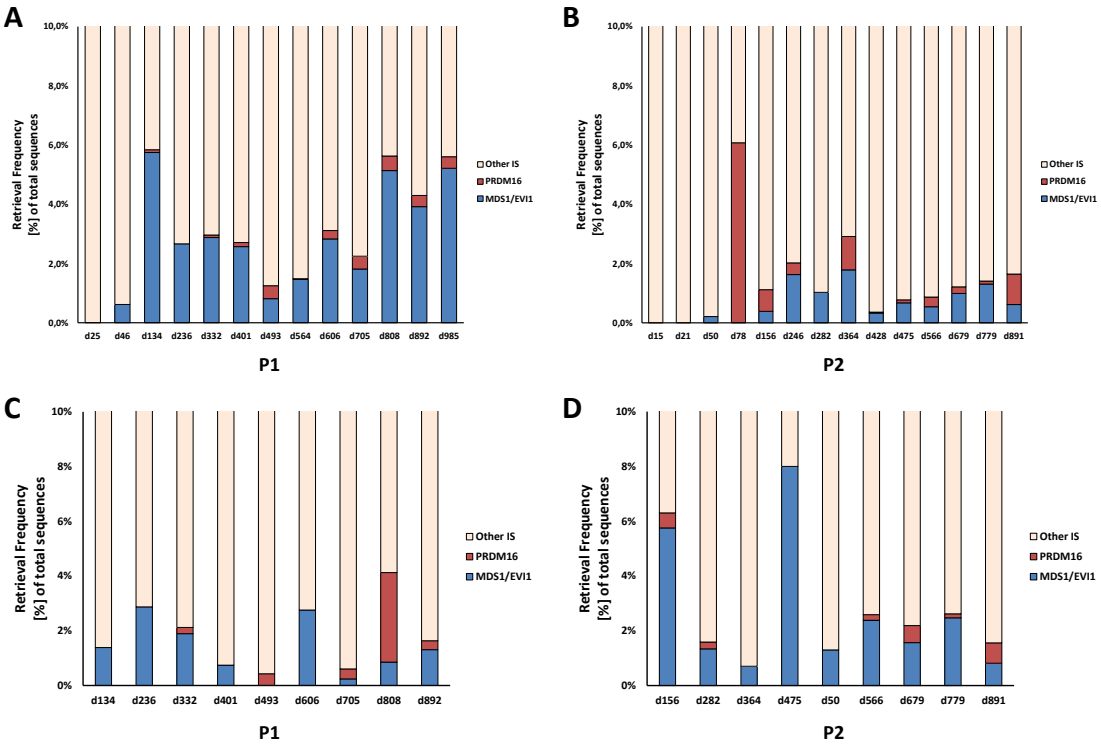

Figure 4-2 Relative clonal contribution of MDS1/EVI1 and PRDM16 containing clones in bone marrow and peripheral blood leukocytes samples of Patient 1 and 2 analyzed by LAM-PCR (A, B) (Boztug et al., 2010) and nrLAM-PCR (C, D) over time.

Further, the contribution of *MDS1/EVI1* and *PRDM16* containing clones to hematopoiesis lineage reconstitution and differentiation, sorted myeloid (CD15+) and lymphoid (CD3+) lineages were analyzed by LAM-PCR. *MDS1/EVI1* and *PRDM16* containing clones were slightly restricted in the myeloid compartment of reconstituted hematopoiesis in both patients (Boztug et al., 2010). In patient 1, *MDS1/EVI1* and *PRDM16* containing clones were exclusively appeared in CD15+ myeloid compartment; few *MDS1/EVI1* and *PRDM16* containing clones could be also detected in CD3+ lymphocytes in patient 2. One particular *MDS1/EVI1* containing clone 169071575 began to increase its frequency of whole CIS in granulocytes compartment over time. In contrast, *LMO2* and *CCND2* containing clones were almost exclusively restricted to lymphocytes in both patients (Boztug et al., 2010). Retroviral insertion sites that were identical between sorted lymphoid or myeloid cell fractions were also retrieved from both patients. This indicated the

hematopoietic stem cells or progenitor cells with multilineage capacity were initially transduced by retroviral vectors.

4.1.3 *Short summary of this section*

To analyze the long-term impact of *MDS1/EVI1* and *PRDM16* retroviral integration sites in hematopoietic stem cell lineages, *MDS1/EVI1* and *PRDM16* loci as vector genome adjunction are monitored and analyzed by (nr)LAM-PCR prospectively more than two years after gene therapy. Vector insertion sites analysis of peripheral blood and bone marrow samples demonstrated sustained highly polyclonal reconstitution over time, although *MDS1/EVI1* and *PRDM16* loci were evaluated as common insertion site (CIS) in this study. Gene-modified clones containing the insertion of *MDS1/EVI1* and *PRDM16* exhibited only very small proportion and remained constantly low level of the whole population. In sorted myeloid (CD15+) and lymphoid (CD3+) lineages, *MDS1/EVI1* and *PRDM16* containing clones were slightly restricted in the myeloid compartment of reconstituted hematopoiesis in both patients. The vector integration pattern remained highly polyclonal and no obvious clonal dominance related to *MDS1/EVI1* and *PRDM16* containing clones have been observed after two years of gene therapy. Prospective monitoring for signs of clonal dominance in gene-corrected hematopoiesis will continue in both patients.

4.2 Influences of Mds1/Evi1 and Prdm16 in long-term repopulating hematopoietic stem cells

We have introduced the concept of measuring physiological and aberrant clonal contributions of stem and progenitor cells to complex configurations of the peripheral blood by using highly sensitive and specific retrovirus integration site (RIS) analysis of gene marked cells. This approach has enabled us to decipher clonal contributions not only in mouse models, but also in large animals and clinical gene transfer studies, providing access to a new quality of clonal contribution data, and thereby progress in understanding human hematopoiesis.

To monitor clonal contribution and to study the hypothesis that biological side effects of vector insertion might confer a cells selection advantage to affected cells *in vivo*, we have performed a large-scale mapping of RIS in a clinical trial using HSCs directed gene therapy to treat X-linked chronic granulomatous disease (X-CGD, see introduction). 5 months after transplantation and recovery of marrow function, we observed that RIS distribution became non-random and that vector insertional activation of *MDS1/EVI1*, *PRDM16* and *SETBP1* triggered a 3- to 5- fold expansion of the transduced cell pool, leading to a sustained long-term transgene expression for the following 1.5 years, before turning into malignant transformation. Thus, studies on the influence of *Mds1/Evi1* and *Prdm16* gene expression on morphologically normal hematopoiesis was followed in this thesis.

4.2.1 The MDS1/EVI1 and PRDM16 genomic locus and its gene products

The *EVI1* genomic locus was initially identified as a retrovirus integration site in AKXD murine myeloid tumors (Mucenski et al., 1988). The human homolog was then localized to chromosome 3 band q26 (Morishita et al., 1990). The murine and human *EVI1* complementary DNAs are highly homologous in nucleotide and amino acid sequence. The human *EVI1* gene spans 60kb and contains 12 exons (Figure 4-3, A), capable of generating multiple alternative 5-mRNA variants and several alternatively

spliced transcripts (Wieser, 2007). The human *MDS1* gene was first identified as a component of the *AML1-MDS1-EVI1* translocation 3;21 found in some spontaneous human myeloid leukemias (Fears et al., 1996). *MDS1* spans 500 kb, contains five exons, three coding, and is located 3 kb apart to the first exon of *EVI1* (Figure 4-3, A). Transcripts are produced that contain only *EVI1* exons, only *MDS1* exons, or an alternatively spliced product containing the first two exons of *MDS1* fused to exon 2 and all remaining exons of *EVI1* (Figure 4-3, A). This fusion gene product is expressed in both normal and leukemic tissues (Fears et al., 1996). Both *MDS1/EVI1* transcripts and *EVI1* transcripts are more abundant than the *MDS1* transcripts, and in most tissues *MDS1/EVI1* and *EVI1* transcripts are coordinately regulated (Wimmer et al., 1998).

Three different proteins can be produced from the *MDS1/EVI1* locus, namely: MDS1, EVI1 or MDS1/EVI1. The major EVI1 protein derived from all the alternatively spliced variations is 1,051 amino acids in length with a predicted molecular weight of 118kd. The MDS1/EVI1 protein contains 188 additional amino acids encoded by the first two *MDS1* exons, and the entire *EVI1* exons, thereby producing a so-called "PR" domain (PRD1-BF1/BLIMPI1-RIZ homology), a 134 amino acid region with high homology to the SET domain, the structural hallmark of histone methytransferases.

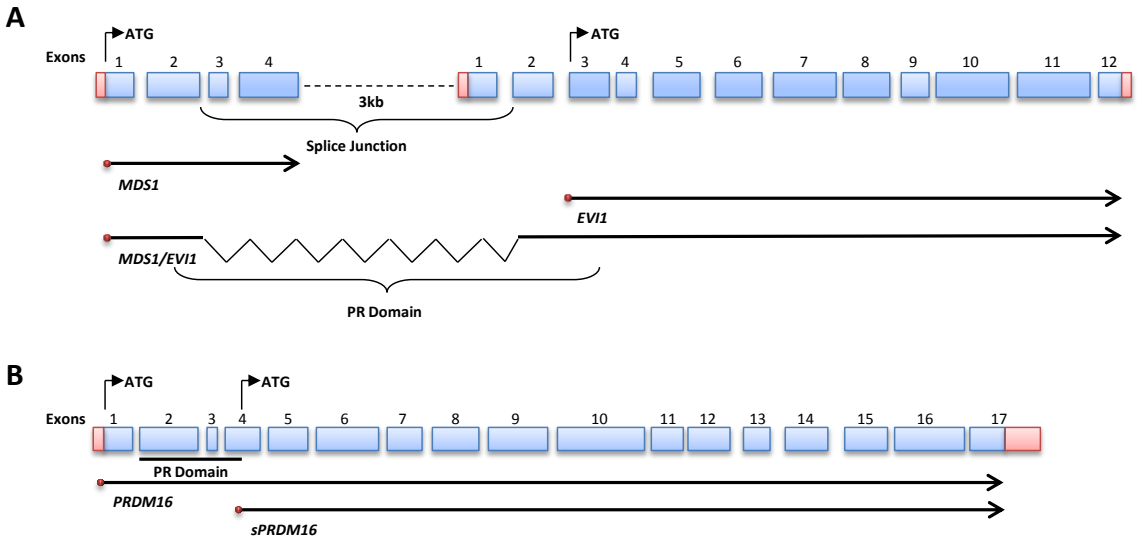

Figure 4-3 Schematic representaion of genomic locus of MDS1/EVI1, PRDM16 and their truncated forms. Exons are represented as solid boxes with light blue. Dashed lines represent splicing events occuring to produce messenger RNA. Proteins are represented as black lines.

PR domain containing 16 (*PRDM16*; also known as *MEL1*) is involved in rare AML-associated translocations t(1;3)(p36;q21) and t(1;21)(p36;q15). In both cases, expression of *PRDM16* is altered, either as consequence of *cis* juxtaposition to the enhance element of *RPN1* at 3q21 (Mochizuki et al., 2000; Shimizu et al., 2000) or to its fusion with *AML1* (*AML1/PRDM16*) at 21q15 (Sakai et al., 2005). The *PRDM16* locus encodes 2 proteins: PRDM16 and the short isoform, sPRDM16, differ at their N-terminal in the presence or absence of the PR domain (Nishikata et al., 2003) (Figure 4-3, B).

The PR-domain negative isoform of MDS1/EVI1 and PRDM16 has the potential to be oncogenic (Buonamici et al., 2004; Shing et al., 2007). However, a direct comparison of the impact of the 2 PRDM16 and MDS1/EVI1 isoforms in hematopoiesis is still missing, and whether transient co-expressing of *MDS1/EVI1* and *PRDM16* could influence HSC engraftment and proliferation remains unclear.

4.2.2 *Lentiviral expression system for Mds1/Evi1, Prdm16 and their truncated forms*

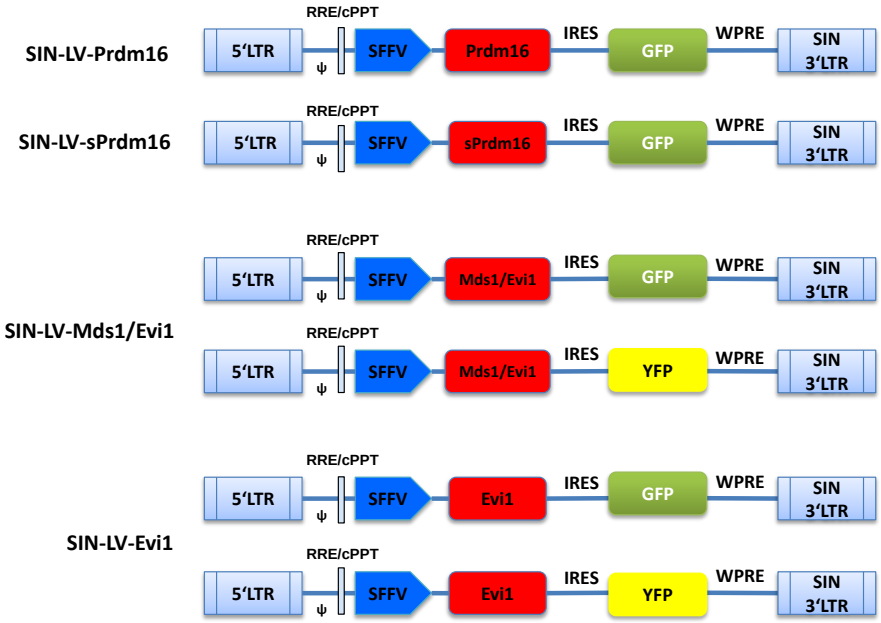

Figure 4-4 Schematic representation of SIN lentiviral vector constructs expressing murine Prdm16, sPrdm16, Mds1/Evi1 and Evi1, respectively. Murine *Prdm16* and *sPrdm16* coding sequences (CDS), as well as *Mds1/Evi1* and *Evi1* CDS were cloned into SFFV promoter (spleen focus-forming virus) containing transfer lentiviral vector backbone, expressing Green Fluorescence Protein (GFP) or Yellow Fluorescence Protein (YFP) simultaneously,

respectively. LTR: long terminal repeat; IRES: internal ribosomal entry site; WPRE: woodchuck hepatitis virus post-transcriptional regulatory element; The 3' ΔU3 comprises the SIN deletion, ΔU3. The R region contains the vector polyadenylation signal.

To construct the lentiviral vectors, endogenous protein coding sequences of murine *Mds1/Evi1*, *Prdm16* and their PR domain lacking variants *Evi1* and *sPrdm16* were directionally cloned into multi-cloning sites of third generation pCCL pptSFFV-GFP-PRE SIN lentiviral backbone, which contains either an IRES-eGFP or an IRES-eYFP cassette as tracing marker (Figure 4-4). The transgene coding sequences were verified by Sanger sequencing. To produce the functional lentiviral vectors, transfer vector constructs together with envelope and packaging plasmids were transient co-transfected into 293T cells mediated by PEI or calcium phosphate. 48 hours after transfection, virus particles containing medium were harvested and concentrated by ultracentrifugation, resulting into ~100-fold enrichment of functional virus particles (Figure 4-5).

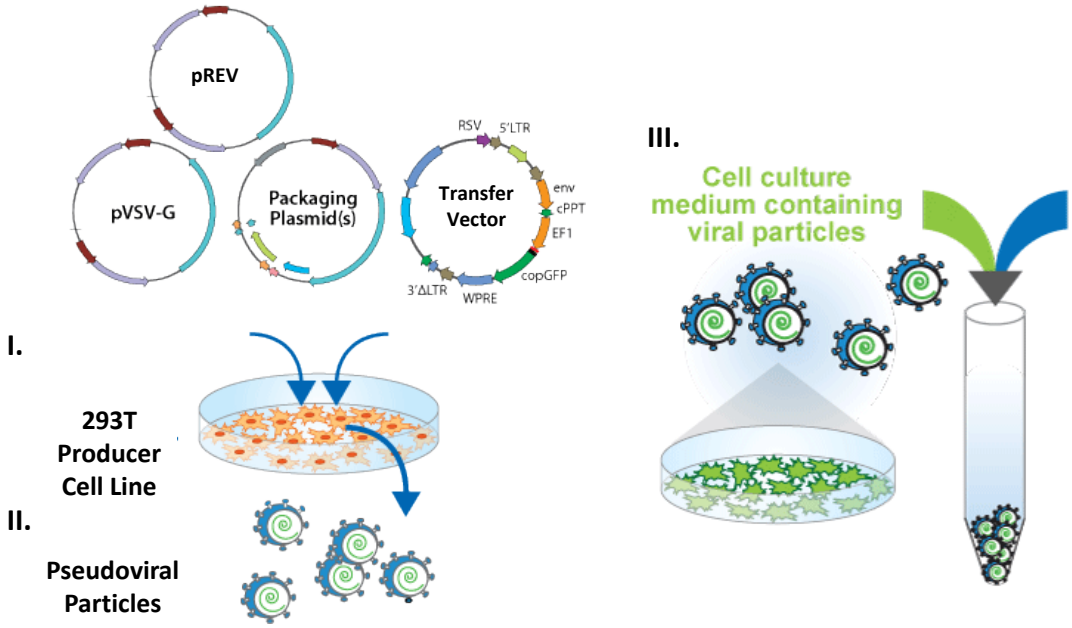

Figure 4-5 Schematic representation of the process of lentiviral vector production (modified from www.invitrogen.com). I. Transient transfection in 293T cells. II. Harvest pseudoviral particles from supernatant. III. Concentration of viral partilcles through ultracentrifugation.

Functional titers of LVs were determined by serial transduction in HeLa cells. In all LV constructs either eGFP or eYFP were present, and the percentage of transduced cells could be easily determined by FACS based on fluorescence gene expression. To estimate the transduction ability, whole LV particles produced after

ultracentrifugation, including partially packaged, empty or destroyed particles were also titrated using a HIV-1 specific p24 ELISA kit. Reliable titers in the range from 10^7 to 10^8 TU/ml were achieved for *Mds1/Evi1, Prdm16* as well as *Evi1* and *sPrdm16* constructs.

4.2.3 *Lentiviral mediated transgene expression in murine cell lines*

To verify the transgene expression mediated by LVs, murine SC-1 cells were transduced by the produced LVs in presence of 8µg/ml Polybrene. 48 hours after transduction, cells were harvested and mRNA and Protein were isolated as well. To verify the transgene expression at the mRNA level, RT-PCR was performed using specific primer pairs complementary to the C-terminus of the transgene (full length), N-terminus (PR domain specific), eGFP as well as house keeper GAPDH region, respectively. All negative controls, non-template controls, as well as non-RT controls showed no PCR products after gel electrophoresis, indicated that all the PCR products in RT positive samples were derived from transgene expressing LVs. PR domain specific PCR demonstrated the PR domain presence or absence in the respective expressed isoforms (Figure 4-6, A).

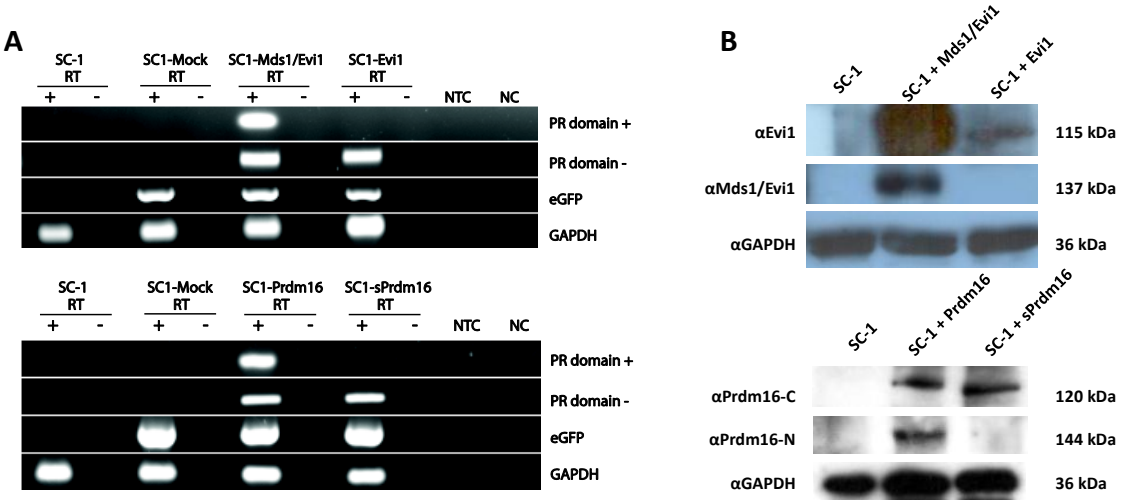

Figure 4-6 Overexpression of *Mds1/Evi1* and *Prdm16* in lentiviral vector transduced SC-1 cells. (A) RT-PCR analysis of transgene expression at mRNA level. cDNAs were synthesized with (RT+) or without reverse transcriptase (RT-). GAPDH primer was used as positive control, eGFP primer was used as functional provirus control, C-terminus and N-terminus primers were used to detect PR domain present or absent isoforms, respectively. **(B)** Western blot analysis of transgene expression at protein level. GAPDH antibody was used as loading control, C-terminal and N-terminal antibodies were used to detect PR domain present or absent isoforms.

To confirm the LV mediated transgene expression also at the protein level, 10µg whole lysate of transduced cells were loaded into Tris-Glycin SDS PAGE gel. After electrophoresis, SDS page gels were blotted on a PVDF membrane and probed with N-terminal (PR domain specific) or C-terminal (full length) antibodies against Mds1/Evi1 or Prdm16, respectively. Specific expression of PR domain containing or lacking isoforms driven by the respective LVs was shown (Figure 4-6, B). The expression of Mds1/Evi1 and Prdm16 driven by the produced LVs was also verified in the murine C1498 cell line (data not shown).

4.2.4 *Effects of Prdm16 and sPrdm16 on HSCs and progenitors*

To study the direct impact of *Prdm16* and *sPrdm16* on hematopoietic differentiation and proliferation, *Prdm16* and *sPrdm16* were ectopically expressed after lentiviral gene transfer in a murine cell population enriched in HSCs and progenitors (lineage-depleted cells "lin-"). 4 days after transduction, FACS analysis for eGFP confirmed transgene expression in all transduced lin- cells (Figure 4-7). Surprisingly, the initial transduction efficiencies between control vector (LV-109) and transgene expressing vectors (LV-Prdm16 and LV-sPrdm16) varied from 64.7% ± 2.6% to only 2.2% ± 0.8% and 2.9 ± 1.1%, respectively. At day 7 post transduction, the proportion of eGFP positive cells in control vector transduced lin- cells was increased to 75% ± 1.4%, but in Prdm16 and sPrdm16 transduced cells, the already low initial percentage further decreased to 0.8% ± 0.5% and 1.5% ± 0.7%, respectively. At day 14 post transduction, the percentages of eGFP positive cells in control, Prdm16 and sPrdm16 transduced samples were dropped to 46.3% ± 0.5%, 0.1% ± 0.1% and 0.1% ± 0.1%, respectively. This finding suggests that transgene expression was probably limited by transgene size, methylation of vector promoter or toxicity of ectopic transgene expression. Notably, the latter has also been noticed in other groups (personal communication Prof. C. Baum, MHH, Hannover)

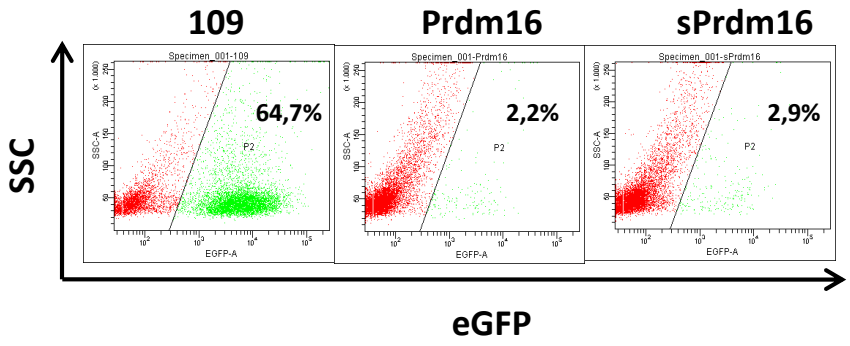

Figure 4-7 Gene transfer efficiency of Prdm16 and sPrdm16 LVs in lin- cells. Lineage depleted bone marrow cells were transduced by Prdm16 and sPrdm16 expressing lentiviral vectors with MOI=50, which were titrated in HeLa cells. The transduced cells were cultured *in vitro* for 72 hours in cytokine containing medium. Transduction efficiency was measured by FACS based on eGFP expression.

The effects of *Prdm16* and *sPrdm16* expression on progenitor life span were assessed by serial replating of lin- cells in methylcellulose based cytokines containing medium. In the first round of plating, although the frequency of clonogenic progenitors detected transduced lin- cells was essentially identical for control and Prdm16/sPrdm16 transduced cells, Prdm16/sPrdm16 transduced progenitors formed much larger granulocyte, erythroid, macrophage, and megakaryocyte colonies (GEMM) at the very beginning time point (Day 1) (39/Prdm16, 59/sPrdm16, 15/untransduced and 22/mock transduced) (Figure 4-8). When directly comparing the cells expressing *Prdm16* and *sPrdm16*, *sPrdm16* expressing cells formed larger portion of GEMM colonies then those expressing *Prdm16* in Day1 and Day 7. Notably, overexpressing of *Prdm16* and *sPrdm16* increased both the number and the size of colonies, suggesting that their expression affects both clonogenicity and proliferation of progenitors. After the first observation period, no colonies were detected in all samples, therefore no serial replating assays were conducted. Together with the observation of low gene transfer efficiency, these findings suggest that *Prdm16* and *sPrdm16* expression might block differentiation and extends survival of hematopoietic progenitor cells, but further optimization of experimental procedure and improvement of transduction efficiency was needed.

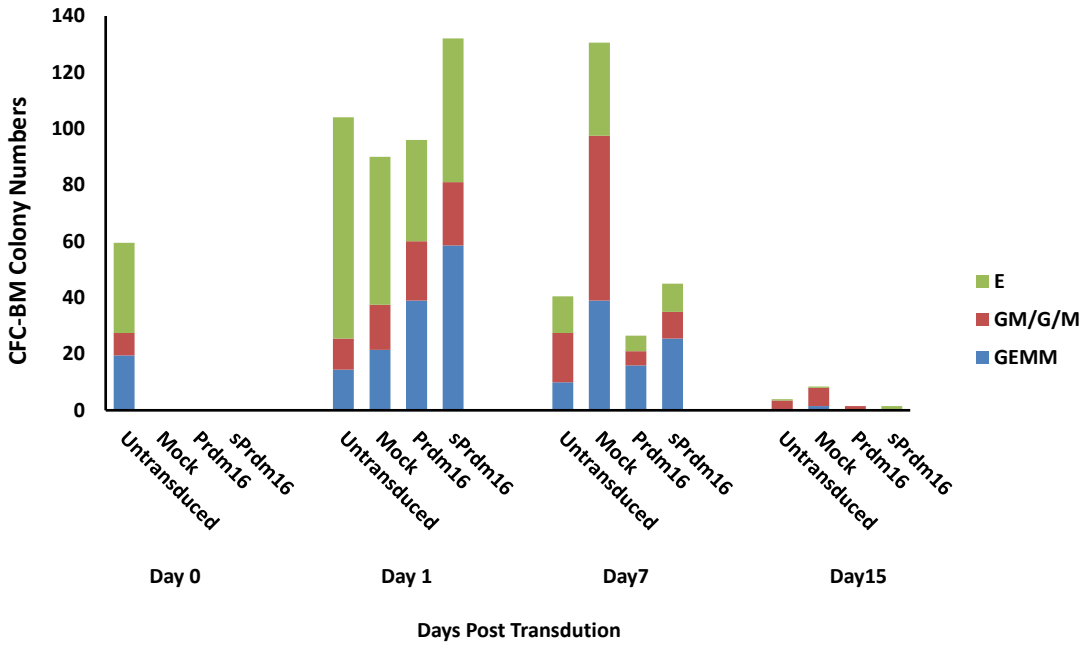

Figure 4-8 Increased hematopoietic stem/progenitor cells in Prdm16 and sPrdm16 over expressed Lineage negative bone marrow (lin- BM) cells. Total number of CFU-Bone marrow derived from 800 FACS-sorted unstranduced, mock, Prdm16 or sPrdm16 transduced Lin- BM cells post transduction. E: Colony forming units-erythroid; GM: Colony forming units-granulocyte and megakaryocyte; G: Colony forming units-granulocyte; M: Colony forming units-megakaryocyte; GEMM: Colony forming units- granulocyte, erythrocyte, monocyte and megakaryocyte.

4.2.5 *Long term persistence of LV transduced clones in hematopoiesis in vitro*

To investigate the effect of *Prdm16* and *sPrdm16* expression on HSCs, we used the long-term culture assay that assesses the expansion of primitive HSCs *in vitro*. Transduced lin- cells were cultured in medium containing IL-3, IL-6 and SCF for 4 weeks. In this period, a portion of cells were monitored for eGFP expression by FACS. After counting the cell numbers of each transduced cell population at each observation time point, only one third of cells were further cultured under the same condition to monitor the clonal expansion of the transduced cells. In the first experiment, the proportion of untransduced and mock transduced remained constant at input level over 2 weeks, while *Prdm16* and *sPrdm16* transduced cells began to expand. 2 weeks later, in contrast to untransduced and mock transduced cells, *Prdm16* and *sPrdm16* transduced cells expanded from input level by nearly 1100- to 1300- folds. After long-term culture, cell morphology showed a higher proportion of immature myeloid cells upon expression of *Prdm16* and *sPrdm16*. In

the second experiment, no significant differences in the proliferation or differentiation were observed between *Prdm16*, *sPrdm16* and eGFP transduced cells.

In both long-term culture experiments, the eGFP expression level in transduced cells was decreasing overtime under detection limit. This finding suggests that initially transduced committed progenitors terminally differentiate, while low frequent HSCs self-renew. To verify the presence of transduced cells in long-term cultured samples, LTR test PCR was performed on genomic DNA of transduced cells 1 and 3 weeks after transduction (Figure 4-9). LTR specific PCR products appeared in all three analyzed samples of both time points, indicating the long-term persistence of LVs in transduced primary samples.

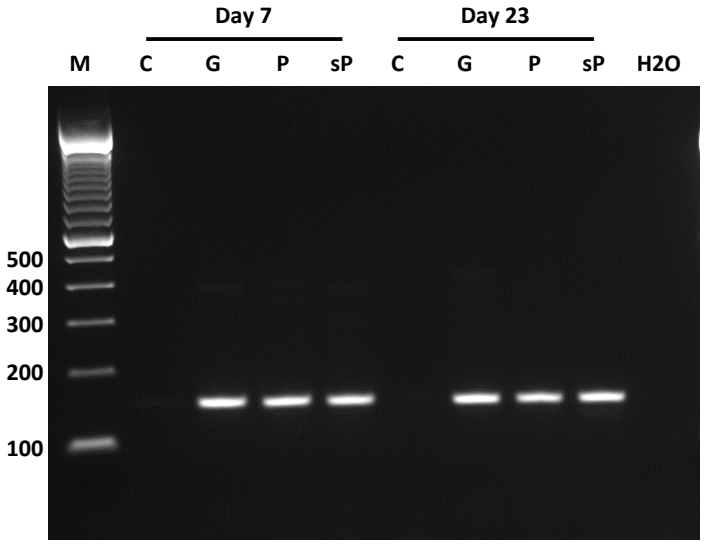

Figure 4-9 LTR test for long-term cultured Prdm16 and sPrdm16 transduced lin- cells. 7 days and 23 days after transduction, genomic DNA of samples were isolated and PCRs were performed using LTR specific PCR. Nested PCRs were performed to increase the signal strength. C: control, not transduced lin-cells; G: eGFP expressing control LV transduced samples; P: Prdm16 expressing LV transduced samples; sP: sPrdm16 expressing LV transduced samples.

4.2.6 *Optimization of lentiviral vector production*

In the context of LV transduction of primary hematopoietic cells, the general gene transfer efficiency was contrary to our expectation. This could be explained by transgene toxicity, promoter methylation, inadequate LV packaging as well as inaccurate LV titration, etc. To improve the titer of LVs, several optimization steps were performed.

First, the quality of FBS used in LV production as well as downstream HSCs/progenitor cell based assays were critically analyzed. 11 FBS batches from different vendors were used for LV production, using the control construct and compared the respective titers either on p24 or functional GFP assays. As shown in Figure 4-10 A, p24 titers demonstrated analogous results to functional GFP titers, suggesting the procedure of LV production was reliable and reproducible; both titration methods could be further used to estimate the qualities and quantities of produced LVs. Sample 8 showed the best p24 titer as well as GFP titer (Figure 4-10, A), including minimal effects on HSCs differentiation and proliferation (data not shown). Although the infectivity (GFP titer/p24 titer) was not superior, in combination with the data of HSCs assays, FBS 8 was selected for further experiments.

To stabilize the virus mRNA transcripts upon LV entry, as well as to improve the transgene expression level, Kozak consensus sequence “gccacc” was cloned in front of the start codon of each transgene construct. Further optimization steps including 293T cell numbers, ratio and amount of plasmids used in LV production, condition for ultracentrifugation were performed. After extensive optimization works, the general functional GFP titers on HeLa cells were successfully increased with 20- to 100- fold (Figure 4-10, B). Thus, the usage of higher MOIs in primary cell transduction experiments became feasible.

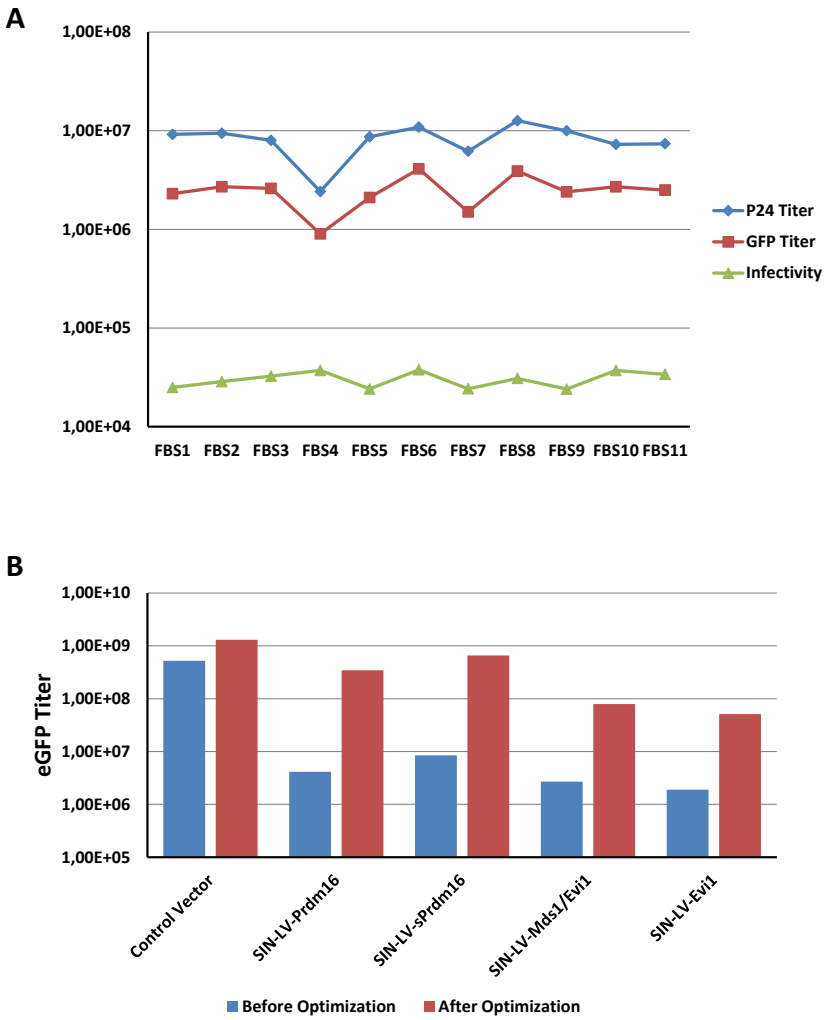

Figure 4-10 Significant improvement of lentiviral vector titers after optimization. (A) 11 batches FBS from different vendors were tested for LV production. Functional LV titers were measured by FACS based on eGFP expression, total LV titers were measured by p24 ELISA assay. LV infectivity of each production was calculated by dividing GFP titer with p24 titer. **(B)** Summary of titer improvement of transgene expressing LVs before optimization (blue bar) and after optimization (red bar).

4.2.7 *Fractionation of the LSK HSC pool into distinct LT-, ST- HSC and MPPs*

To further improve the LV transduction efficiency in the context of primary cell gene transfer, more defined LT-, ST- HSCs and MPPs separated by FACS using cell surface marker Sca1, cKit, Flt3 and CD34 were evaluated in this thesis (Figure 4-11).

Although representing only 0.05% to 0.1% of total adult BM cells, the LSK compartment contains all LT-HSCs. Whereas the LSK LT-HSCs population lacks CD34 and flt3 expression, more than 90% of LSK cells are CD34+ with short-term

repopulating stem and progenitor cell activities. However, the potential heterogeneity of the LSKCD34+ ST-HSC pool can be further divided into LSKCD34+flt3- and LSKCD34+flt3+ after co-staining with anti-CD34 and anti-flt3 antibodies (Yang et al., 2005). These three distinct LSK populations, LSKCD34-flt3-, LSKCD34+flt3- and LSKCD34+flt3+, representing 0.007%, 0.04% and 0.05% of total BM cells, exhibit also distinct HSCs properties, reflecting LT-, ST- and MPPs, respectively.

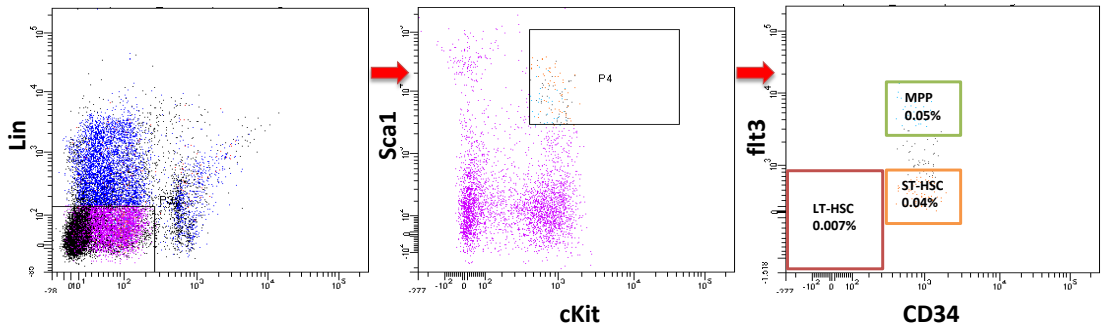

Figure 4-11 LT-, ST- HSC and MPPs sub-fractionation of LSK cells based on expression of CD34 and Flt-3. Lin- BM cells (left panel) expressing high levels of c-kit and Sca-1 (middle panel) were investigated for expression of CD34 and flt3 (right panel). Percentages indicate frequencies (mean of 3 independent experiments) within total BM of LSKCD34-flt3- (lower left quadrant), LSKCD34+flt3- (lower right quadrant) and LSKCD34+flt3+ (upper right quadrant) cells. Boxes denote the sorting strategy used for each of the populations. Cells used in these studies were double sorted, always resulting in 98% or higher purity (data not shown).

4.2.8 *Short summary of this section*

Protein coding sequences of murine *Mds1/Evi1*, *Prdm16* and their PR domain lacking variants *Evi1* and *sPrdm16* were cloned and verified by sequencing. Lentiviral vectors (LVs) were produced and reliable titers in the range from 10^7 to 10^9 TU/ml were achieved after optimization for all LVs. Sufficient transgene expression was detected for all constructs by Western-Blot and RT-PCR after transduction of murine SC-1 cells. To study the direct impact of the *Prdm16* and *sPrdm16* expression on hematopoietic differentiation and proliferation, lineage depleted murine primary bone marrow cells were transduced with Prdm16 or sPrdm16 LVs (MOI 50) and monitored for 30 days *in vitro*. Long-term eGFP expression was observed in all three transduced cells populations for up to 30 days. No significant differences in the proliferation potential or differentiation capacity were observed between Prdm16, sPrdm16 and eGFP transduced cells. To improve the transduction efficiency, more

defined LT- and ST- HSCs separated by FACS using cell surface marker Sca1, cKit, Flt3 and CD34 were tested for further studies.

4.3 Comprehensive characterization of integration deficient lentiviral vector (IDLV) integration profiles in vitro

The application of IDLV which can persist in the nucleus with surprising robustness might address growing concerns about vector and gene therapy biosafety issues. Murine leukemia virus based integrase mutant class I deficient vectors have shown to persist as circular 1-LTR, 2-LTR and oligomeric episomal vector forms (Hagino-Yamagishi et al., 1987). Residual integration events have been attributed to non-homologous recombination, often coupled with the loss of nucleotides at the site of integration (Gaur and Leavitt, 1998). One pre-clinical study on the successful long-term functional correction of retinal degeneration in two rodent disease models have proven the therapeutic potential of these vectors, at least in post-mitotic or otherwise quiescent cells (Yanez-Munoz et al., 2006).

Episomal forms of IDLVs are not able to replicate and are diluted out in rapid dividing cells. The resulting transient expression of specific transgenes can represent an important biosafety feature. A transient transgene expression should however be sufficient to induce critical changes in the cell fate of transduced cells during their first few replication cycles, e.g. when trying to modify proliferation and differentiation capacities of repopulating HSCs, or iPS cells. The transient presence of these nuclear forms further allows the use of IDLVs as a hybrid gene ferry thus to combine the non-toxic nucleic acid transfer properties of LV particles with the reliable and stable integration of animal transposases (Ivics et al., 1997), I-*SceI* nuclease (Cornu and Cathomen, 2007) and Zin finger nuclease for targeted genome modifications (Lombardo et al., 2007).

The main objectives of this study were I) to study the base line functional properties of IDLVs, including residual integration frequency and distribution analysis, II) to test whether IDLV derived transient *MDS1/EVI1* and *PRDM16* expression is relevant to influence the fate of long-term hematopoietic repopulating cells.

4.3.1 *Construction of integration deficient lentiviral vector*

Transgene expression from IDLVs probably depends on both the number and the structure of extra-chromosomal viral DNA (Philpott and Thrasher, 2007). Based on the previous study, we chose to introduce class I mutation into IN at positions D64 (Figure 4-12, B), which is part of the DDE catalytic triad and is absolutely essential for integration (Leavitt et al., 1996), resulting a mutant D64V gag/pol packaging plasmid. The expression of Mds1/Evi1, Prdm16 and their PR domain absence forms are driven by internal SFFV promoter, eGFP is co-expressed as a surrogate marker for the intracellular presence of vector genomes (Figure 4-12, A).

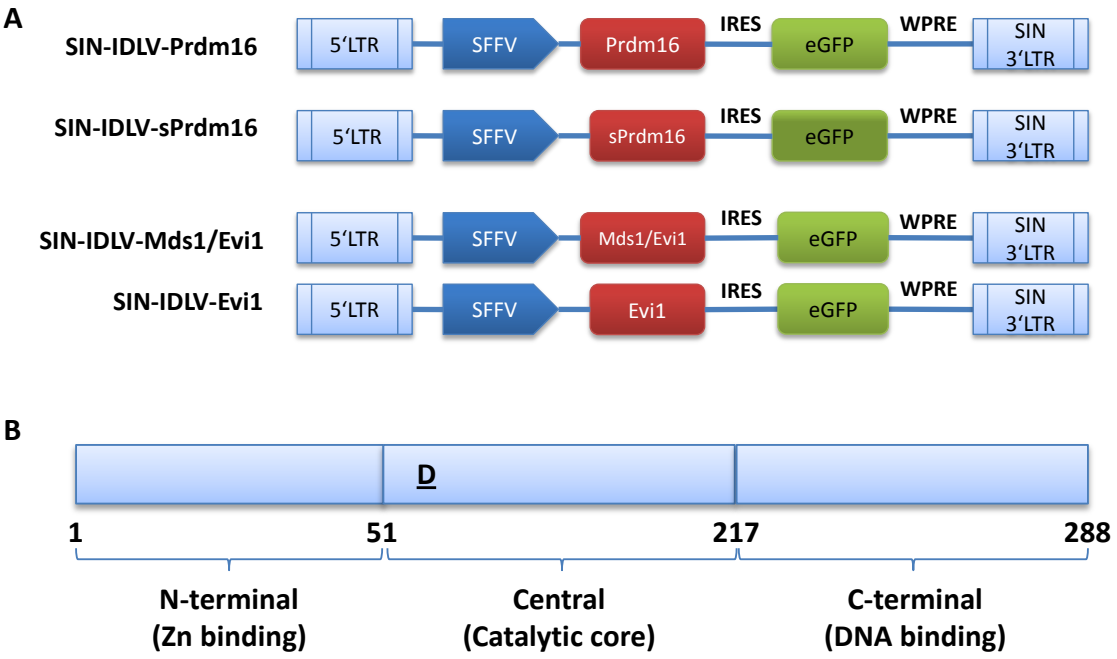

Figure 4-12 Schematic representation of IDLV constructs expressing murine Prdm16, sPrdm16, Mds1/Evi1 and Evi1, respectively. **(A)** Murine Prdm16 and sPrdm16 Coding sequences (CDS), as well as Mds1/Evi1 and Evi1 CDS were cloned into SFFV promoter (spleen focus-forming virus) drived transfer vector backbone expressing Green Fluorescence Protein (GFP) simultaneously. LTR: long terminal repeating element; IRES: internal ribosome entry site; WPRE: woodchuck hepatitis virus post-transcriptional regulatory element; the 3' ΔU3 contains the SIN deletion ΔU3; the R region contains the vector polyadenylation signal. **(B)** HIV-1 integrase consists of three domains: the N-terminal domain, the C-terminal domain, and a catalytic core domain. Mutation was introduced in D64 residue within the catalytic core.

To confirm that introduction of D64V mutation in IN did not interfere with virus production, the concentration of p24 was measured. The p24 enzyme-linked immunosorbent assay (*ELISA*) results in Table 4-1 showed minor variation between the WT and D64V vector stocks, indicating that introduction of the class I IN mutation did not interfere with virus budding and production. Viral titers were

quantified by measuring transducing units in HeLa cells. Albeit IDLVs achieved lower titers in comparison to wild type LVs, which might be explained by dilution of transduced events during culturing, all stocks showed relative high titers in a range between 10^7 to 10^9 TU/ml. therefore, the D64V IN mutation allow efficient viral genome encapsidation, reverse transcription of RNA genome, nuclear entry of the viral complementary DNA, and efficient transcription and translation of the transgene.

Virus	Transgene	Viral particles (pg p24/ml)	eGFP Titer (TU/ml)
LV	109	4,81E+09	1,10E+09
IDLV	109	4,03E+09	1,30E+08
LV	Prdm16	1,71E+09	3,30E+08
IDLV	Prdm16	5,59E+09	1,39E+07
LV	sPrdm16	4,47E+09	6,30E+08
IDLV	sPrdm16	2,24E+10	1,10E+07
LV	Mds1/Evi1	1,76E+09	1.70E+07
IDLV	Mds1/Evi1	6,84E+08	1,00E+07
LV	Evi1	8,08E+09	1,55E+08
IDLV	Evi1	2,23E+09	5,00E+07

Table 4-1 Quantification of p24 and eGFP expression levels in vitro. Titer measurements were obtained from serial dilutions of at least three separate batches of each vector. Viral stock p24 concentrations were measured by performing a p24 ELISA assay using serial dilutions of the vectors. eGFP titer were measured by performing transduction of HeLa cells using the same serial dilution of vector stocks used in p24 ELISA assay. FACS was performed on the third day after transduction and titers were calculated in transducing units/ml (TU).

4.3.2 *Transient transgene expression driven by IDLVs*

In order to evaluate transgene expression overtime, SC-1 and C1498 cells were transduced by LVs and IDLVs expressing eGFP at the MOI of 25, 50 and 100, respectively, and eGFP expression was measured by flow cytometry. As shown in Figure 4-13, all of the vectors showed a maximal eGFP expression level at the day 2 after transduction, demonstrated a typical LV-mediated transgene expression kinetic described before (Brussel and Sonigo, 2003, 2004; Butler et al., 2001). After transduction of both SC-1 and C1498 cell lines with wild type LVs, eGFP expression was stable and sustained over time, indicating that stable viral integration into host genome and sustained transgene expression. Of note, expression of eGFP from

transduced C1498 cells was lower as compared with SC-1 cells. This was probably attributable to the variation of transduction efficiency in suspension cell line and adherent cell line. In contrast, by 14 days after transduction, background levels of expression was seen in samples transduced with IDLVs, indicating the episomal forms were diluted in the dividing cell population over time, but residual integration still occurred. Interestingly, cells transduced by IDLVs with all three MOIs showed similar kinetics of transgene expression diminishment, suggested that IDLV background integration was not associated with initiate load of vectors.

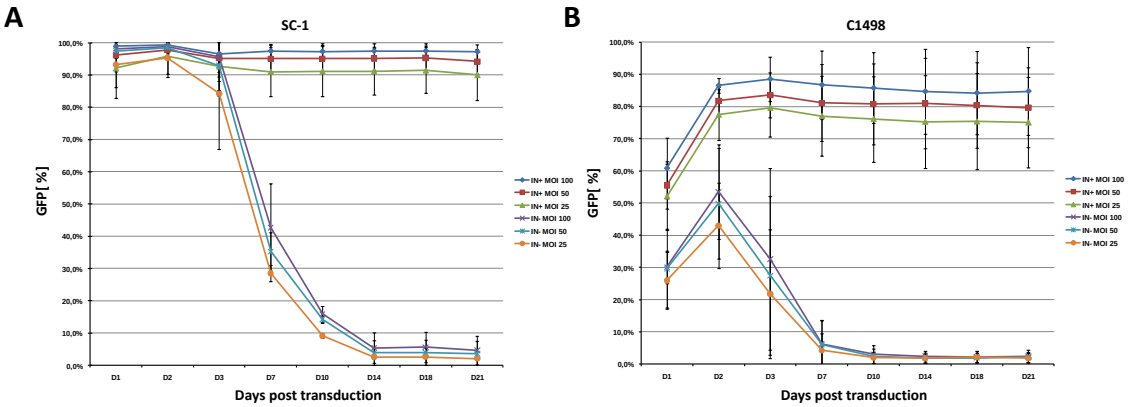

Figure 4-13 Transgene expression driven by IDLV was transient in dividing cells. (A) SC-1 and **(B)** C1498 cells were transduced with eGFP expressing IDLV and LV vectors with MOIs of 25, 50 and 100, respectively. Data represent means ± standard deviation (SD); n=3.

4.3.3 *Quantification of IDLV background integration*

Data obtained from measuring eGFP expression in SC-1 and C1498 cells suggested that all IDLVs were integration deficient, the eGFP expression levels were diminishing over time. Next, we wanted to quantify more precisely the efficiency of residual integration. IDLV transduced cells were cultured over a period of 4 weeks to dilute out the un-integrated episomal forms, e.g. 1-LTR circle, 2-LTR circle and linear forms.

The absolute integration events of IDLV in SC-1 and C1498 cells were shown in Figure 4-13, the frequencies of stable eGFP expression ranged between 1.8% and 2.4% in C1498 cells and between 2.0% and 4.6% in SC-1 cells, in a dose dependent manner. For the relative integration frequency between IDLV and LV, we isolated genomic DNA from IDLV transduced and LV transduced samples, performed qPCR using WPRE

primer sets, which is specific present in provirus sequences. The level of background integration of the IDLVs was calculated by comparison the absolute virus copy numbers in the same amount (5ng) of genomic DNA from IDLV and LV transduced samples. As shown in Figure 4-14, the level of background integration of the D64V mutant was in general 10^{-2} to 10^{-3} fold lower than integration of the wild type vector. Notably, the residual integration rate of IDLVs in C1498 cells was higher than in SC-1 cells. It could be explained by tumor genetic background of C1498 cells, which have higher genomic arrangement and instability, more IDLVs could be captured by those cells through non-homologous end joining mechanism upon vector entry.

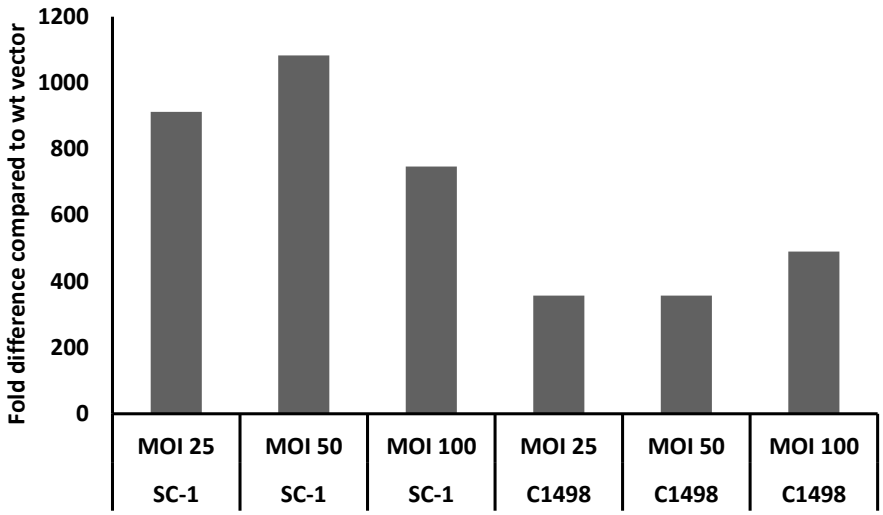

Figure 4-14 Quantification of background integration of IDLVs. Difference of integration efficiency was calculated by comparing the copy number of provirus in 500ng of genomic DNA of IDLV and LV transduced SC-1 and C1498 cells with MOI 25, 50 and 100, respectively.

4.3.4 *Enrichment of residual integrated events*

In order to precisely analyze IDLV integration events genome-widely, enhance the sensitivity of LAM-PCR, we isolated eGFP positive cells, referred to IDLV transduced cells, based on eGFP expression by FACS sorting. As mentioned above, the bulk IDLV transduced cells were cultured in vitro for 4 weeks after transduction to dilute out the un-integrated forms. 20,000 monoclonal eGFP positive cells were sorted by FACSAria in DKFZ flow cytometry core facility. Sorted cells were cultured in vitro for additional 2 week for cell expansion, a portion of cells from each samples were analyzed by FACS for sorting purity. As shown in Figure 4-15, all the samples

showed relative high eGFP positive cells proportion (≥ 85%). In comparison to wild type LV transduced samples, which eGFP positive cells were also over 85% percent in whole cell population, direct comparison between sorted IDLV transduced cells and LV transduced cells became feasible.

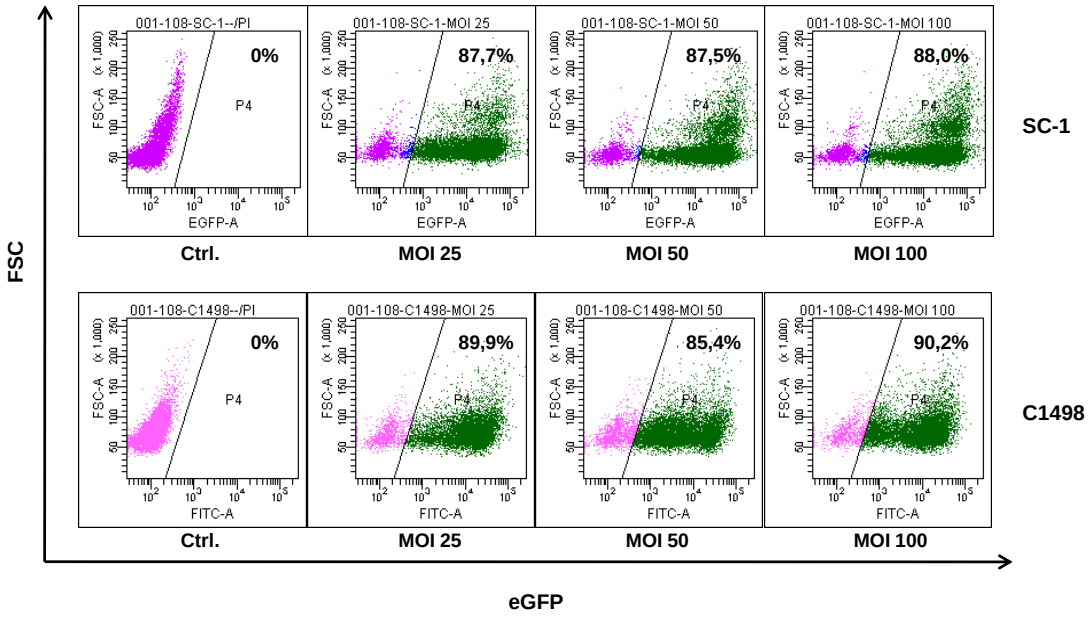

Figure 4-15 Purity of FACS sorted IDLV transduced SC-1 and C1498 cells. 20,000 eGFP positive cells were sorted by FACSAria based on eGFP expression; sorted cells were cultured for additional 2 weeks to get enough cells for further experiments.

4.3.5 *Clonality of IDLV transduced samples*

To uncover specific features of the IDLV IS distribution *in vitro*, and gain a comprehensive overview of IDLV integration preference and distribution, we performed comparative large scale IS analysis on the IDLV and LV transduced SC-1 and C1498 samples using (nr)LAM-PCR combined with high-throughput pyrosequencing (454/Roche).

The clonality of IDLV or LV transduced samples was determined by 5' LAM-PCR using NlaIII and HpyCH4IV first. 100ng genomic DNA was used as starting material. The results from both transduced cell lines showed a polyclonal (Figure 4-16, A, HpyCH4IV only, Tsp509I not shown) virus integration pattern. A distinct vector related fragment was present - the band for internal control fragment (300bp). The

3'LAM-PCR approach using Tsp509I and MseI for the both transduced cell lines resembled the result same to 5'LAM-PCR: a polyclonal pattern of IDLV IS (Figure 4-16, B, NlaIII only, MseI not shown). A 400bp internal control band could be also detected in this procedure. To overcome the obstacle of using restrictive enzymes and obtain many IS as possible, nrLAM-PCR for both 5' and 3' directions were also performed using 500ng genomic DNA as starting material, as mentioned in 4.1.1, nrLAM-PCR resulted in a variety of amplicon lengths for each insertion, appearing as a 'smear' on electrophoresis (data not shown).

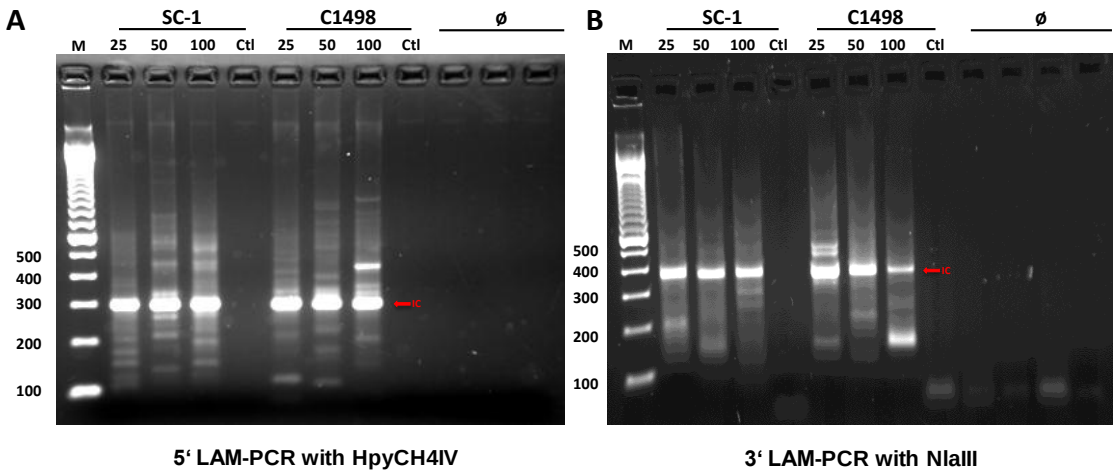

Figure 4-16 LAM-PCR analyses of IDLV integrants in the genome of transduced cells. (A) 5' strand and **(B)** 3' strand LAM-PCR analysis of FACS-sorted eGFP positive cells 6 weeks post transduction. IC, internal control derived from the vector sequences.

4.3.6 *Distribution analysis of IDLV IS in vitro*

LAM-PCR and high-throughput sequencing of integration sites was performed on genomic DNA isolated from LV and sorted IDLV transduced SC-1 and C1498 cells with MOI of 25, 50 as well as 100, respectively. In total, 905 unique insertion sites were retrieved from combined IDLV transduced SC-1 cells with all three virus loads, of which 533 IS could be mapped exactly to the mouse genome. Similarly, 437 unique insertion sites were retrieved from IDLV transduced C1498 cells, among them 271 were unique mappable. For LV thansduced cells, much more IS were isolated using the same experiment conditions: 2419 of 2788 and 898 of 1062 unique IS could be mapped exactly into mouse reference genome for SC-1 and C1498 cells, respectively

(Table 4-2). This reflecting the conclusion that IDLV integrate less frequently into host genome. The relative high number of retrieved IS of IDLV was resulted by enrichment of positive transduced cells using FACS. Of note, this was first time to retrieve such many IS of IDLV, allowing us to dissect the preference and distribution of IDLV integration genome widely, as well as the safety aspect for IDLVs.

	Total IS	Unique mappable IS	In gene	In gene +/-10kb
SC1 (IDLV)	905 (100)	533 (59)	229 (43)	299 (56)
C1498 (IDLV)	437 (100)	271 (62)	110 (41)	134 (49)
SC1 (LV)	2788 (100)	2419 (87)	1802 (74)	1935 (80)
C1498 (LV)	1062 (100)	898 (85)	557 (62)	613 (68)

Table 4-2 Large scale integration site analysis of IDLV by LAM-PCR and nrLAM-PCR. Cumulative genomic distribution of IS detected in IDLV and LV transduced SC-1 and C1498 cells. LAM-PCR amplicons were sequenced by next generation sequencing (Roche 454)**.** IS are shown as absolute number (percent) of the exactly mappable sequences for each category.

The chromosomal distribution of ISs was analyzed to determine the relationships among chromosome size, gene density, and insertion frequency. For each of the mouse chromosomes, the number of integrations was not related to the size of the chromosome, whereas a correlation between gene content and insertion numbers in both IDLV transduced samples and LV transduced samples was shown (Figure 4-17, A). This suggested that integration of IDLV and LV is dependent on the gene density of a chromosome, rather than its size.

To analyze the positional preferences of IDLV ISs *in vitro*, we mapped all the unique mappable IS on the mouse genome, as shown in Figure 4-17 B, IDLV ISs close to randomly distributed of each chromosome, no IS clustering near to centromeric or telomeric region was observed.

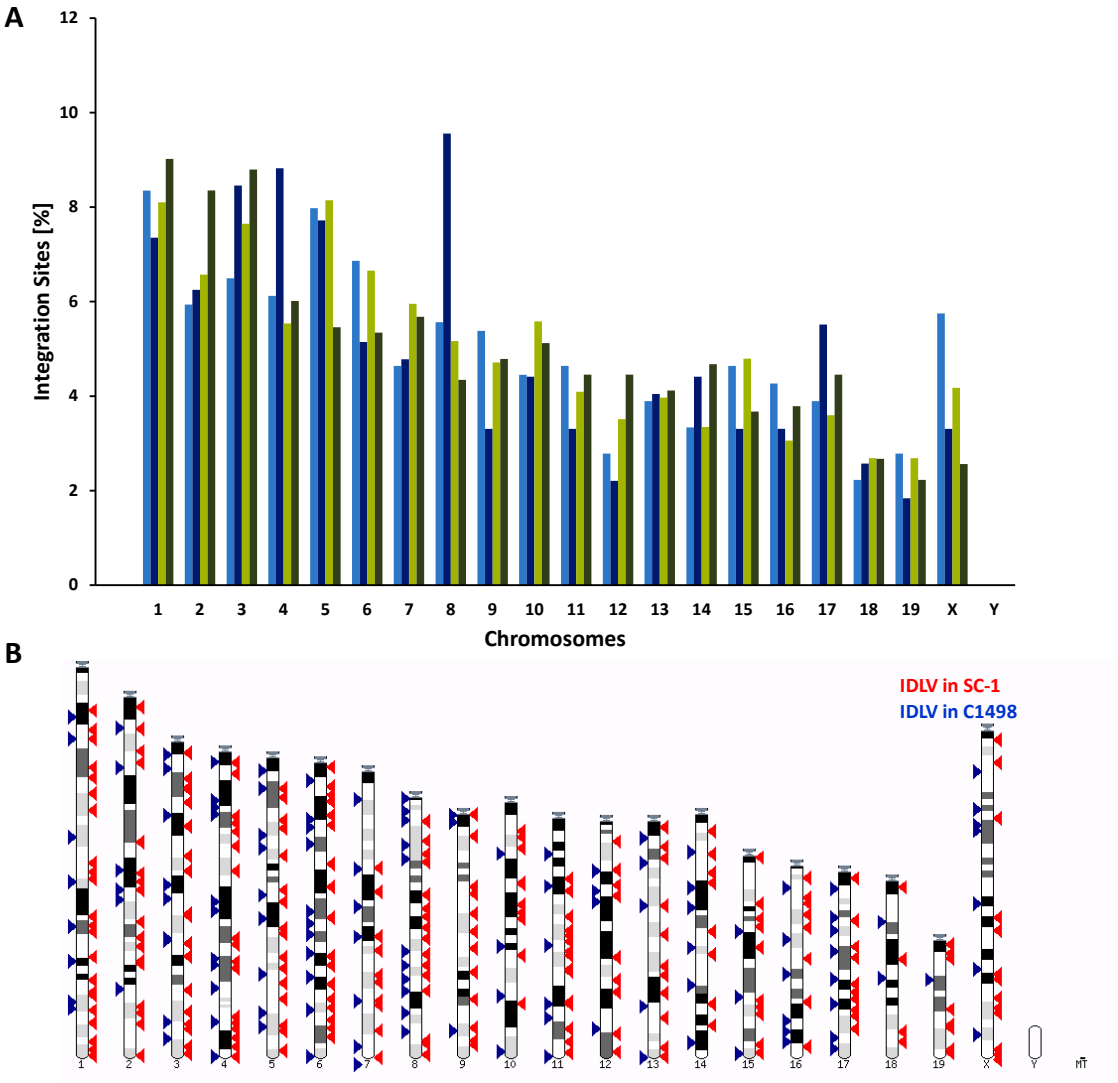

Figure 4-17 Genomic distribution of IDLV integration sites. (A) Distribution of ISs among chromosomes. Values for ISs are shown as a percentage of all ISs from the corresponding samples. Light blue bars, IDLV in SC-1 cells; dark blue bars, IDLV in C1498 cells; light green bars, LV in SC-1 cells; dark green bars, LV in C1498 cells. **(B)** Positions of IDLV ISs among chromosomes (Generated by Ensembl KaryoView). Each triangle represent a IS in responding samples. Red triangle, IDLV in SC-1 cells; Blue triangle; IDLV in C1498 cells.

4.3.7 *Close to random in vitro IDLV integration frequency in gene coding regions*

Previous studies revealed that integration sites from MLV and HIV-1 based vectors showed distinct differences in respect to annotated features of human genome. MLV based vectors showed significant clustering around the transcription start sites (TSS) or RefSeq genes and in the proximity to CpG-islands whereas HIV-1 based vectors disfavored TSS and CpG islands, but showed preferences for

integration inside transcribed regions of RefSeq genes (Schroder et al., 2002; Wu et al., 2003).

To dissect the IDLV IS pattern in vitro, IS for each of IDLV and LV transduced samples were mapped to the mouse genome, and nearby features were assessed (Figure 4-18). Given the number of genes in the mouse genome and assuming random integration, 25% of ISs would be expected to fall into or within a 10-kbp window around RefSeq genes, which account for approximately one third of the mouse genome. As shown in Figure 4-18 C, and Table 4-2, the actual frequency of IDLV IS in gene region (42.5% for SC-1 and 40.4% for C1498) and including the 10kb window up- and downstream (55.5% for SC-1 and 49.2% for C1498) was significantly lower than LV IS pattern, which showed a typical favorable integration pattern inside of RefSeq genes as expected (74.5% for SC-1 and 62% for C1498, p<0.0001, chi-square test). Both IDLV and LV insertion patterns were significantly different from the random integrations sites, of which only 32% were within RefSeq genes, a percentage identical to the average estimation of the mouse genome content (25.5%-37.8%, median 31.5%).

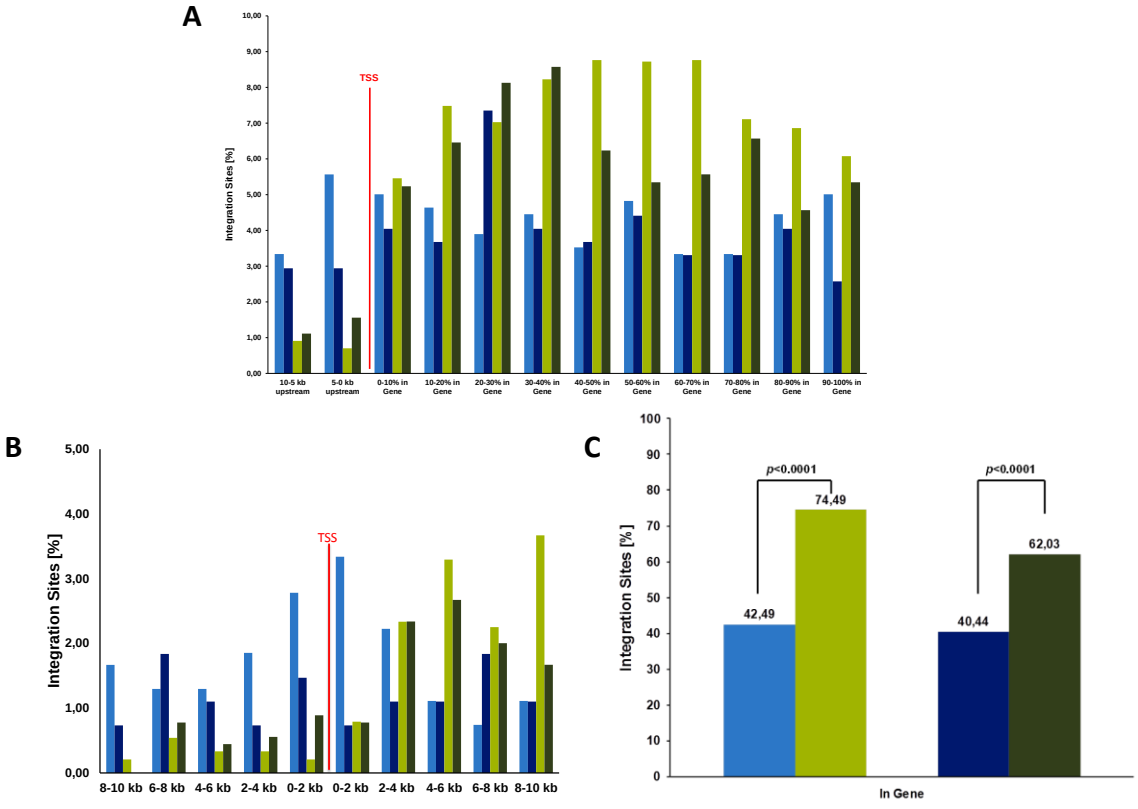

Figure 4-18 IDLV integration sites distribution in mouse RefSeq genes and the surrounding 10kb region. (A) The percentage of IDLV and LV IS in mouse RefSeq genes and **(B)** in RefSeq genes including surrounding 10 kbp region is shown. **(C)** Comparison of IDLV and ID IS inside of RefSeq gene region. ISs locations inside genes are expressed as the percentage of the overall length of each individual target gene. Light blue bars, IDLV in SC-1 cells; dark blue bars, IDLV in C1498 cells; light green bars, LV in SC-1 cells; dark green bars, LV in C1498 cells. TSS, transcription start sites.

We next examined whether specific regions of the transcription units were most likely sites of integration than others. We analyzed the distribution of the integration events within the transcription unit by dividing the distance of each integration site from the transcription start site by the gene length. The resulting ratio, reported as the total number of integration events in RefSeq genes for each vector, provides the percentage of integrations within ten equal sections of transcription units. While both IDLV and LV targets entire transcription unit with no noticeable preference, as expected, LV integration pattern exhibited a strong preference for integration near transcription start sites (TSS) - 63 of 3317 (2%, SC-1 and C1498) of the LV integration sites, as compared 64 of 811 (8%, SC-1 and C1498) for IDLV, are located within 10kb upstream of TSS (Figure 4-18 A).

To further explore IDLV preferential integration in the vicinity of TSS, we determined the distance to the nearest 5' and 3' ends of a RefSeq gene for each integration site. Interestingly, whereas IDLV integration events do not favor locations upstream or downstream of transcription units, 64 of 811 (8%, SC-1 and C1498) of the total IDLV integration sites landed within a 10-kbp region upstream of a RefSeq gene, as compared to 5% expected with the random integration sets ($p<0.0001$). The frequency of insertions within 10kb downstream of the 3' end is almost identical for the IDLV and the in silico-generated random sets (3.6% versus 4.8%, $p=0.057$).

We then looked at the proviral integration within a 2kb window on either side of transcription start sites. This survey revealed a strong tendency for IDLV vectors to integrate close to transcription start sites, with 27 of 539 (5%, SC-1) of the total IDLV integration events occurring within 2kb upstream or downstream, as compared to 24 of 2419 (1%, SC-1) for LV ($p<0.0001$). Interestingly, this finding was not observed in C1498 cell line, where 4 of 272 (1.5%) of the total IDLV IS as compared to 15 of 898 (1.7%) of the total LV IS located within 2kb window (Figure 4-18 B).

Taken together, these data demonstrated a distinct integration pattern between IDLV and LV *in vitro* ($p<0.0001$, Chi-square test): while the latter appear to integrate predominantly within transcription units, IDLV vectors integrate close to randomly into genome, when it presents, favor integration within a 10kb window centered on transcription start sites.

4.3.8 *Network analysis using ingenuity pathway analysis*

To define and compare specific physiological functions and networks included in the IDLV and LV IS datasets, we performed comparative ingenuity pathway analyses (IPA). Ingenuity analysis also offers further insights into gene data sets associated with specific diseases. In both IDLV and LV transduced samples, IS are mainly found in genes involved in genetic disorder and immunological disease. In “genetic disorder”, “immunological disease”, “cell death”, “cell cycle” and “caner” related gene classes, IDLV transduced samples showed significant lower ($p<0.05$, student *t*-test) overrepresentation as compared to LV transduced samples (Figure 4-19).

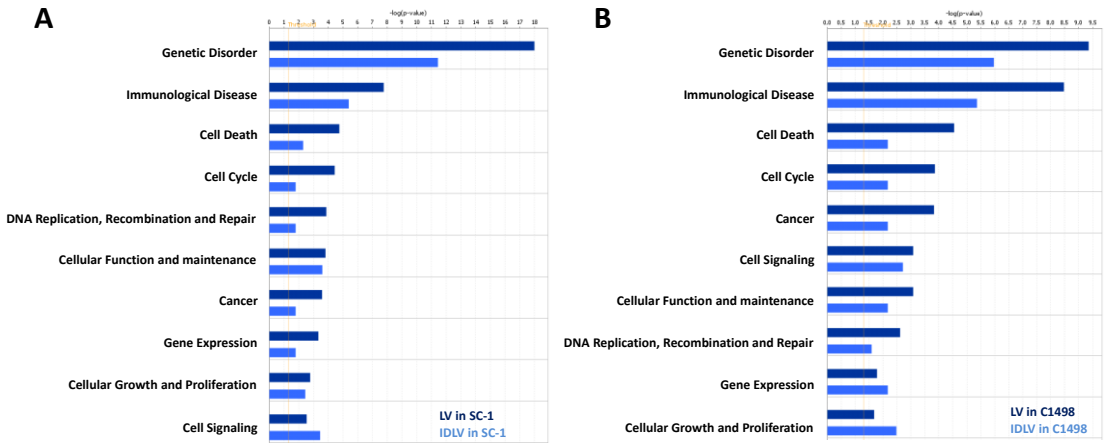

Figure 4-19 IPA analysis of LV and IDLV integrations. Genes of IDLV and LV transduced **(A)** SC-1 and **(B)** C1498 cells with an insertion within the gene or in the neighboring 10 kb were classified according to physiological function. The y-axis indicates the category to which the analyzed genes contribute. Only the significant categories were shown. For each analyzed gene group the statistical significance (student *t*-test) of overrepresented genes in a pathway is given on the x-axis.

4.3.9 *No proto-oncogenes were found in IDLV IS sets*

It is generally recognized that known preferences of gamma-retroviral vectors to target regulatory regions of genes increase the risk for genetic alterations of

neighboring genes (Baum et al., 2003; Nienhuis et al., 2006). In contrast, it has been shown that independent of the integration pattern also vector design and dose influence the likelihood of vector induced cellular transformation (Maruggi et al., 2009; Montini et al., 2009). To address this potential problem for IDLV application in clinical gene transfer, we compared the integration site datasets generated in our experiment with Retrovirus and Transposon tagged Cancer Gene Database (RTCGD, http://variation.osu.edu/rtcgd/index.html), which summarized potential or known proto- and oncogenes from multiple high-throughput insertional mutagenesis screening projects, using retroviral vectors or transposons as tools to discover potent cancer gene in mouse hematopoietic tumors or other tumor types (Akagi et al., 2004; Collier et al., 2005; Starr et al., 2009). No match was found in 503 known RTCGD genes from 804 unique mappable IDLV integrants, especially the overrepresentated *MDS1/EVI1* and *PRDM16* loci found in other vector. The T-cell acute lymphocytic leukemia (T-ALL) oncogenes *LMO2, TAL1, TAN1, LCK, LMO1, HOX11, HOX11L2, LYL1, TAL2, C-MYC* and *CCND2* were not present in the IDLV IS data sets.

4.3.10 *Establishment of Two-directional LAM-PCR*

We used (nr)LAM-PCR strategies that could detect LTR- and non-LTR-mediated vector integration, which we estimated allowed scanning of 50% to 70% of the vector genome for host DNA-vector junctions. But LAM-PCR on genomic DNA of these nonclonal samples only analyses one end of the integration events, so no conclusions are possible regarding the presence or absence of the flanking 5-bp host DNA duplication, LTR sequence deletion, genomic DNA deletion as well as addition . To further understanding the mechanism of IDLV integration, we used two strategies to overcome this obstacle: First, we performed 5’ and 3’ strand LAM-PCR separately, combined the IS we obtained from individual experiment, and then selected the unique IS which has both strand vector-genomic DNA junction; Second, we established a mixed, so called two-directional LAM-PCR, which both strand LAM-PCR could be performed in the same reaction tube.

The concept of two-directional LAM-PCT is instead of using only one strand LTR primer for linear PCR, by performing the same experiment with both strand LTR primer sets. For exponential PCR, we used equal amount of both strand LTR primer sets and linker sets instead of one strand sets. To avoid any methodological bias introduced by increasing the primer amounts, the same concentration of primer sets were used as single strand based LAM-PCR.

To evaluate the feasibility of this method, we performed both single strand and two-directional LAM-PCR on the single HeLa cells single clones transduced with lentiviral vector, which their integration sites are known. As shown in Figure 4-20, 5' strand LAM-PCR was performed using MseI as restriction enzyme, resulted an internal control band with around 110bp; 3' strand LAM-PCR was performed using MseI as well, resulted an internal control band with around 290bp; when two-directional LAM-PCR was performed using MseI, both internal control of 110bp and 290bp were shown, indicating that the combined method is capable to detect both direction of vector-genomic DNA junction in the same reaction, no noticeable changes of IS pattern and decreases of signal density were seen in the gel electrophoresis. Two-directional based nr LAM-PCR was also tested and functional (data not shown).

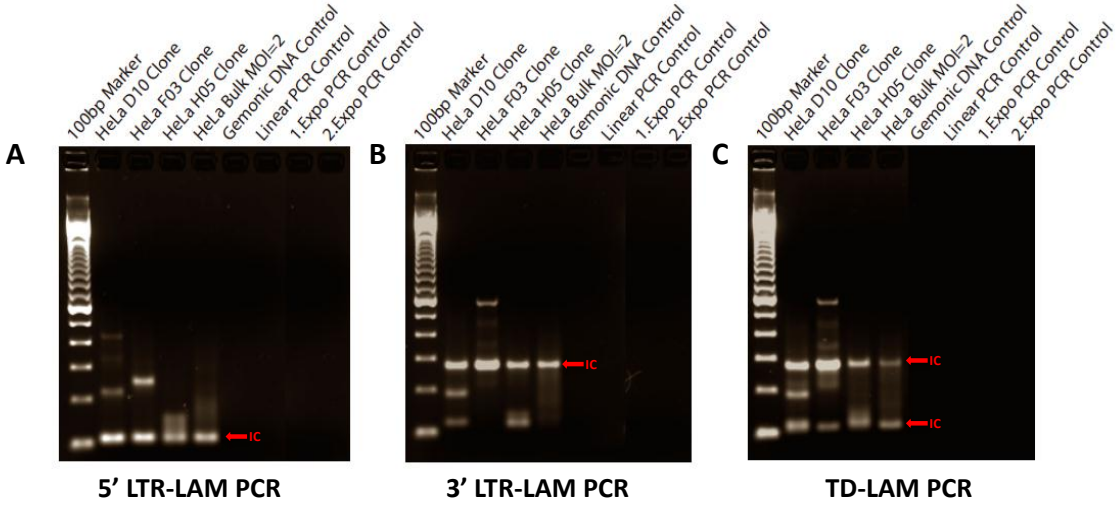

Figure 4-20 Two-directional LAM-PCR. (A) 5', **(B)** 3' LAM-PCR and **(C)** two-directional LAM-PCR were performed on LV transduced HeLa single cell clone D10, F03, H05 as well as bulk samples transduced with MOI=2. IC: internal control

To further evaluate the two-directional LAM-PCR, the amplicons generated from 5’ and 3’ strand LAM-PCR and two-directional LAM-PCR were sequenced by high-throughput sequencing using 454 platform. As summarized in Figure 4-21, both in conventional and nr LAM-PCR, two-directional based LAM-PCR generated less sequence counts in comparison to single-directional one, except one sample of HeLa clone D10, on the 3’ strand, 9 ISs with 300 sequence counts were retrieved in two-directional LAM-PCR sample, whereas only 5 ISs with 277 sequence counts were retrieved in 3’ strand LAM-PCR sample.

Of note, the frequency of IS retrieved by different method was different. It was not surprisingly in single strand based conventional LAM-PCR because of the usage of restriction enzyme and inequipotential distribution of motifs for enzyme cutting among the genome. IS retrieved by nr LAM-PCR showed less variant then conventional one, but for unknown reasons, in HeLa clone F03 and H05, two-directional based nr LAM-PCR didn't retrieve any sequence counts in 5’ strand after sequencing. Taking account with the less efficiency of nr LAM-PCR, few ISs were discovered by nr LAM-PCR in contrast to conventional method.

A

IS	LAM-MseI				nrLAM			
			2D				2D	
	5' Strand	3' Strand	5' Strand	3' Strand	5' Strand	3' Strand	5' Strand	3' Strand
CDH12				0,33				
CHST1		0,36						
CLMN			3,57					
DAP				0,33				
FGF9					0,18			
FLRT2		0,36		1,00				
KIAA0947[1] 5487373 (chr. 5, +)	84,62	93,14	85,71	94,35	32,60	35,69	100,00	45,40
KIAA0947[2] 5487558 (chr. 5, -)				0,33				
MYH7					0,18			
NCAM1			3,57					
PCNXL2	10,77	0,36	3,57	1,99	14,47	42,12		17,79
PPAPDC1A					0,18			
PRL				0,33				
SLITRK1[1] 83740807 (chr. 13, -)	4,62	5,78	3,57		51,10	22,19		36,81
SLITRK1[2] 83739582 (chr. 13, +)				0,33				
SLITRK1[3] 83740759 (chr. 13, +)				0,66				
TBX22					0,37			
IS\Seq.Count	3\130	5\277	5\28	9\300	7\546	3\311	1\5	3\163

10%-100%
1%-10%
0.1%-1%

B

IS	LAM-MseI				nrLAM			
			2D				2D	
	5' Strand	3' Strand	5' Strand	3' Strand	5' Strand	3' Strand	5' Strand	3' Strand
INSIG1[1] 155094853 (chr. 7, +)	96,88	100,00	100,00	77,78	100,00	100,00		100,00
INSIG1[2] 155094101 (chr. 7, -)	0,39							
INSIG1[3] 155094170 (chr. 7, -)	1,56							
INSIG1[4] 155094594 (chr. 7, -)	0,78							
NCAM1	0,39							
SPATS2				22,22				
IS\Seq.Count	5\256	1\92	1\49	2\63	1\622	1\387		1\389

C

IS	LAM-MseI				nrLAM			
			2D				2D	
	5' Strand	3' Strand	5' Strand	3' Strand	5' Strand	3' Strand	5' Strand	3' Strand
ACTBL2				0,71				
BCL9				0,71				
C7orf10				0,71				
DENND1B	0,26							
EBF1[1] 158448525 (chr. 5, -)	0,52			0,71	42,10	26,53		53,28
EBF1[2] 158515982 (chr. 5, +)	0,26							
EBF1[3] 158516074 (chr. 5, +)	2,59							
MTERF				0,71				
NDRG1				0,71				
PRELID2[1] 144907073 (chr. 5, -)	96,11	98,45	100,00	95,71	57,90	73,47		46,72
PRELID2[2] 144906347 (chr. 5, +)		1,56						
TMEM189-UBE2V1	0,26							
IS\Seq.Count	6\386	2\193	1\66	7\140	2\905	2\294		2\122

Figure 4-21 Sequencing summary of two-directional LAM-PCR. LV integrants retrieved in **(A)** HeLa D10, **(B)** HeLa F03 and **(C)** HeLa H05 were listed and compared between different methods. The percentage of each IS retrieved from individual sample and method was calculated by dividing IS with total sequence count. Red represents 10-100%; orange, 1-10%; yellow, 0.1-1%, respectively.

4.3.11 *Canonical and noncanonical integration of LV and IDLV*

In addition to providing viral-host DNA junction sequences, the two-directional LAM-PCR allowed us to determine the sequence of host DNA flanking the integrated provirus on both side. HIV-1 provirus are characteristically flanked by a 5-bp direct repeat of host cell DNA, the consequence of a 5-bp staggered cleavage of host DNA during IN-mediated strand transfer. To distinguish noncanonical integration from residual integrase activity-mediated integration, we screened our IDLV and LV integration data for the presence of the characteristic 5-bp direct repeat of host DNA flanking the proviral terminal CA dinucleotide as a hallmark of integrase activity. For IDLVs, we retrieved the 5'- and 3'-vector-host genome junctions in 22 instances from the bulk cell populations (statistical considerations make it extremely unlikely that such junctions could from two distinct integration events) and in 2 instances from single cell-derived clones. 17 of 24 integrants (71%) showed loss of the CA dinucleotide at least at one end of the vector. All 24 integrants revealed a partial

deletion of LTR and/or genomic sequences, and no 5-bp direct repeat could be detected in any of these vector-genome junctions (Figure 4-22, B).

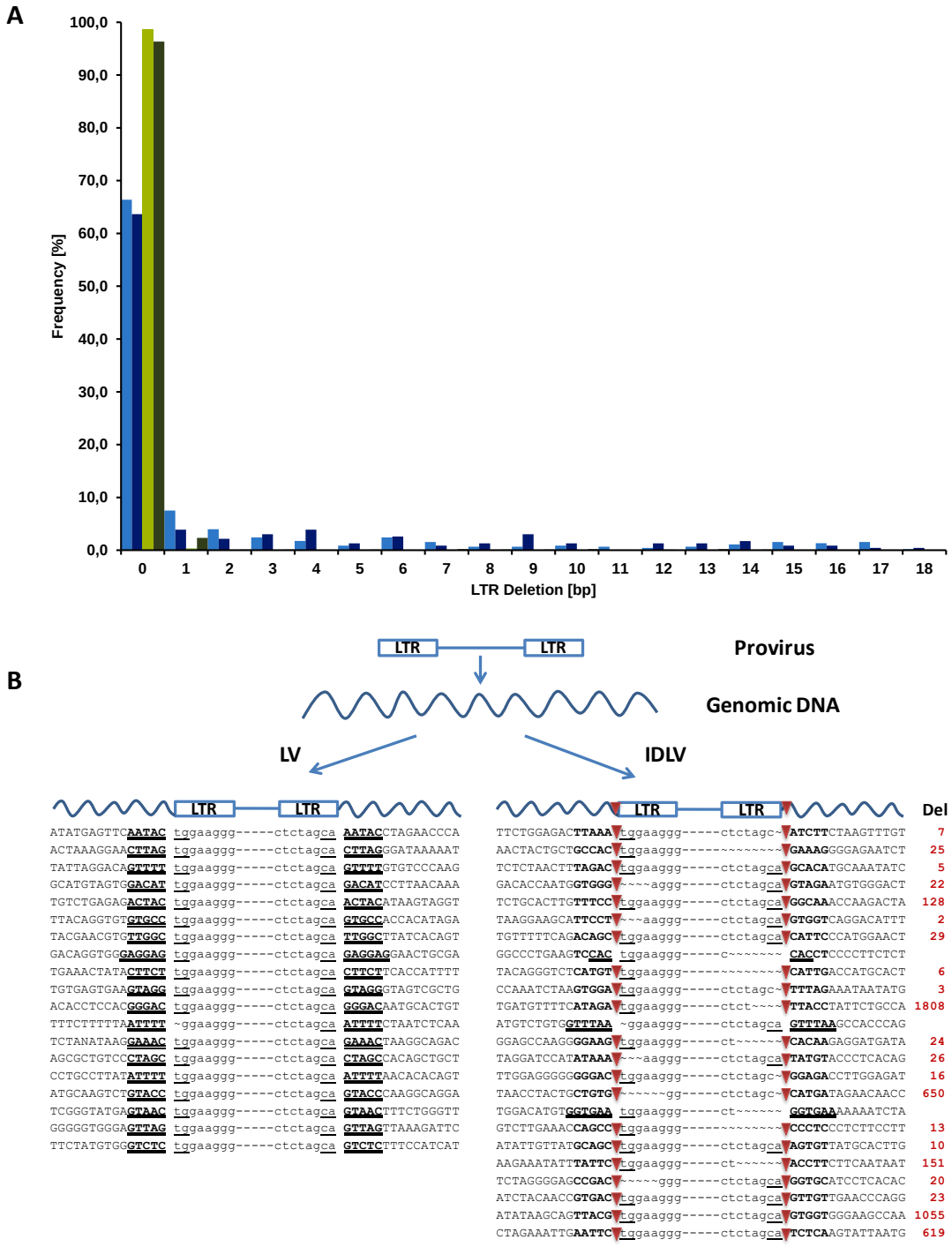

Figure 4-22 Canonical and noncanonical integration of LV and IDLV. **(A)** Frequency of LTR deletions in IDLV and LV transduced SC-1 and C1498 cells. Light blue: IDLV in SC-1; dark blue: IDLV in C1498; light green: LV inSC-1 and dark green: LV in C1498. **(B)** LTR deletions (swung dash) and genomic deletions (red triangles) identified in IDLV and LV integrants from which both the 5' and 3' genomic flanking region were identified by LAM-PCR (Matrai et al., 2011). The terminal CA dinucleotide (underlined, GT at end of 5' strand LTR, complementary to CA at end of 3' strand LTR) is shown if exist. The 5-bp direct repeat of host DNA flanking the provirus is delineated with double underline if exist. The number of deleted bp in host DNA is indicated (Del). LTR: long terminal repeat. LV: lentiviral vector. IDLV: integration deficient lentiviral vector.

In contrast, 18 of 19 LV integrations (94.7%) for which both sequence ends were mapped revealed neither CA dinucleotide deletions nor LTR sequence deletions, and they harbored the typical 5-bp direct repeat without genomic deletions (Figure 4-22, B).

To further investigate the nature of the strand transfer reaction, we screened for LTR sequence deletions in our complete IS data sets, as well as for target genomic DNA deletions in those two directional clones. Surprisingly, one-third of all IDLV integrations had a partial loss of the LTR sequence, whereas LTR deletions accounted for only approximately 2% of all LV sequences (Figure 4-22, A). Due to the limitation of primer sets used in this experiment, up to 18 bps deletion of LTR sequence could be detected.

The 5-bp direct repeat of host DNA normally flanking HIV-1 provirus is due to the left and right ends of incoming virus ligating to the host DNA in a 5'-staggered fashion. A deletion of target DNA could occur if an integration event involved a 3'-staggered cleavage of the target DNA, as opposed to the normal 5'-staggered cleavage. 21 of 24 two directional clones underwent a deletion during the viral integration process, with deletions ranging from 2 to 1808 bps (Figure 4-22, B).

4.3.12 *Integrated IDLV viral DNA is oligomeric*

Gene expression from non-integrated lentiviral DNA can be from linear DNA, 1- or 2-LTR circles, or even concatemers of the exogenously acquired DNA formed by intermolecular ligation (Perucho et al., 1980). To detect potential existing concatemeric integrated IDLV forms, we have used a combination of a nested PCR and LAM-PCR for the analysis of sorted IDLV transduced long-term cell lines. In theory, if concatemeric IDLV formations exist, 3 types of LTR concatemers could be detected: "Tail to Head"; "Head to Head" and "Tail to Tail". We designed 4 primers locate in different LTR region. All three potential concatemer formations could be detected using different primer combinations (Figure 4-23, A). We performed nested PCR for sorted IDLV transduced SC-1 and C1498 samples with MOI 25, 50 and 100, respectively, multiple bands were shown after agarose electrophoresis (Figure 4-23,

B). All the amplicons were tagged with LAM-PCR primers and sequenced with high-throughput pyrosequencing. Interestingly only "Tail to Head" formation could be retrieved from all assayed samples, suggesting that GFP expression derived either from remaining 2-LTR circles of IDLV or from "Tail to Head" IDLV concatemers. Statistically it is very unlikely that only "Tail to Head" concatemers could be detected if all three formations exist. The present of "Tail to Head" formation need be further evaluated by other experimental approach (such as southern blot) if it exists and the mechanism for "Tail to Head" formation remains unclear.

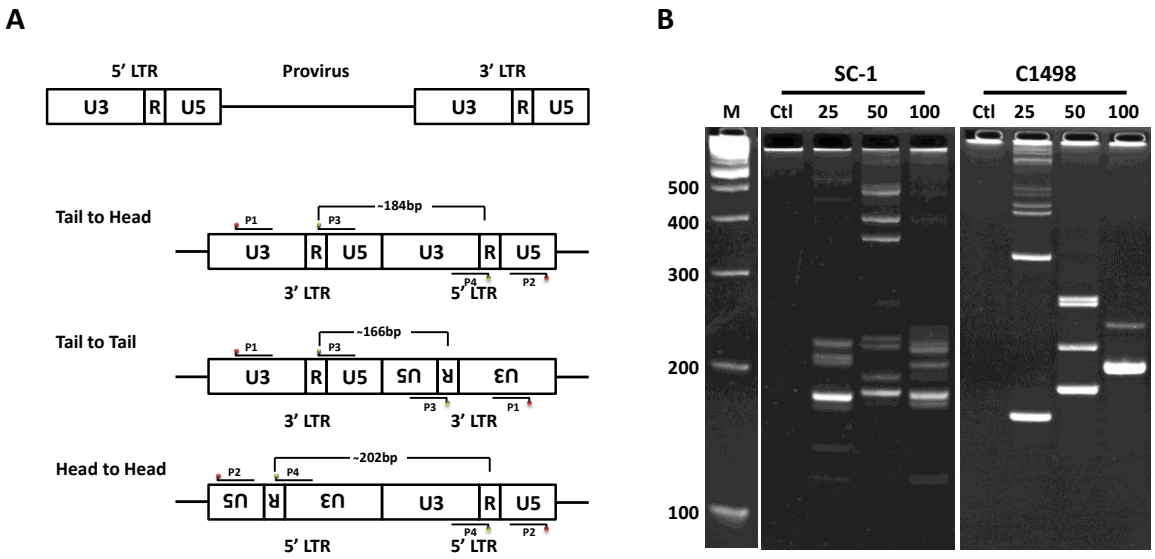

Figure 4-23 Integrated IDLV provirus is oligomeric. (A) Strategy to detect potential IDLV concatemers formations: "Tail to Head" formation should result a184 bp band; "Tail to Tail" formation should result a 166 bp band; and "Head to Head" should result a 202 bp band. LTR: long term repeat element. **(B)** Detection of concatemers in sorted IDLV transduced SC-1 and C1498 cells. M: 100bp marker; Ctl: control; 25: MOI=25; 50: MOI=50; 100: MOI=100.

4.3.13 *Integrated IDLV viral DNA is intact*

Some episomal DNA molecular as well as episomal virus, such as Epstein Barr Virus (EBV) and human papillomaviruses (HPV), often break anywhere on the episome and integrate into host genome through non-LTR-mediated mechanism. To address this aspect in IDLV transduced cells, we performed a modified LAM-PCR , so called internal LAM-PCR, using 3 primer pairs covered almost 75% of IDLV backbone to screening for the potential episomal breakage flanking host genomic DNA. One primer pair locates inside of viral packaging signal to monitor the formation of

provirus; the other two primer pairs locate beginning as well as ending of eGFP sequences to monitor the functional provirus. A combination of HpaI/FspI/StuI restriction enzymes, whose cutting motifs are absent in vector backbone, but often present in human genome, was used in this protocol.

Sorted and bulk unsorted IDLV transduced SC-1 and C1498 with MOI 100 were analyzed in this experiment. PCR amplicons were only seen in sorted SC-1 and C1498 cells assayed with packaging signal and GFP front primers (Figure 4-24). The amplicons were analyzed by short-gun cloning following by Sanger sequencing, as well as by pyrosequencing with 454 platform, only virus sequences were retrieved from both methods. It indicated that no breakage was occurred during residual IDLV integration, the provirus of integrated IDLV remains intact.

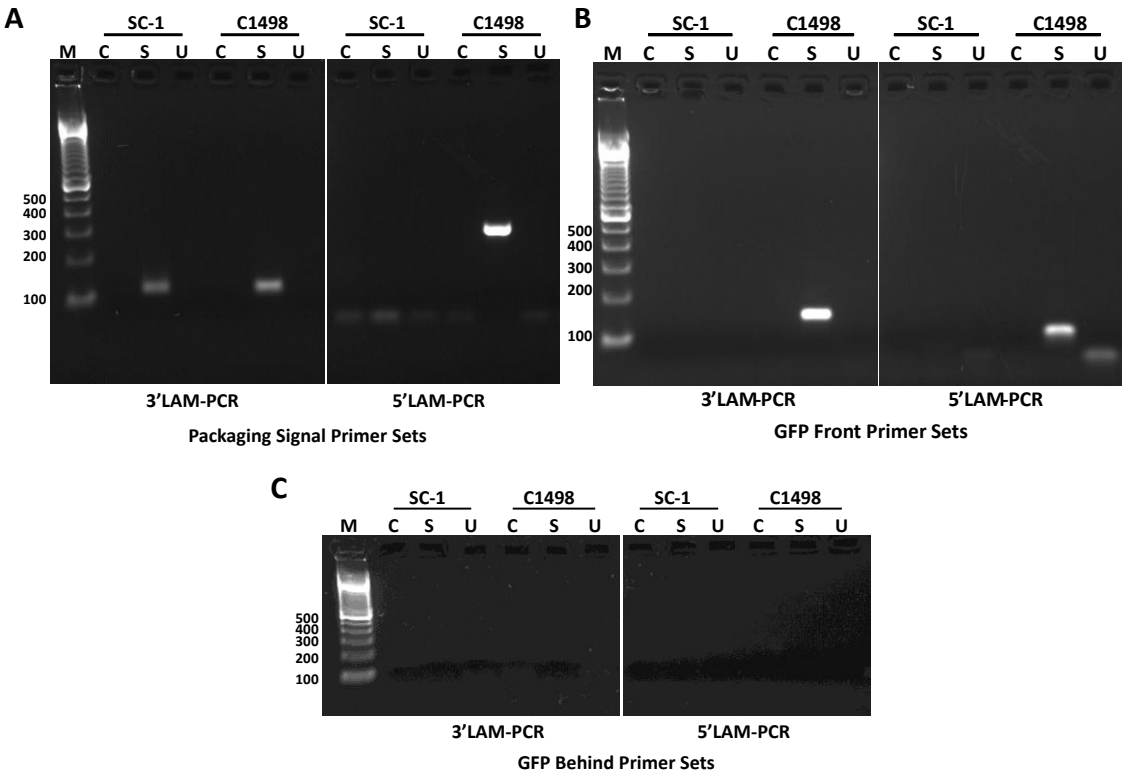

Figure 4-24 Integrated IDLV provirus is intact. Detection of IDLV breakage in sorted and unsorted IDLV trausduced SC-1 and C1498 cells by internal LAM-PCR using packaging signal primer sets (A), GFP front primer sets (B) and GFP behind primer sets (C). C: H_2O control; S: Sorted sample; U: Unsorted sample; M: 100bp marker

4.3.14 *Target site selection may not be random for IDLV*

While the location of retroviral integration within the host cell genome is not a random event, DNA structural characteristics, and not nucleotide sequence, appear to be the primary determinants that influence target site selection (Pryciak et al., 1992; Pryciak and Varmus, 1992). Previous work has shown that a favored palindromic sequence can be detected when many HIV integration sites are aligned, the inverted repeat structure likely originating from the symmetry of the IN-DNA complexes responsible for covalent DNA joining at each end (Berry et al., 2006). We asked whether the conserved favored primary sequences characteristic of HIV integration sites were present also in IDLV data sets. The primary chromosomal DNA sequence immediately flanking each integration site was scanned for base composition using the WebLogo sequence logo tool with ± 20bp resolution (Crooks et al., 2004).

For both LV and IDLV transduced SC-1 and C1498 cells, a host 'G' residue is disproportionately represented immediately flanking each end of the provirus, and there is a tendency for a 'T' at positions -2 and +2. In LV transduced SC-1 samples, 324 of 1283 cases (25.3%) for 3' strand and 242 of 1136 cases (21.3%) for 5' strand host DNA represent a weak favored 'GT' sequence selection, respectively. Similar results were found in LV transduced C1498 cells, 88 of 344 cases (25.6%) for 3' strand and 119 of 553 cases (21.5%) for 5' strand host DNA immediately flanking LV LTR were demonstrated (Figure 4-25, A and B). IDLV, however, demonstrated a striking preference for a 'GT' residue immediately adjacent to the conserved viral CA dinucleotide at the left viral end (44.7% for SC-1 and 47.8% for C1498, respectively). Similarly, a 'GT' residue was found immediately adjacent to the right viral end with 28.7% in SC-1 and 25.7 in C1498 cells (Figure 4-25, C and D). If we exclude the clones with truncated LTR ends, which may integrate via mechanism different than that used by clones with intact LTR ends, the percentage of 'GT' residue immediately adjacent to the left viral end raised to 62.4% for SC-1 and 63.4% for C1498 cells, and to 40.5% for SC-1 and 42.4% for the right viral ends.

The 3' end of each strand of unprocessed HIV-1 DNA ends in 'CAGT'. During a normal infection, IN-mediated 3'-processing removes the terminal 'GT' dinucleotide

from the 3' end of each strand of viral DNA. Each 3' end of the viral DNA is left with a terminal 'CA' dinucleotide. If the IN was mutated, the mutant virions failed to undergo 3'-processing prior to provirus formation, the so called flanking 'GT' residue would actually be of viral, not host, origin. These data clearly demonstrated that the residual IDLV integration was guided by different mechanism to wile type LV.

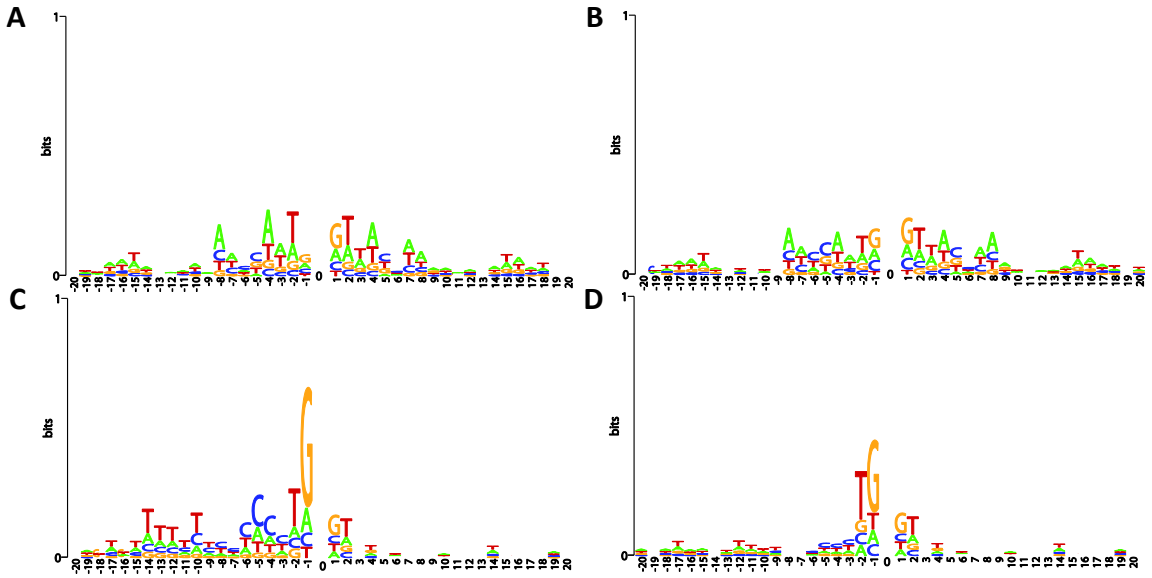

Figure 4-25 Primary sequence at integration sites of IDLV and LV *in vitro*. A sequence logo was generated using the WebLogo tool to identify preferred base pair usage at the LV transduced **(A)** SC-1 and **(B)** C1498 as well as IDLV transduced **(C)** SC-1 and **(D)** C1498. Position 0 denotes the first base of flanking LTR sequence 3' of the integration site. That is, the host DNA sequence is adjacent to the top strand of viral DNA on the right but from the bottom strand of viral DNA on the left. For sequence to the left of the provirus, host DNA at -1 is immediately adjacent to the provirus and -20 is located 20bp from the provirus; for the right side, host DNA at +1 is immediately adjacent to the provirus and +20 is located 20bp from the provirus.

4.3.15 *Short summary of this section*

Our large-scale IDLV IS data revealed IDLV represent the lacking of representative 5-bp direct repeat of host DNA flanking the provirus, deletion of host DNA and LTR sequences, abnormal 3'-processing and no concatemer integration. IDLV Integration pattern demonstrated random distribution of chromosomes, non-clustered around TSS sites, and less preference integration into Refseq genes as compare to lentiviral vectors (LV) integration pattern. No proto- and oncogenes loci as integration sites have been found in IDLV transduced cells. Taken together, the residual integration of IDLVs might be mediated by an integrase-independent mechanism, the risk of

insertional mutagenesis from IDLV is highly minimized and it provides one of the potentially safest tools for efficient gene delivery for clinical applications.

5. Discussion

5.1 Integration "hot spots" in MDS1/EVI1 and PRDM16 in WAS gene therapy

Gamma-retroviral DNA permanently inserts into the genome, usually near promoters of active genes. Thus, the major safety concern in gene therapy patients is that permanent insertion may occur in or near an oncogene to promote its upregulation and to lead to malignant transformation and leukemia. Retroviral insertion into the *LMO2* locus in children with X-linked severe combined immunodeficiency (X-SCID) was associated with development of leukemia after gene therapy. Therefore, we performed extensive insertion site analyses in an ongoing clinical gene therapy trial with patients suffering from Wiskott-Aldrich syndrome. Gamma-retroviral vector insertions in the *MDS1/EVI1* and *PRDM16* loci were founded to be highly overrepresented. To analyze the long-term impact of *MDS1/EVI1* and *PRDM16* retroviral integration sites in hematopoietic stem cell lineages, *MDS1/EVI1* and *PRDM16* loci as vector genome adjunction are monitored and analyzed by LAM-PCR and (nr)LAM-PCR prospectively more than two years after gene therapy in patient 1 and 2.

Patient 1 presented with hematoma and petechiae after birth, and within the first year of life, hemorrhagic colitis, eczema, and therapy-refractory autoimmune hemolytic anemia developed. In addition, he had autoimmune neutropenia and recurrent infections (bronchitis, otitis media, and enterocolitis). Patient 2 presented with thrombocytopenia and hemorrhagic diathesis soon after birth. He had recurrent episodes of upper respiratory tract infection, life-threatening salmonella septicemia, and eczema that was not responsive to local and systemic therapy. Molecular studies revealed mutations in *WAS*, no residual protein or somatic chimerism was detectable in both patients (Boztug et al., 2010).

In both patients, vector insertion site analysis of peripheral blood and bone marrow samples demonstrated sustained highly polyclonal reconstitution over time,

MDS1/EVI1 and *PRDM16* loci were evaluated as common insertion sites (CIS) in this study. Both conventional and non-restrictive LAM-PCR results revealed that gene-modified clones containing the insertion of *MDS1/EVI1* and *PRDM16* exhibited only minor contributions to hematopoietic repopulation within the first 2 years after treatment and remained constant. In Patient 1, *MDS1/EVI1* containing clones were first been detected at 7 weeks after gene therapy with a contribution of 0.6% in total 222 retrieved ISs. The proportion of *MDS1/EVI1* containing clones increased slowly and peaked at 4 months after gene therapy with 5.7% in total ISs detected. At 33 months, the last observation time point in this thesis, it remained constant with 5.2% in total retrieved ISs. In patient 2, *MDS1/EVI1* containing clones remained at very low level from undetectable to 1.8% in whole retrieved ISs during the entire observation period (up to 30 months). In both patients, *Prdm16* containing clones contributed minor in whole retrieved ISs, ranged from undetectable to 1.1%, with an exception in patient 2 at 2.5 months after gene therapy, which had 6.1% in 44 detected ISs. In sorted myeloid (CD15+) and lymphoid (CD3+) lineages, *MDS1/EVI1* and *PRDM16* containing clones were mostly restricted in the myeloid compartment of reconstituted hematopoiesis in both patients, suggested that MDS1/EVI1 and PRDM16 may contribute mainly to myeloid progenitor differentiation and proliferation.

EVI1 has been described as a proto-oncogene since its first discovery in 1988 (Morishita et al., 1988). Overexpression and aberrant expression of *EVI1* has been associated with human acute myelogenous leukemia (AML), myelodysplastic syndrome (MDS) and chronic myelogenous leukemia (CML), and more recently has been shown to represent a poor prognostic indicator (Metais and Dunbar, 2008). All of these malignancies involve erratic cellular development and differentiation in the bone marrow leading to dramatic alterations in the normal population of blood cells (Goyama et al., 2008; Yuasa et al., 2005). *EVI1* has also been found to play a role in solid ovarian and colon tumors, although it is not yet well characterized (Shackelford et al., 2006). It has been hypothesized that it acts as a survival factor in tumor cell

lines, preventing therapeutic-induced apoptosis and rendering the tumor cells more resistant to current treatments (Liu et al., 2006).

Prdm16 is expressed very selectively in the earliest stem and progenitor compartments, and, consistent with this expression pattern, is critical for the establishment and maintenance of the HSC pool during development and after transplantation (Deneault et al., 2009). *Prdm16* deletion enhances apoptosis and cycling of HSCs (Aguilo et al., 2011; Chuikov et al., 2010). Expression analysis revealed that Prdm16 regulates a remarkable number of genes both enhancing and suppressing HSC function, and affecting quiescence, cell cycling, renewal, differentiation, and apoptosis to various extents (Aguilo et al., 2011). Elevated *PRDM16* expression, because of promoter hypomethylation, is frequently observed in karyotypically normal AML (Shing et al., 2007). Deletion of the PR domain, which shows homology with a SET chromatin remodeling domain and is also present in MDS1/EVI1 (Mochizuki et al., 2000), appears important for the leukemogenic properties of human PRDM16. Translocations involving PRDM16 invariably delete the PR domain, whereas PR-deleted Prdm16 causes AML in p53−/− mice (Shing et al., 2007). Overexpression of Prdm16 expands HSCs in vitro. However, these expanded HSCs cause a myeloproliferative disease after transplantation (Deneault et al., 2009).

Strikingly, all the models investigating the impact of MLV vector transduction on primary hematopoietic cells results strongly suggest that hematopoietic stem and progenitor cells with insertional activation of *MDS1/EVI1* have a significant advantage. First, primary murine bone marrow myeloid progenitor cells are all immortalized by vector insertions either directly or indirectly activating the *Evi1* or the homologous *Prdm16* locus (Du et al., 2005). Second, extraordinary overrepresentation of *MDS1/EVI1* and *PRDM16* clones contributing to the myeloid lineage were seen over the long term in rhesus macaques (Kim et al., 2009) and human CGD patients (Ott et al., 2006), respectively. Third, serial transplantation in the mouse model increased the dominance of clones with insertions in *Mds1/Evi1* and *Prdm16*, as well as a limited number of other genes that encode proteins implicated in self-renewal and stem cell survival (Modlich et al., 2009).

The vector integration pattern in the analyzed WAS patients remained highly polyclonal and no obvious clonal dominance have been observed after two years of gene therapy. But this work accentuates the need for safer gene therapy vectors that will not activate oncogenes. Ongoing efforts to make gene therapy safer include: 1) assuring that the inserted virus does not activate adjacent genes by removing all promoter activity except for that which drives expression of the therapeutic gene; 2) insulating expression of the transgene from the surrounding genomic DNA; and 3) targeting integration to "safe" genomic regions.

5.2 Effects of Mds1/Evi1 and Prdm16 in hematopoiesis

The increasing number of gene therapy clinical trials resulting in patient clinical benefit highlights the rising use of gene therapy to treat a variety of disease conditions. Many trials have used retroviral vectors which are successful as they integrate their genetic material into patient chromosomes. This allows therapeutic DNA to be passed to daughter cells during mitosis, enabling stable gene expression in dividing tissues. However, five cases of acute lymphoblastic leukemia in two SCID-X1 clinical trials caused by integration of the retroviral genome near proto-oncogenes including *LMO2* have raised concerns about the safety of integrating vectors.

In patients with CGD, rhesus macaques, and mice receiving hematopoietic stem and progenitor cells transduced with retrovirus vectors, a highly non-random pattern of vector integration has been reported. The most striking finding has been overrepresentation of integrations in two specific genomic loci, a complex containing the *MDS1* and the *EVI1* genes, as well as the locus of the *PRDM16* gene. Most evidence suggests that this overrepresentation is primarily due to a modification of primitive myeloid cell behavior by overexpression of *MDS1/EVI1, EVI1* or *PRDM16*, as a specific preferential integration site of viral vectors. These phenomena could be accounted for via two general underlying mechanisms. Either these loci are extraordinary "hot spot" for retroviral integration in hematopoietic stem and progenitor cells, or progenitor cells that happen to acquire an insertion in those loci

have a unique propensity to engraft, persist, expand and/or contribute to hematopoiesis after transplantation. Although components of both mechanisms may be operative, the preponderance of evidence suggests that in *vitro* and in *vivo* selection for survival and hematopoietic output from clones with insertional activation of the *MDS1/EVI1* complex and *PRDM16* gene can account for the findings.

To test whether non-integrating vectors derived transient expression of *MDS1/EVI1* and/or *PRDM16* is sufficiently stable to influence long-term repopulating hematopoietic cells, we established a lentiviral based transient and stable expression system for this purpose. Protein coding sequences of murine Mds1/Evi1, Prdm16 and their PR domain lacking variants Evi1 and sPrdm16 were directionally cloned into LV constructs and verified by sequencing. Functional LVs were produced by transient co-transfection of 293T cells with envelope, packaging and transfer vector constructs. Reliable titers in the range from 10^7 to 10^9 TU/ml were achieved after optimization. Sufficient transgene expression was detected for all constructs by Western-Blot and RT-PCR after transduction of murine SC-1 cells. IDLVs expressing Mds1/Evi1, Prdm16 and their PR domain lacking variants Evi1 and sPrdm16 were also constructed and produced.

To study the direct impact of *Prdm16* and *sPrdm16* expression on hematopoietic differentiation and proliferation, lin- murine primary BM cells were transduced with Prdm16 or sPrdm16 LVs and cultured for 30 days in *vitro*. Long-term eGFP expression was observed in all three transduced cells populations for up to 30 days. This suggested that lentiviral mediated stable transgene expression in hematopoietic stem and progenitor cells are feasible. However, no significant differences in the proliferation potential or differentiation capacity were observed between Prdm16, sPrdm16 and eGFP transduced cells. This could be accounted for the poor gene transfer efficiency in primary murine BM cells mediated by LV and the dilution steps during the process of the experiment. In each observation time point, only certain numbers of transduced cells were kept further in culture due to the volume capacity of in *vitro* medium and space. Hematopoiesis is a highly hierarchy system, the HSCs constitute only 1:10.000 of cells in myeloid tissue. In this experiment, although lin-

BM cells were assayed, the frequency of HSCs is still below 0.01%. Combined with the relative low transduction efficiency of <3%, there was only a small likelihood to have 'true' stem cells transduced in this setup.

There are several possible ways of improving gene delivery efficiency to stem cells. First, we employed several optimization steps to increase our lentiviral titers. Several batches of FBS from different vendors and sources were tested since the FBS quality is reported as a crucial factor for LV production. Kozak consensus sequence was cloned in front of the transgene to enhance the translation procedure as well. The reliable functional titers increased approximately about 3- to 4- fold after optimization. Second, we tried different transduction protocols, including spinoculation; magnetic beads based transduction; two-round transduction procedure and others. Magnetic beads based transduction was evaluated as the most stable and reliable procedure and selected as standard protocol for LV transduction in primary cells. Third, protocol to separate more defined MPP, LT- and ST- HSCs by FACS using cell surface markers (i.e. CD34, Sca1, cKit and Flt3) was established. Taking advantage of the short transduction procedure of LVs, we could significantly increase the transduction efficiency in more defined cell types. These developments and improvements in gene transfer parameters enabled further studies on the impact of transient expression of *MDS1/EVI1* and *PRDM16* on proliferation, engraftment and expansion of HSCs *in vitro* and *in vivo*, beyond the scope of this thesis.

5.3 LV and IDLV differ in their integrations characteristics

Integration deficient lentiviral vectors (IDLV) use the advantages of lentiviral based vectors without the potential risk of insertional mutagenesis. As these vectors do not integrate into the host chromosomes and lack a replication signal, the viral DNA will be diluted out during cell division. Therefore, IDLVs will be particularly useful, for example, in post mitotic tissue gene therapy (Wanisch and Yanez-Munoz, 2009).

IDLVs can be produced through the use of integrase mutations that specifically prevent proviral integration, resulting in the generation of increased levels of 1- and 2-LTR circular episomes in transduced cells. Compared to integration proficient lentiviral vectors IDLVs have a significantly lower risk of causing insertional mutagenesis and a lower risk of generating replication competent retroviruses (RCRs).

Few studies on integration of IDLVs have been published (Gaur and Leavitt, 1998; Nightingale et al., 2006). Available data include analyses of residual proviruses and host DNA–vector junctions in a very limited data set of integration events, commonly not exceeding 10 events, and quantitation of integration frequencies is based on using selectable transgenes or eGFP expression. Thus, little is known about the existing residual integration events of IDLV, and comprehensive characterization of IDLV integration profiles in mammalian genomes is still missing.

In this large-scale LAM-PCR analysis of integration events from IDLVs, we used a high dose (MOI 25, 50 and 100, respectively) vectors carrying integrase D64V mutation and expressing eGFP to transduce two murine cell lines, C1498 and SC-1. Despite the low level of eGFP expression from all three doses of IDLVs after 4 weeks, we did consistently observe low levels (from <0.5%-2.3%) of cells persistently expressing eGFP, depending on the type of cell lines and vector dose. We hypothesized that inadvertent integration was responsible for the small amount of sustained gene expression we observed. Those eGFP expressing cells were then cultured for another 4 weeks to dilute out the residual nonintegrated viral episomes, sorted by FACS and analyzed for IDLV integration pattern.

The highly sensitive conventional LAM-PCR technology as well as the newly developed non-restricted LAM-PCR has been employed as a powerful and standard tool to identify lentiviral integration sites. With minimal amounts of DNA input, LAM-PCR has the potential to characterize integration site preferences of retroviral and lentiviral vectors, and to determine the clonal complexity in gene therapy research. We retrieved 905 integrants from IDLV transduced SC-1 cells and 437 integrants

from C1498 cells, respectively. Among them, 533 (SC-1) and 271 (C1498) integrants could be mapped to a definite position in the mouse genome. In comparison to IDLV, much more integrants from LV transduced samples (2788/2419 for SC-1, 1062/898 for C1498, total/unique mappable) were retrieved. These large IS data sets allowed a confident LV and IDLV IS analysis. The chromosomal IS distribution of the experimental groups resembled near random in either IDLV or LV. In some cases the distribution of IDLV IS showed a significant deviation in two cell lines, also an under- or overrepresentation of RefSeq genes at certain chromosomes was found (chromosome 2, 4, 8, 10 and X). It might be explained by different integration mechanism between the IDLVs and LVs, as well as the different genetic background between cell lines.

Previous studies have found that two thirds of LV IS are commonly located in genes, these results were observed consistently in *in vitro* and *ex vivo* experimental settings with different HIV-1 based LV and SIN-LV in varying cell populations (Beard et al., 2007; Felice et al., 2009; Laufs et al., 2006; Lewinski et al., 2005; Mitchell et al., 2004; Montini et al., 2009; Schroeder et al., 2002; Wang et al., 2009). Only recently, our group found that these results cannot be extrapolated to all cell types and tissues. Instead, a reduced IS frequency into gene coding regions was noticed in postmitotic tissue (eye and brain) after *in vivo* injection of LVs (Bartholomae et al., 2011). The reduced IS in gene coding regions went in line with a reduced expression of Psip1, which results the LV preintegration complex to specific, gene associated genomic loci. Underexpression of the Psip1 and gene associated preference of lentiviral vector integration loci could also be found in terminally differentiated macrophages and growth arrested cell lines (Barr et al., 2006; Ciuffi et al., 2006). Taken together, the discrepancies of lentiviral integration pattern does not seem to simply relate to the mitotic active versus quiescent cellular state, but might also reflect cell type / tissue specific differences. In our study, Integration pattern of IDLV showed significant lower gene associated preference than LV, indicating that the residual IDLV integration events are underlying different mechanism to LV.

Remarkably, a significantly reduced LV IS frequency was observed for the close upstream genomic region of the transcription start site. This was supported by the finding that nearly no LV IS were found within a 1kb distance to a CpG island, which are typically located at TSS (Saxonov et al., 2006). This was consistent with the previous experiments where no LV IS was found within the 1kb distance to a CpG island in postmitotic rat eye (Bartholomae et al., 2011). In comparison to LV, IS distribution of IDLV is nearly uniform in the genome, including gene and CpG island rich regions. Interestingly, the IS pattern of IDLV showed significantly to be more clustered around the TSS region compared to LV, suggesting the integration of IDLV might more easily occur through opened chromatin structure and might be mediated by other mechanism.

5.4 IDLVs integrate only to background levels

The peculiarities showed by proviruses from IN mutant HIV-1 and IDLVs, together with the lack of effect on residual integration frequencies from adding att mutations to IN mutants, have led to the conclusion that integration events from IN mutants are unlikely to be IN-mediated (Gaur and Leavitt, 1998; Nightingale et al., 2006). In support of this, integration of an IDLV has been observed in a pre-existing DSB in human osteosarcoma U-2 OS cells consistent with the known frequent capture of extrachromosomal DNA fragments and vector genomes by pre-existing genomic DSBs (Cornu and Cathomen, 2007; Lin and Waldman, 2001). However, the lack of IN involvement in some of the residual integration events has not been formally demonstrated, and it could be argued that residual IN activity could have aberrant characteristics and hence the products would not necessarily display all the features of wild type proviruses.

An analysis of relative integration frequencies of replication defective, integrating and IN mutant HIV-1 variants encoding a hygromycin resistance gene showed that mutations in the catalytic triad led to reductions of 3–4 logs in the number of drug-resistant colonies with the latter (Leavitt et al., 1996). Similar results have been also obtained in this thesis in the context of SIN IDLVs. Relative

integration frequency in C1498 and SC-1 cells have been estimated as about 350- to 1000- fold lower than wild type for D64V mutant, regardless the cell lines and vector doses.

5.5 Noncanonical integration of IDLVs

The product of HIV reverse transcription is a linear double stranded DNA molecule (with a discontinuous plus strand) (Miller et al., 1995). Linear double stranded DNA can be integrated non-specifically in the host genome. When there is a double strand break in the host genome, the DNA is repaired either by homologous recombination or non-homologous end joining. The last pathway involves the ligation of two double stranded DNA molecules, not requiring any homology between them. Using this mechanism, the cells may then integrate the viral linear DNA produced upon reverse transcription or viral DNA circles that were broken at double stranded break points in the genome. In either case it would be reasonable to find host DNA–vector junctions involving the ends of the vector LTRs, and assays for the analysis of integration junctions are normally biased to detect such LTR-mediated events.

The viral locus responsibility for retroviral IN function was originally described in 1984 landmark papers (Donehower and Varmus, 1984; Panganiban and Temin, 1984). One of these studies already analyzed that residually integrated IN mutant MoMLV proviruses (Donehower and Varmus, 1984). Only in very rare cases had IN mutant proviruses the structures expected from wild-type virus, i.e., two nucleotides lost from each end of viral DNA and identical 4-bp (or 5-bp) flanking sequences of host origin (Hagino-Yamagishi et al., 1987). For the wild type LVs, we predominantly observed the hallmark signature of integrase-mediated integration, in the most majority (>98%) perfectly intact LTR sequences at the integration site. Strikingly, sequencing of integration junctions isolated by LAM-PCR showed that the integrated proviruses from IN mutants did not confirm the canonical characteristics of integrated LVs: (i) the 5-bp direct repeat of host DNA was often not maintained; (ii) integration often occurred with truncations of vector DNA ends, perhaps due to

aberrant 3'-processing; (iii) in some cases integration caused a deletion of host DNA at the insertion site; and (iv) target site selection of mutant IN viruses showed sequence preferences not observed with wild-type proviruses (Gaur and Leavitt, 1998).

Regarding the structure of the integrated IDLV proviruses, a study in human colon adenocarcinoma HT-29 cells transduced with HIV-1 vector showed gross reorganizations by Southern blot in about 50% of G418-resistant colonies (Nightingale et al., 2006). Inverse PCR-based strategies on genomic DNA of nongrossly reorganized proviruses from IN mutant vectors, followed by DNA sequencing, demonstrated noncanonical patterns including deletions, duplications or reorganizations of vector sequence and lack of the characteristic 5-bp duplication of flanking host DNA (Gaur and Leavitt, 1998; Nightingale et al., 2006). From large-scale data sets of 800 and 3300 unique mappable IS from IDLVs and LVs transduced samples, respectively, we retrieved 24 'clones' from IDLVs and 19 'clones' from LVs, which showed both 5'- and 3'- vector/genomic junctions (statistical considerations makes it extremely unlikely that such junctions could originate from distinct integration events). For the LVs, our data demonstrated the hallmark of signature of integrase-mediated integration, with perfectly intact LTR-sequences (18/19) and duplication of 5 bp of the cellular DNA sequences at the integration sites, no genomic junction deletions/additions were observed. In contrast, the IDLV data set showed abnormalities of the LTR/genomic structure, including CA dinucleotide deletions (24/24), LTR sequences deletions, as well as genomic junction deletions/additions. In agreement with these data, screening for LTR sequences in our complete IS data sets showed that one-third of all IDLV integrations had a partial loss of the LTR sequences, whereas LTR deletions accounted for only approximately 2% of all LV sequences.

Circular IDLV episomes may be subjected to DNA damage non-LTR-mediated genomic integration. Very few attempts have been made to detect such host DNA–vector junctions not involving the LTR ends, but the stable detection of 2-LTR HIV-1 junction in dividing cells is supporting evidence for non-LTR-mediated integration of

viral circles (Brussel and Sonigo, 2003). In an effort to elucidate the mechanism underlying the rare integration by IDLVs, we analyzed the integrated vectors by LAM-PCR based internal PCR using 3 sets of primers that covered nearly 75% of the vector genome. The amplicons were sequenced per deep sequencing for 5000 raw sequencing reads and validated per Sanger sequencing in 96-well plate format. Since the sorting procedure particularly enriched the intact form of integrated provirus, unsorted samples were also investigated in the same setting. After extensive searching either in sorted or unsorted samples, no virus/genomic junctions have been identified, suggesting that neither rearrangement in the integrated DNA nor random breakage of circular IDLV episomes was present.

Another hypothesis that IDLVs might integrate into the host genome through concatemer formats by homologous integration (Hagino-Yamagishi et al., 1987). To address this question, nested PCR on the LTR region was performed based on three concatemer hypothetical formats: “Head to Head”, “Tail to Head” and “Tail to Tail”. The amplicons of two samples were sequenced by deep sequencing, and analyzed for LTR formations. Surprisingly, only “Tail to Head” format has been detected in around 2000 reads. These results are most simply explained by the presence of 2 - LTR circle in the analyzed samples, although the cells were cultured for 8 weeks post transduction. It is very unlikely that only one format but no other formats were present in case of existing concatemers. It might also be argued that the nested PCR is performed only within LTR sequences; no LTR/genomic junctions could be detected. But by combining the data with only intact provirus (disregard LTR sequences) presented in the IDLVs transduced samples, we could claim that IDLVs integrate into the host genome in a more or less intact and oligomeric formation.

5.6 Target site selection of IDLV integration

Target site selection by retroviruses is not random (Scherdin et al., 1990; Shih et al., 1988), but structural aspects of the target DNA, and not DNA sequence, appear to be most critical for determining if a given stretch of DNA will be the site of integration (Pryciak et al., 1992; Pryciak and Varmus, 1992). Both LV and IDLV

demonstrated a marked preference for a 'G' residue immediately adjacent to the conserved 'CA' dinucleotide found at the end of each strand of viral DNA, along with a tendency for the 'G' to be followed by a 'T' residue (Figure 4-25). Notably, IDLV demonstrated more preference in this phenomenon compared to LV. Regions of HIV-1 IN near D64 and E152 have recently been demonstrated to come into close contact with viral *att* sites and the target DNA (Heuer and Brown, 1997). That observation makes it intriguing to think that the mutations that we introduced directly affected target site selection. Possibly an unknown cellular protein interacts with the mutant IN proteins to affect target site selection. Alternatively, the 'G' preference seen in Figure 4-25 may indicate that IN is so crippled by the mutations that very few, if any, of the integration (strand transfer) reactions occurred via IN-mediated transesterification. In that case, integration sites may be influenced by simple base pairing between the viral ends and target DNA, with cellular ligases performing the actual integration events. This hypothesis predicts a 'GT' sequence immediately adjacent to the 'CA' dinucleotide at each viral end. Consistent with this, we found a favored 'GT' residue immediately flanking the virus at both viral ends. If this hypothesis is true, it could predict the predominant flanking host nucleotide for other retroviruses with similar IN mutations.

Bi-directional clones sequencing revealed that in some cases has few nucleotides to the left and right of the integrated viral DNA that are not of host origin. The extra DNA places a 'GT' dinucleotide immediately adjacent to the 'CA' dinucleotide found at the end of each viral DNA strand. While we do not know the exact origin of the extra DNA, the 'GT' immediately flanking each viral 'CA' dinucleotide is most likely from a viral genome that did not undergo 3'-processing prior to integration. The additional extra nucleotides on the left and right have no obvious source. Others have described extra nucleotides at the ends of HIV-1 DNA thought to be due to reverse transcriptase-mediated addition of non-template-encoded bases (Miller et al., 1997; Patel and Preston, 1994). While this is possible, the extra sequence could also be related to the mechanism underlying the recombination event that lead to this particular clone. While the target site sequence data strongly suggest that

mutations to the conserved acidic residues of the D,D(35)E motif produce a bias toward integrating at a host 'G' residue, they do not allow us to conclude that the G residue flanking the provirus is of host origin. We cannot distinguish between (i) the integration of 3'-processed viral DNA via a mechanism that retains the host 'G' residue and (ii) the integration of unprocessed viral DNA via a mechanism that deletes the host 'G' residue. Consequently, we were unable to address the effects of our mutants on 3'-processing during viral infection.

5.7 In vivo applications of IDLV

There are only two published analysis of genomic integration of IDLVs *in vivo*, that were performed in our group on genomic DNA from mouse and rat eyecups subretinally injected as well as from mouse liver intravenously injected with D64V based IDLVs (Matrai et al., 2011; Yanez-Munoz et al., 2006). We used LAM-PCR strategies that could detect LTR- and non- LTR-mediated integration, which allowed scanning of 30–50% of the vector genome for host DNA–vector junctions. An extensive analysis showed a single integration event in a rat eyecup 2.5 months post injection, LTR-mediated and with canonical vector sequence. As our LAM-PCR search for integration events in these samples only revealed a single case in rat eyecups and 35 cases in mouse liver samples, we concluded that the high levels of IDLV 2-LTRs corresponded to episomal molecules, present at much higher levels than integrated vectors.

In conclusion, D64V based IDLVs have residual integration frequencies 2-4 log below those of their wild-type counterparts, at least in cell culture. Initial data sets from in vivo experiments are equally encouraging. Estimates of stable integration frequency of IDLVs in cultured cells (0.1–2.3%) are within the range described for plasmid transfection. Integrated proviruses from IDLVs are frequently noncanonical regarding terminal nucleotides, deletions in host cell DNA, and abnormal or missing DNA flanking repeats. IDLV integration pattern demonstrated random distribution on chromosomes, non-clustering around TSS sites, and less preferred integration into RefSeq genes as compared to LV integration pattern. Notably no proto-oncogene loci

as integration sites have been found in IDLV transduced cells. The rare but real integrations that do occur with IDLVs may also occur with other transient gene delivery methods, such as adenoviral vectors and nonviral plasmid-mediated gene delivery. Even this minor level of residual integration may be unacceptable in some clinical applications, like those involving mitosis-promoting genes, which may require non-DNA-based methods to prevent any genomic modification. However, the risk of insertional mutagenesis from IDLVs is highly reduced compared to integrating LVs, and from a biosafety standpoint IDLVs should be considered highly suitable at least for applications involving quiescent tissues.

6. References

Adolfsson, J., Borge, O.J., Bryder, D., Theilgaard-Monch, K., Astrand-Grundstrom, I., Sitnicka, E., Sasaki, Y., and Jacobsen, S.E. (2001). Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. Immunity *15*, 659-669.

Aguilo, F., Avagyan, S., Labar, A., Sevilla, A., Lee, D.F., Kumar, P., Lemischka, I.R., Zhou, B.Y., and Snoeck, H.W. (2011). Prdm16 is a physiologic regulator of hematopoietic stem cells. Blood *117*, 5057-5066.

Ailles, L., Schmidt, M., Santoni de Sio, F.R., Glimm, H., Cavalieri, S., Bruno, S., Piacibello, W., Von Kalle, C., and Naldini, L. (2002). Molecular evidence of lentiviral vector-mediated gene transfer into human self-renewing, multi-potent, long-term NOD/SCID repopulating hematopoietic cells. Mol Ther *6*, 615-626.

Aiuti, A., Cassani, B., Andolfi, G., Mirolo, M., Biasco, L., Recchia, A., Urbinati, F., Valacca, C., Scaramuzza, S., Aker, M.*, 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.

Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A., and Copeland, N.G. (2004). RTCGD: retroviral tagged cancer gene database. Nucleic Acids Res *32*, D523-527.

Akashi, K., Traver, D., Miyamoto, T., and Weissman, I.L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature *404*, 193-197.

Allsopp, R.C., Cheshier, S., and Weissman, I.L. (2001). Telomere shortening accompanies increased cell cycle activity during serial transplantation of hematopoietic stem cells. J Exp Med *193*, 917-924.

Allsopp, R.C., Morin, G.B., DePinho, R., Harley, C.B., and Weissman, I.L. (2003a). Telomerase is required to slow telomere shortening and extend replicative lifespan of HSCs during serial transplantation. Blood *102*, 517-520.

Allsopp, R.C., Morin, G.B., Horner, J.W., DePinho, R., Harley, C.B., and Weissman, I.L. (2003b). Effect of TERT over-expression on the long-term transplantation capacity of hematopoietic stem cells. Nat Med *9*, 369-371.

Amado, R.G., and Chen, I.S. (1999). Lentiviral vectors--the promise of gene therapy within reach? Science *285*, 674-676.

Anderson, W.F. (1998). Human gene therapy. Nature *392*, 25-30.

Antonchuk, J., Sauvageau, G., and Humphries, R.K. (2002). HOXB4-induced expansion of adult hematopoietic stem cells ex vivo. Cell *109*, 39-45.

Apolonia, L., Waddington, S.N., Fernandes, C., Ward, N.J., Bouma, G., Blundell, M.P., Thrasher, A.J., Collins, M.K., and Philpott, N.J. (2007). Stable gene transfer to muscle using non-integrating lentiviral vectors. Mol Ther *15*, 1947-1954.

Arai, F., Hirao, A., Ohmura, M., Sato, H., Matsuoka, S., Takubo, K., Ito, K., Koh, G.Y., and Suda, T. (2004). Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. Cell *118*, 149-161.

Babior, B.M. (2004). NADPH oxidase. Curr Opin Immunol *16*, 42-47.

Bach, F.H., Albertini, R.J., Joo, P., Anderson, J.L., and Bortin, M.M. (1968). Bone-marrow transplantation in a patient with the Wiskott-Aldrich syndrome. Lancet *2*, 1364-1366.

Barr, S.D., Ciuffi, A., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2006). HIV integration site selection: targeting in macrophages and the effects of different routes of viral entry. Mol Ther *14*, 218-225.

Bartholomae, C.C., Arens, A., Balaggan, K.S., Yanez-Munoz, R.J., Montini, E., Howe, S.J., Paruzynski, A., Korn, B., Appelt, J.U., MacNeil, A.*, et al.* (2011). Lentiviral Vector Integration Profiles Differ in Rodent Postmitotic Tissues. Mol Ther.

Baum, C., Dullmann, J., Li, Z., Fehse, B., Meyer, J., Williams, D.A., and von Kalle, C. (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. Blood *101*, 2099-2114.

Baum, C., Forster, P., Hegewisch-Becker, S., and Harbers, K. (1994). An optimized electroporation protocol applicable to a wide range of cell lines. Biotechniques *17*, 1058-1062.

Beard, B.C., Dickerson, D., Beebe, K., Gooch, C., Fletcher, J., Okbinoglu, T., Miller, D.G., Jacobs, M.A., Kaul, R., Kiem, H.P.*, et al.* (2007). Comparison of HIV-derived Lentiviral and MLV-based Gammaretroviral Vector Integration Sites in Primate Repopulating Cells. Mol Ther *15*, 1356-1365.

Becker, A.J., Mc, C.E., and Till, J.E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452-454.

Berry, C., Hannenhalli, S., Leipzig, J., and Bushman, F.D. (2006). Selection of target sites for mobile DNA integration in the human genome. PLoS computational biology *2*, e157.

Bhardwaj, G., Murdoch, B., Wu, D., Baker, D.P., Williams, K.P., Chadwick, K., Ling, L.E., Karanu, F.N., and Bhatia, M. (2001). Sonic hedgehog induces the proliferation of primitive human hematopoietic cells via BMP regulation. Nat Immunol *2*, 172-180.

Bhatia, M., Bonnet, D., Wu, D., Murdoch, B., Wrana, J., Gallacher, L., and Dick, J.E. (1999). Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells. J Exp Med *189*, 1139-1148.

Bloquel, C., Fabre, E., Bureau, M.F., and Scherman, D. (2004). Plasmid DNA electrotransfer for intracellular and secreted proteins expression: new methodological developments and applications. J Gene Med *6 Suppl 1*, S11-23.

Bosticardo, M., Marangoni, F., Aiuti, A., Villa, A., and Grazia Roncarolo, M. (2009). Recent advances in understanding the pathophysiology of Wiskott-Aldrich syndrome. Blood *113*, 6288-6295.

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

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. (2008). Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell stem cell *2*, 151-159.

Brenner, S., and Malech, H.L. (2003). Current developments in the design of onco-retrovirus and lentivirus vector systems for hematopoietic cell gene therapy. Biochim Biophys Acta *1640*, 1-24.

Brenner, S., Ryser, M.F., Choi, U., Whiting-Theobald, N., Kuhlisch, E., Linton, G., Kang, E., Lehmann, R., Rosen-Wolff, A., Rudikoff, A.G.*, et al.* (2006). Polyclonal long-term MFGS-gp91phox marking in rhesus macaques after nonmyeloablative transplantation with transduced autologous peripheral blood progenitor cells. Mol Ther *14*, 202-211.

Brussel, A., and Sonigo, P. (2003). Analysis of early human immunodeficiency virus type 1 DNA synthesis by use of a new sensitive assay for quantifying integrated provirus. J Virol *77*, 10119-10124.

Brussel, A., and Sonigo, P. (2004). Evidence for gene expression by unintegrated human immunodeficiency virus type 1 DNA species. J Virol *78*, 11263-11271.

Buonamici, S., Li, D., Chi, Y., Zhao, R., Wang, X., Brace, L., Ni, H., Saunthararajah, Y., and Nucifora, G. (2004). EVI1 induces myelodysplastic syndrome in mice. J Clin Invest *114*, 713-719.

Burns, J.C., Matsubara, T., Lozinski, G., Yee, J.K., Friedmann, T., Washabaugh, C.H., and Tsonis, P.A. (1994). Pantropic retroviral vector-mediated gene transfer, integration, and expression in cultured newt limb cells. Dev Biol *165*, 285-289.

Butler, S.L., Hansen, M.S., and Bushman, F.D. (2001). A quantitative assay for HIV DNA integration in vivo. Nat Med *7*, 631-634.

Calvi, L.M., Adams, G.B., Weibrecht, K.W., Weber, J.M., Olson, D.P., Knight, M.C., Martin, R.P., Schipani, E., Divieti, P., Bringhurst, F.R.*, et al.* (2003). Osteoblastic cells regulate the haematopoietic stem cell niche. Nature *425*, 841-846.

Capecchi, M.R. (1980). High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. Cell *22*, 479-488.

Cherry, S.R., Biniszkiewicz, D., van Parijs, L., Baltimore, D., and Jaenisch, R. (2000). Retroviral expression in embryonic stem cells and hematopoietic stem cells. Mol Cell Biol *20*, 7419-7426.

Christensen, J.L., and Weissman, I.L. (2001). Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. Proc Natl Acad Sci U S A *98*, 14541-14546.

Chuikov, S., Levi, B.P., Smith, M.L., and Morrison, S.J. (2010). Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. Nat Cell Biol *12*, 999-1006.

Ciuffi, A., Mitchell, R.S., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2006). Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. Mol Ther *13*, 366-373.

Coffin, J.M. (1997). Retroviruses (Plainview, NY: Cold Spring Harbor Laboratory Press).

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D.A. (2005). Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature *436*, 272-276.

Conneally, E., Eaves, C.J., and Humphries, R.K. (1998). Efficient retroviral-mediated gene transfer to human cord blood stem cells with in vivo repopulating potential. Blood *91*, 3487-3493.

Cooper, M.J., Lippa, M., Payne, J.M., Hatzivassiliou, G., Reifenberg, E., Fayazi, B., Perales, J.C., Morrison, L.J., Templeton, D., Piekarz, R.L.*, et al.* (1997). Safety-modified episomal vectors for human gene therapy. Proc Natl Acad Sci U S A *94*, 6450-6455.

Cornu, T.I., and Cathomen, T. (2007). Targeted genome modifications using integrase-deficient lentiviral vectors. Mol Ther *15*, 2107-2113.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

De Clercq, E. (2002). Strategies in the design of antiviral drugs. Nat Rev Drug Discov *1*, 13-25.

De Palma, M., Montini, E., Santoni de Sio, F.R., Benedicenti, F., Gentile, A., Medico, E., and Naldini, L. (2005). Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. Blood *105*, 2307-2315.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C.*, 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.

Deneault, E., Cellot, S., Faubert, A., Laverdure, J.P., Frechette, M., Chagraoui, J., Mayotte, N., Sauvageau, M., Ting, S.B., and Sauvageau, G. (2009). A functional screen to identify novel effectors of hematopoietic stem cell activity. Cell *137*, 369-379.

Derry, J.M., Ochs, H.D., and Francke, U. (1994). Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. Cell *78*, 635-644.

Devriendt, K., Kim, A.S., Mathijs, G., Frints, S.G., Schwartz, M., Van Den Oord, J.J., Verhoef, G.E., Boogaerts, M.A., Fryns, J.P., You, D.*, et al.* (2001). Constitutively activating mutation in WASP causes X-linked severe congenital neutropenia. Nat Genet *27*, 313-317.

Dewey, R.A., Avedillo Diez, I., Ballmaier, M., Filipovich, A., Greil, J., Gungor, T., Happel, C., Maschan, A., Noyan, F., Pannicke, U.*, et al.* (2006). Retroviral WASP gene transfer into human hematopoietic stem cells reconstitutes the actin cytoskeleton in myeloid progeny cells differentiated in vitro. Exp Hematol *34*, 1161-1169.

Dinauer, M.C., Li, L.L., Bjorgvinsdottir, H., Ding, C., and Pech, N. (1999). Long-term correction of phagocyte NADPH oxidase activity by retroviral-mediated gene transfer in murine X-linked chronic granulomatous disease. Blood *94*, 914-922.

Donehower, L.A., and Varmus, H.E. (1984). A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. Proc Natl Acad Sci U S A *81*, 6461-6465.

Du, Y., Jenkins, N.A., and Copeland, N.G. (2005). Insertional mutagenesis identifies genes that promote the immortalization of primary bone marrow progenitor cells. Blood *106*, 3932-3939.

Du, Z.W., Hu, B.Y., Ayala, M., Sauer, B., and Zhang, S.C. (2009). Cre recombination-mediated cassette exchange for building versatile transgenic human embryonic stem cells lines. Stem Cells *27*, 1032-1041.

Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K.J., and Engelhardt, J.F. (1998). Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J Virol *72*, 8568-8577.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J Virol *72*, 8463-8471.

Dykstra, B., Ramunas, J., Kent, D., McCaffrey, L., Szumsky, E., Kelly, L., Farn, K., Blaylock, A., Eaves, C., and Jervis, E. (2006). High-resolution video monitoring of hematopoietic stem cells cultured in single-cell arrays identifies new features of self-renewal. Proc Natl Acad Sci U S A *103*, 8185-8190.

Ehrhardt, A., Haase, R., Schepers, A., Deutsch, M.J., Lipps, H.J., and Baiker, A. (2008). Episomal vectors for gene therapy. Curr Gene Ther *8*, 147-161.

Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. Hum Gene Ther *16*, 1241-1246.

Emmons, R.V., Doren, S., Zujewski, J., Cottler-Fox, M., Carter, C.S., Hines, K., O'Shaughnessy, J.A., Leitman, S.F., Greenblatt, J.J., Cowan, K.*, et al.* (1997). Retroviral gene transduction of adult peripheral blood or marrow-derived CD34+ cells for six hours without growth factors or on autologous stroma does not improve marking efficiency assessed in vivo. Blood *89*, 4040-4046.

Engelman, A., Englund, G., Orenstein, J.M., Martin, M.A., and Craigie, R. (1995). Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. J Virol *69*, 2729-2736.

ESGT (2006). One of three successfully treated CGD patients in a Swiss-German gene therapy trial died due to his underlying disease: A position statement from the European Society of Gene Therapy (ESGT). J Gene Med *8*, 1435.

Evans-Galea, M.V., Wielgosz, M.M., Hanawa, H., Srivastava, D.K., and Nienhuis, A.W. (2007). Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. Mol Ther *15*, 801-809.

Fears, S., Mathieu, C., Zeleznik-Le, N., Huang, S., Rowley, J.D., and Nucifora, G. (1996). Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia

and produces a new member of the PR domain family. Proc Natl Acad Sci U S A *93*, 1642-1647.
Felgner, P.L., Gadek, T.R., Holm, M., Roman, R., Chan, H.W., Wenz, M., Northrop, J.P., Ringold, G.M., and Danielsen, M. (1987). Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proc Natl Acad Sci U S A *84*, 7413-7417.
Felice, B., Cattoglio, C., Cittaro, D., Testa, A., Miccio, A., Ferrari, G., Luzi, L., Recchia, A., and Mavilio, F. (2009). Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS One *4*.
Ford, C.E., Hamerton, J.L., Barnes, D.W., and Loutit, J.F. (1956). Cytological identification of radiation-chimaeras. Nature *177*, 452-454.
Gabriel, R., Eckenberg, R., Paruzynski, A., Bartholomae, C.C., Nowrouzi, A., Arens, A., Howe, S.J., Recchia, A., Cattoglio, C., Wang, W.*, et al.* (2009). Comprehensive genomic access to vector integration in clinical gene therapy. Nat Med *15*, 1431-1436.
Gatti, R.A., Meuwissen, H.J., Allen, H.D., Hong, R., and Good, R.A. (1968). Immunological reconstitution of sex-linked lymphopenic immunological deficiency. Lancet *2*, 1366-1369.
Gaur, M., and Leavitt, A.D. (1998). Mutations in the human immunodeficiency virus type 1 integrase D,D(35)E motif do not eliminate provirus formation. J Virol *72*, 4678-4685.
Ghazizadeh, S., Carroll, J.M., and Taichman, L.B. (1997). Repression of retrovirus-mediated transgene expression by interferons: implications for gene therapy. J Virol *71*, 9163-9169.
Glimm, H., and Eaves, C.J. (1999). Direct evidence for multiple self-renewal divisions of human in vivo repopulating hematopoietic cells in short-term culture. Blood *94*, 2161-2168.
Glimm, H., Eisterer, W., Lee, K., Cashman, J., Holyoake, T.L., Nicolini, F., Shultz, L.D., von Kalle, C., and Eaves, C.J. (2001). Previously undetected human hematopoietic cell populations with short-term repopulating activity selectively engraft NOD/SCID-beta2 microglobulin-null mice. J Clin Invest *107*, 199-206.
Glimm, H., Flugge, K., Mobest, D., Hofmann, V.M., Postmus, J., Henschler, R., Lange, W., Finke, J., Kiem, H.P., Schulz, G.*, et al.* (1998). Efficient serum-free retroviral gene transfer into primitive human hematopoietic progenitor cells by a defined, high-titer, nonconcentrated vector-containing medium. Hum Gene Ther *9*, 771-778.
Glimm, H., Kiem, H.P., Darovsky, B., Storb, R., Wolf, J., Diehl, V., Mertelsmann, R., and Von Kalle, C. (1997). Efficient gene transfer in primitive CD34+/CD38lo human bone marrow cells reselected after long-term exposure to GALV-pseudotyped retroviral vector. Hum Gene Ther *8*, 2079-2086.
Glimm, H., Tang, P., Clark-Lewis, I., von Kalle, C., and Eaves, C. (2002). Ex vivo treatment of proliferating human cord blood stem cells with stroma-derived factor-1 enhances their ability to engraft NOD/SCID mice. Blood *99*, 3454-3457.
Goodell, M.A., Brose, K., Paradis, G., Conner, A.S., and Mulligan, R.C. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J Exp Med *183*, 1797-1806.
Goyama, S., Yamamoto, G., Shimabe, M., Sato, T., Ichikawa, M., Ogawa, S., Chiba, S., and Kurokawa, M. (2008). Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. Cell stem cell *3*, 207-220.
Graham, F.L., and van der Eb, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology *52*, 456-467.
Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K.*, et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest *118*, 3132-3142.
Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M.*, et al.* (2003a). A serious adverse

event after successful gene therapy for X-linked severe combined immunodeficiency. N Engl J Med *348*, 255-256.
Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E.*, et al.* (2003b). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415-419.
Hagino-Yamagishi, K., Donehower, L.A., and Varmus, H.E. (1987). Retroviral DNA integrated during infection by an integration-deficient mutant of murine leukemia virus is oligomeric. J Virol *61*, 1964-1971.
Heuer, T.S., and Brown, P.O. (1997). Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. Biochemistry *36*, 10655-10665.
Holyoake, T.L., Nicolini, F.E., and Eaves, C.J. (1999). Functional differences between transplantable human hematopoietic stem cells from fetal liver, cord blood, and adult marrow. Exp Hematol *27*, 1418-1427.
Horwitz, M.E., Barrett, A.J., Brown, M.R., Carter, C.S., Childs, R., Gallin, J.I., Holland, S.M., Linton, G.F., Miller, J.A., Leitman, S.F.*, et al.* (2001). Treatment of chronic granulomatous disease with nonmyeloablative conditioning and a T-cell-depleted hematopoietic allograft. N Engl J Med *344*, 881-888.
Howe, S.J., Mansour, M.R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., Brugman, M.H., Pike-Overzet, K., Chatters, S.J., de Ridder, D.*, 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.
Imai, K., Morio, T., Zhu, Y., Jin, Y., Itoh, S., Kajiwara, M., Yata, J., Mizutani, S., Ochs, H.D., and Nonoyama, S. (2004). Clinical course of patients with WASP gene mutations. Blood *103*, 456-464.
Iscove, N.N., and Nawa, K. (1997). Hematopoietic stem cells expand during serial transplantation in vivo without apparent exhaustion. Curr Biol *7*, 805-808.
Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell *91*, 501-510.
Iwama, A., Oguro, H., Negishi, M., Kato, Y., Morita, Y., Tsukui, H., Ema, H., Kamijo, T., Katoh-Fukui, Y., Koseki, H.*, et al.* (2004). Enhanced self-renewal of hematopoietic stem cells mediated by the polycomb gene product Bmi-1. Immunity *21*, 843-851.
Jacobson, L.O., Simmons, E.L., and Bethard, W.F. (1950). Studies on hematopoietic recovery from radiation injury. J Clin Invest *29*, 825.
Jacobson, L.O., Simmons, E.L., Marks, E.K., and Eldredge, J.H. (1951). Recovery from radiation injury. Science *113*, 510-511.
Kay, M.A., Glorioso, J.C., and Naldini, L. (2001). Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. Nat Med *7*, 33-40.
Kay, M.A., and High, K. (1999). Gene therapy for the hemophilias. Proc Natl Acad Sci U S A *96*, 9973-9975.
Kern, J.A., Wakita, R., and Sliwkowski, M.X. (1999). Neuregulin receptor-mediated gene transfer by human epidermal growth factor receptor 2-targeted antibodies and neuregulin-1. Cancer Gene Ther *6*, 537-545.
Kim, Y.J., Kim, Y.S., Larochelle, A., Renaud, G., Wolfsberg, T.G., Adler, R., Donahue, R.E., Hematti, P., Hong, B.K., Roayaei, J.*, et al.* (2009). Sustained high-level polyclonal hematopoietic marking and transgene expression 4 years after autologous transplantation of rhesus macaques with SIV lentiviral vector-transduced CD34+ cells. Blood *113*, 5434-5443.
Klein, C., Nguyen, D., Liu, C.H., Mizoguchi, A., Bhan, A.K., Miki, H., Takenawa, T., Rosen, F.S., Alt, F.W., Mulligan, R.C.*, et al.* (2003). Gene therapy for Wiskott-Aldrich syndrome: rescue of

T-cell signaling and amelioration of colitis upon transplantation of retrovirally transduced hematopoietic stem cells in mice. Blood *101*, 2159-2166.
Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661-672.
Kren, B.T., Unger, G.M., Sjeklocha, L., Trossen, A.A., Korman, V., Diethelm-Okita, B.M., Reding, M.T., and Steer, C.J. (2009). Nanocapsule-delivered Sleeping Beauty mediates therapeutic Factor VIII expression in liver sinusoidal endothelial cells of hemophilia A mice. J Clin Invest *119*, 2086-2099.
Kustikova, O., Fehse, B., Modlich, U., Yang, M., Dullmann, J., Kamino, K., von Neuhoff, N., Schlegelberger, B., Li, Z., and Baum, C. (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. Science *308*, 1171-1174.
Laufs, S., Guenechea, G., Gonzalez-Murillo, A., Zsuzsanna Nagy, K., Luz Lozano, M., del Val, C., Jonnakuty, S., Hotz-Wagenblatt, A., Jens Zeller, W., Bueren, J.A.*, et al.* (2006). Lentiviral vector integration sites in human NOD/SCID repopulating cells. J Gene Med *8*, 1197-1207.
Lawrence, H.J., Christensen, J., Fong, S., Hu, Y.L., Weissman, I., Sauvageau, G., Humphries, R.K., and Largman, C. (2005). Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. Blood *106*, 3988-3994.
Leavitt, A.D., Robles, G., Alesandro, N., and Varmus, H.E. (1996). Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. J Virol *70*, 721-728.
Lewinski, M.K., Bisgrove, D., Shinn, P., Chen, H., Hoffmann, C., Hannenhalli, S., Verdin, E., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2005). Genome-wide analysis of chromosomal features repressing human immunodeficiency virus transcription. J Virol *79*, 6610-6619.
Li, Z., Dullmann, J., Schiedlmeier, B., Schmidt, M., von Kalle, C., Meyer, J., Forster, M., Stocking, C., Wahlers, A., Frank, O.*, et al.* (2002). Murine leukemia induced by retroviral gene marking. Science *296*, 497.
Liu, Y., Chen, L., Ko, T.C., Fields, A.P., and Thompson, E.A. (2006). Evi1 is a survival factor which conveys resistance to both TGFbeta- and taxol-mediated cell death via PI3K/AKT. Oncogene *25*, 3565-3575.
Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science *295*, 868-872.
Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D.*, et al.* (2007). Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol *25*, 1298-1306.
Lorenz, E., Uphoff, D., Reid, T.R., and Shelton, E. (1951). Modification of irradiation injury in mice and guinea pigs by bone marrow injections. J Natl Cancer Inst *12*, 197-201.
Magli, M.C., Iscove, N.N., and Odartchenko, N. (1982). Transient nature of early haematopoietic spleen colonies. Nature *295*, 527-529.
Malech, H.L., Maples, P.B., Whiting-Theobald, N., Linton, G.F., Sekhsaria, S., Vowells, S.J., Li, F., Miller, J.A., DeCarlo, E., Holland, S.M.*, et al.* (1997). Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease. Proc Natl Acad Sci U S A *94*, 12133-12138.
Mardiney, M., 3rd, Jackson, S.H., Spratt, S.K., Li, F., Holland, S.M., and Malech, H.L. (1997). Enhanced host defense after gene transfer in the murine p47phox-deficient model of chronic granulomatous disease. Blood *89*, 2268-2275.
Maruggi, G., Porcellini, S., Facchini, G., Perna, S.K., Cattoglio, C., Sartori, D., Ambrosi, A., Schambach, A., Baum, C., Bonini, C.*, et al.* (2009). Transcriptional enhancers induce

insertional gene deregulation independently from the vector type and design. Mol Ther *17*, 851-856.
Masuda, T., Kuroda, M.J., and Harada, S. (1998). Specific and independent recognition of U3 and U5 att sites by human immunodeficiency virus type 1 integrase in vivo. J Virol *72*, 8396-8402.
Matrai, J., Cantore, A., Bartholomae, C.C., Annoni, A., Wang, W., Acosta-Sanchez, A., Samara-Kuko, E., De Waele, L., Ma, L., Genovese, P.*, et al.* (2011). Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. Hepatology *53*, 1696-1707.
Matrai, J., Chuah, M.K., and VandenDriessche, T. (2010). Recent advances in lentiviral vector development and applications. Mol Ther *18*, 477-490.
May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K.M., Luzzatto, L., and Sadelain, M. (2000). Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. Nature *406*, 82-86.
Metais, J.Y., and Dunbar, C.E. (2008). The MDS1-EVI1 gene complex as a retrovirus integration site: impact on behavior of hematopoietic cells and implications for gene therapy. Mol Ther *16*, 439-449.
Miao, C.H., Snyder, R.O., Schowalter, D.B., Patijn, G.A., Donahue, B., Winther, B., and Kay, M.A. (1998). The kinetics of rAAV integration in the liver. Nat Genet *19*, 13-15.
Miller, M.D., Farnet, C.M., and Bushman, F.D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol *71*, 5382-5390.
Miller, M.D., Wang, B., and Bushman, F.D. (1995). Human immunodeficiency virus type 1 preintegration complexes containing discontinuous plus strands are competent to integrate in vitro. J Virol *69*, 3938-3944.
Miller, N., and Whelan, J. (1997). Progress in transcriptionally targeted and regulatable vectors for genetic therapy. Hum Gene Ther *8*, 803-815.
Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol *2*, E234.
Miyoshi, H., Blomer, U., Takahashi, M., Gage, F.H., and Verma, I.M. (1998). Development of a self-inactivating lentivirus vector. J Virol *72*, 8150-8157.
Mochizuki, N., Shimizu, S., Nagasawa, T., Tanaka, H., Taniwaki, M., Yokota, J., and Morishita, K. (2000). A novel gene, MEL1, mapped to 1p36.3 is highly homologous to the MDS1/EVI1 gene and is transcriptionally activated in t(1;3)(p36;q21)-positive leukemia cells. Blood *96*, 3209-3214.
Modlich, U., Bohne, J., Schmidt, M., von Kalle, C., Knoss, S., Schambach, A., and Baum, C. (2006). Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. Blood *108*, 2545-2553.
Modlich, U., Navarro, S., Zychlinski, D., Maetzig, T., Knoess, S., Brugman, M.H., Schambach, A., Charrier, S., Galy, A., Thrasher, A.J.*, et al.* (2009). Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. Mol Ther *17*, 1919-1928.
Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M.*, et al.* (2009). 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 *119*, 964-975.
Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi Sergi, L., Benedicenti, F., Ambrosi, A., Di Serio, C.*, et al.* (2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. Nat Biotechnol *24*, 687-696.

Morishita, K., Parganas, E., Bartholomew, C., Sacchi, N., Valentine, M.B., Raimondi, S.C., Le Beau, M.M., and Ihle, J.N. (1990). The human Evi-1 gene is located on chromosome 3q24-q28 but is not rearranged in three cases of acute nonlymphocytic leukemias containing t(3;5)(q25;q34) translocations. Oncogene Res *5*, 221-231.

Morishita, K., Parker, D.S., Mucenski, M.L., Jenkins, N.A., Copeland, N.G., and Ihle, J.N. (1988). Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. Cell *54*, 831-840.

Moritz, T., Dutt, P., Xiao, X., Carstanjen, D., Vik, T., Hanenberg, H., and Williams, D.A. (1996). Fibronectin improves transduction of reconstituting hematopoietic stem cells by retroviral vectors: evidence of direct viral binding to chymotryptic carboxy-terminal fragments. Blood *88*, 855-862.

Morrison, S.J., Wandycz, A.M., Hemmati, H.D., Wright, D.E., and Weissman, I.L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. Development *124*, 1929-1939.

Morrison, S.J., and Weissman, I.L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. Immunity *1*, 661-673.

Mucenski, M.L., Taylor, B.A., Ihle, J.N., Hartley, J.W., Morse, H.C., 3rd, Jenkins, N.A., and Copeland, N.G. (1988). Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. Mol Cell Biol *8*, 301-308.

Na Nakorn, T., Traver, D., Weissman, I.L., and Akashi, K. (2002). Myeloerythroid-restricted progenitors are sufficient to confer radioprotection and provide the majority of day 8 CFU-S. J Clin Invest *109*, 1579-1585.

Nakai, H., Iwaki, Y., Kay, M.A., and Couto, L.B. (1999). Isolation of recombinant adeno-associated virus vector-cellular DNA junctions from mouse liver. J Virol *73*, 5438-5447.

Naldini, L., Blomer, U., Gage, F.H., Trono, D., and Verma, I.M. (1996a). Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. Proc Natl Acad Sci U S A *93*, 11382-11388.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996b). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263-267.

Nienhuis, A.W., Dunbar, C.E., and Sorrentino, B.P. (2006). Genotoxicity of retroviral integration in hematopoietic cells. Mol Ther *13*, 1031-1049.

Nightingale, S.J., Hollis, R.P., Pepper, K.A., Petersen, D., Yu, X.J., Yang, C., Bahner, I., and Kohn, D.B. (2006). Transient gene expression by nonintegrating lentiviral vectors. Mol Ther *13*, 1121-1132.

Nijnik, A., Woodbine, L., Marchetti, C., Dawson, S., Lambe, T., Liu, C., Rodrigues, N.P., Crockford, T.L., Cabuy, E., Vindigni, A*., et al.* (2007). DNA repair is limiting for haematopoietic stem cells during ageing. Nature *447*, 686-690.

Nishikata, I., Sasaki, H., Iga, M., Tateno, Y., Imayoshi, S., Asou, N., Nakamura, T., and Morishita, K. (2003). A novel EVI1 gene family, MEL1, lacking a PR domain (MEL1S) is expressed mainly in t(1;3)(p36;q21)-positive AML and blocks G-CSF-induced myeloid differentiation. Blood *102*, 3323-3332.

Nishikawa, M., Tahara, T., Hinohara, A., Miyajima, A., Nakahata, T., and Shimosaka, A. (2001). Role of the microenvironment of the embryonic aorta-gonad-mesonephros region in hematopoiesis. Ann N Y Acad Sci *938*, 109-116.

Nowell, P.C., Cole, L.J., Habermeyer, J.G., and Roan, P.L. (1956). Growth and continued function of rat marrow cells in x-radiated mice. Cancer Res *16*, 258-261.

Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Miura, Y., and Suda, T. (1992). In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. Blood *80*, 3044-3050.

Orkin, S.H. (1995). Hematopoiesis: how does it happen? Curr Opin Cell Biol *7*, 870-877.
Orkin, S.H. (1996). Development of the hematopoietic system. Curr Opin Genet Dev *6*, 597-602.
Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kuhlcke, K., Schilz, A., Kunkel, H.*, 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.
Panet, A., and Cedar, H. (1977). Selective degradation of integrated murine leukemia proviral DNA by deoxyribonucleases. Cell *11*, 933-940.
Panganiban, A.T., and Temin, H.M. (1984). The retrovirus pol gene encodes a product required for DNA integration: identification of a retrovirus int locus. Proc Natl Acad Sci U S A *81*, 7885-7889.
Paruzynski, A., Arens, A., Gabriel, R., Bartholomae, C.C., Scholz, S., Wang, W., Wolf, S., Glimm, H., Schmidt, M., and von Kalle, C. (2010). Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. Nat Protoc *5*, 1379-1395.
Passegue, E., Wagner, E.F., and Weissman, I.L. (2004). JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. Cell *119*, 431-443.
Patel, P.H., and Preston, B.D. (1994). Marked infidelity of human immunodeficiency virus type 1 reverse transcriptase at RNA and DNA template ends. Proc Natl Acad Sci U S A *91*, 549-553.
Perucho, M., Hanahan, D., and Wigler, M. (1980). Genetic and physical linkage of exogenous sequences in transformed cells. Cell *22*, 309-317.
Pfeifer, A., Brandon, E.P., Kootstra, N., Gage, F.H., and Verma, I.M. (2001). Delivery of the Cre recombinase by a self-deleting lentiviral vector: efficient gene targeting in vivo. Proc Natl Acad Sci U S A *98*, 11450-11455.
Philpott, N.J., and Thrasher, A.J. (2007). Use of nonintegrating lentiviral vectors for gene therapy. Hum Gene Ther *18*, 483-489.
Potter, H., Weir, L., and Leder, P. (1984). Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. Proc Natl Acad Sci U S A *81*, 7161-7165.
Pryciak, P.M., Sil, A., and Varmus, H.E. (1992). Retroviral integration into minichromosomes in vitro. EMBO J *11*, 291-303.
Pryciak, P.M., and Varmus, H.E. (1992). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell *69*, 769-780.
Reya, T. (2003). Regulation of hematopoietic stem cell self-renewal. Recent Prog Horm Res *58*, 283-295.
Reya, T., Duncan, A.W., Ailles, L., Domen, J., Scherer, D.C., Willert, K., Hintz, L., Nusse, R., and Weissman, I.L. (2003). A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature *423*, 409-414.
Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J Virol *61*, 336-343.
Rollins, S.A., Birks, C.W., Setter, E., Squinto, S.P., and Rother, R.P. (1996). Retroviral vector producer cell killing in human serum is mediated by natural antibody and complement: strategies for evading the humoral immune response. Hum Gene Ther *7*, 619-626.
Rossi, D.J., Bryder, D., Seita, J., Nussenzweig, A., Hoeijmakers, J., and Weissman, I.L. (2007). Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature *447*, 725-729.

Roth, J.A., Nguyen, D., Lawrence, D.D., Kemp, B.L., Carrasco, C.H., Ferson, D.Z., Hong, W.K., Komaki, R., Lee, J.J., Nesbitt, J.C*., et al.* (1996). Retrovirus-mediated wild-type p53 gene transfer to tumors of patients with lung cancer. Nat Med *2*, 985-991.

Rother, R.P., Fodor, W.L., Springhorn, J.P., Birks, C.W., Setter, E., Sandrin, M.S., Squinto, S.P., and Rollins, S.A. (1995). A novel mechanism of retrovirus inactivation in human serum mediated by anti-alpha-galactosyl natural antibody. J Exp Med *182*, 1345-1355.

Ryser, M.F., Roesler, J., Gentsch, M., and Brenner, S. (2007). Gene therapy for chronic granulomatous disease. Expert Opin Biol Ther *7*, 1799-1809.

Sadat, M.A., Pech, N., Saulnier, S., Leroy, B.A., Hossle, J.P., Grez, M., and Dinauer, M.C. (2003). 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 *14*, 651-666.

Sakai, I., Tamura, T., Narumi, H., Uchida, N., Yakushijin, Y., Hato, T., Fujita, S., and Yasukawa, M. (2005). Novel RUNX1-PRDM16 fusion transcripts in a patient with acute myeloid leukemia showing t(1;21)(p36;q22). Genes, chromosomes & cancer *44*, 265-270.

Sauvageau, G., Iscove, N.N., and Humphries, R.K. (2004). In vitro and in vivo expansion of hematopoietic stem cells. Oncogene *23*, 7223-7232.

Sauvageau, G., Thorsteinsdottir, U., Eaves, C.J., Lawrence, H.J., Largman, C., Lansdorp, P.M., and Humphries, R.K. (1995). Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations in vitro and in vivo. Genes Dev *9*, 1753-1765.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A *103*, 1412-1417.

Scherdin, U., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J Virol *64*, 907-912.

Scherer, F., Anton, M., Schillinger, U., Henke, J., Bergemann, C., Kruger, A., Gansbacher, B., and Plank, C. (2002). Magnetofection: enhancing and targeting gene delivery by magnetic force in vitro and in vivo. Gene Ther *9*, 102-109.

Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H., and von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods *4*, 1051-1057.

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

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

Schwarzwaelder, K., Howe, S.J., Schmidt, M., Brugman, M.H., Deichmann, A., Glimm, H., Schmidt, S., Prinz, C., Wissler, M., King, D.J*., 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.

Segal, A.W. (2005). How neutrophils kill microbes. Annu Rev Immunol *23*, 197-223.

Seger, R.A., Gungor, T., Belohradsky, B.H., Blanche, S., Bordigoni, P., Di Bartolomeo, P., Flood, T., Landais, P., Muller, S., Ozsahin, H*., et al.* (2002). Treatment of chronic granulomatous disease with myeloablative conditioning and an unmodified hemopoietic allograft: a survey of the European experience, 1985-2000. Blood *100*, 4344-4350.

Shackelford, D., Kenific, C., Blusztajn, A., Waxman, S., and Ren, R. (2006). Targeted degradation of the AML1/MDS1/EVI1 oncoprotein by arsenic trioxide. Cancer Res *66*, 11360-11369.

Shih, C.C., Stoye, J.P., and Coffin, J.M. (1988). Highly preferred targets for retrovirus integration. Cell *53*, 531-537.

Shimizu, S., Suzukawa, K., Kodera, T., Nagasawa, T., Abe, T., Taniwaki, M., Yagasaki, F., Tanaka, H., Fujisawa, S., Johansson, B.*, et al.* (2000). Identification of breakpoint cluster regions at 1p36.3 and 3q21 in hematologic malignancies with t(1;3)(p36;q21). Genes, chromosomes & cancer *27*, 229-238.
Shing, D.C., Trubia, M., Marchesi, F., Radaelli, E., Belloni, E., Tapinassi, C., Scanziani, E., Mecucci, C., Crescenzi, B., Lahortiga, I.*, et al.* (2007). Overexpression of sPRDM16 coupled with loss of p53 induces myeloid leukemias in mice. J Clin Invest *117*, 3696-3707.
Siminovitch, L., McCulloch, E.A., and Till, J.E. (1963). The Distribution of Colony-Forming Cells among Spleen Colonies. J Cell Physiol *62*, 327-336.
Smith, L.G., Weissman, I.L., and Heimfeld, S. (1991). Clonal analysis of hematopoietic stem-cell differentiation in vivo. Proc Natl Acad Sci U S A *88*, 2788-2792.
Sorrentino, B.P. (2004). Clinical strategies for expansion of haematopoietic stem cells. Nat Rev Immunol *4*, 878-888.
Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58-62.
Starr, T.K., Allaei, R., Silverstein, K.A., Staggs, R.A., Sarver, A.L., Bergemann, T.L., Gupta, M., O'Sullivan, M.G., Matise, I., Dupuy, A.J.*, et al.* (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science *323*, 1747-1750.
Stein, S., Ott, M.G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Kramer, A., Schwable, J., Glimm, H.*, et al.* (2010). Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat Med *16*, 198-204.
Stieger, K., Belbellaa, B., Le Guiner, C., Moullier, P., and Rolling, F. (2009). In vivo gene regulation using tetracycline-regulatable systems. Adv Drug Deliv Rev *61*, 527-541.
Taniyama, Y., Tachibana, K., Hiraoka, K., Aoki, M., Yamamoto, S., Matsumoto, K., Nakamura, T., Ogihara, T., Kaneda, Y., and Morishita, R. (2002). Development of safe and efficient novel nonviral gene transfer using ultrasound: enhancement of transfection efficiency of naked plasmid DNA in skeletal muscle. Gene Ther *9*, 372-380.
Thorsteinsdottir, U., Mamo, A., Kroon, E., Jerome, L., Bijl, J., Lawrence, H.J., Humphries, K., and Sauvageau, G. (2002). Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. Blood *99*, 121-129.
Thrasher, A.J. (2002). WASp in immune-system organization and function. Nat Rev Immunol *2*, 635-646.
Thrasher, A.J., and Burns, S.O. (2010). WASP: a key immunological multitasker. Nat Rev Immunol *10*, 182-192.
Thrasher, A.J., and Kinnon, C. (2000). The Wiskott-Aldrich syndrome. Clin Exp Immunol *120*, 2-9.
Till, J.E., and Mc, C.E. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat Res *14*, 213-222.
Vargas, J., Jr., Gusella, G.L., Najfeld, V., Klotman, M.E., and Cara, A. (2004). Novel integrase-defective lentiviral episomal vectors for gene transfer. Hum Gene Ther *15*, 361-372.
Varnum-Finney, B., Brashem-Stein, C., and Bernstein, I.D. (2003). Combined effects of Notch signaling and cytokines induce a multiple log increase in precursors with lymphoid and myeloid reconstituting ability. Blood *101*, 1784-1789.
Varnum-Finney, B., Xu, L., Brashem-Stein, C., Nourigat, C., Flowers, D., Bakkour, S., Pear, W.S., and Bernstein, I.D. (2000). Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. Nat Med *6*, 1278-1281.
Vijaya, S., Steffen, D.L., and Robinson, H.L. (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J Virol *60*, 683-692.

Vink, C.A., Gaspar, H.B., Gabriel, R., Schmidt, M., McIvor, R.S., Thrasher, A.J., and Qasim, W. (2009). Sleeping beauty transposition from nonintegrating lentivirus. Mol Ther *17*, 1197-1204.

Visser, J.W., and Van Bekkum, D.W. (1990). Purification of pluripotent hemopoietic stem cells: past and present. Exp Hematol *18*, 248-256.

von Kalle, C., Glimm, H., Schulz, G., Mertelsmann, R., and Henschler, R. (1998). New developments in hematopoietic stem cell expansion. Curr Opin Hematol *5*, 79-86.

von Kalle, C., Kiem, H.P., Goehle, S., Darovsky, B., Heimfeld, S., Torok-Storb, B., Storb, R., and Schuening, F.G. (1994). Increased gene transfer into human hematopoietic progenitor cells by extended in vitro exposure to a pseudotyped retroviral vector. Blood *84*, 2890-2897.

Wagers, A.J., Sherwood, R.I., Christensen, J.L., and Weissman, I.L. (2002). Little evidence for developmental plasticity of adult hematopoietic stem cells. Science *297*, 2256-2259.

Wagner, E., Zenke, M., Cotten, M., Beug, H., and Birnstiel, M.L. (1990). Transferrin-polycation conjugates as carriers for DNA uptake into cells. Proc Natl Acad Sci U S A *87*, 3410-3414.

Wang, G.P., Levine, B.L., Binder, G.K., Berry, C.C., Malani, N., McGarrity, G., Tebas, P., June, C.H., and Bushman, F.D. (2009). Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol Ther *17*, 844-850.

Wanisch, K., and Yanez-Munoz, R.J. (2009). Integration-deficient lentiviral vectors: a slow coming of age. Mol Ther *17*, 1316-1332.

Weil, W.M., Linton, G.F., Whiting-Theobald, N., Vowells, S.J., Rafferty, S.P., Li, F., and Malech, H.L. (1997). Genetic correction of p67phox deficient chronic granulomatous disease using peripheral blood progenitor cells as a target for retrovirus mediated gene transfer. Blood *89*, 1754-1761.

Weissman, I.L. (2000). Stem cells: units of development, units of regeneration, and units in evolution. Cell *100*, 157-168.

Wieser, R. (2007). The oncogene and developmental regulator EVI1: expression, biochemical properties, and biological functions. Gene *396*, 346-357.

Williams, D.A. (1999). Retroviral-fibronectin interactions in transduction of mammalian cells. Ann N Y Acad Sci *872*, 109-113; discussion 113-104.

Wilson, A., Murphy, M.J., Oskarsson, T., Kaloulis, K., Bettess, M.D., Oser, G.M., Pasche, A.C., Knabenhans, C., Macdonald, H.R., and Trumpp, A. (2004). c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. Genes Dev *18*, 2747-2763.

Wimmer, K., Vinatzer, U., Zwirn, P., Fonatsch, C., and Wieser, R. (1998). Comparative expression analysis of the antagonistic transcription factors EVI1 and MDS1-EVI1 in murine tissues and during in vitro hematopoietic differentiation. Biochem Biophys Res Commun *252*, 691-696.

Wu, A.M., Till, J.E., Siminovitch, L., and McCulloch, E.A. (1968). Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system. J Exp Med *127*, 455-464.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749-1751.

Xu, L.C., Kluepfel-Stahl, S., Blanco, M., Schiffmann, R., Dunbar, C., and Karlsson, S. (1995). Growth factors and stromal support generate very efficient retroviral transduction of peripheral blood CD34+ cells from Gaucher patients. Blood *86*, 141-146.

Yamane, T., Kunisada, T., Tsukamoto, H., Yamazaki, H., Niwa, H., Takada, S., and Hayashi, S.I. (2001). Wnt signaling regulates hemopoiesis through stromal cells. J Immunol *167*, 765-772.

Yanez-Munoz, R.J., Balaggan, K.S., MacNeil, A., Howe, S.J., Schmidt, M., Smith, A.J., Buch, P., MacLaren, R.E., Anderson, P.N., Barker, S.E.*, et al.* (2006). Effective gene therapy with nonintegrating lentiviral vectors. Nat Med *12*, 348-353.

Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Mansson, R., Sigvardsson, M., and Jacobsen, S.E. (2005). Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. Blood *105*, 2717-2723.

Yang, N.S., Burkholder, J., Roberts, B., Martinell, B., and McCabe, D. (1990). In vivo and in vitro gene transfer to mammalian somatic cells by particle bombardment. Proc Natl Acad Sci U S A *87*, 9568-9572.

Yuasa, H., Oike, Y., Iwama, A., Nishikata, I., Sugiyama, D., Perkins, A., Mucenski, M.L., Suda, T., and Morishita, K. (2005). Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression. EMBO J *24*, 1976-1987.

Zavidij, O., Ball, C.R., Herbst, F., Fessler, S., Schmidt, M., von Kalle, C., and Glimm, H. (2010). Hematopoietic activity of human short term repopulating cells in mobilized peripheral blood cell transplants is restricted to the first 5 months after transplantation. Blood.

Zeira, E., Manevitch, A., Khatchatouriants, A., Pappo, O., Hyam, E., Darash-Yahana, M., Tavor, E., Honigman, A., Lewis, A., and Galun, E. (2003). Femtosecond infrared laser-an efficient and safe in vivo gene delivery system for prolonged expression. Mol Ther *8*, 342-350.

Zhang, J., Niu, C., Ye, L., Huang, H., He, X., Tong, W.G., Ross, J., Haug, J., Johnson, T., Feng, J.Q*., et al.* (2003). Identification of the haematopoietic stem cell niche and control of the niche size. Nature *425*, 836-841.

Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L., and Trono, D. (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. J Virol *72*, 9873-9880.

7. Appendix

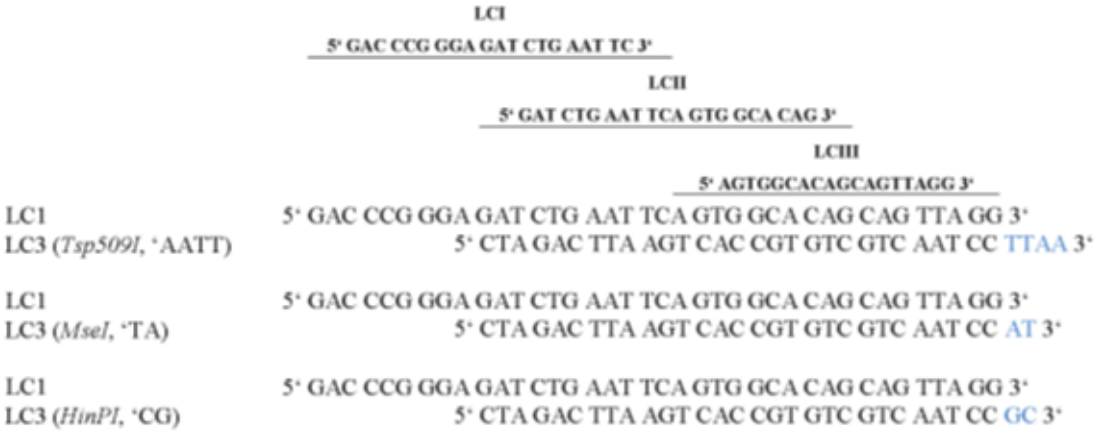

Figure 7-1 Linker cassette design for the LAM-PCR. For the production of linker cassettes oligonucleotides (LC1 + one LC3) were chosen, which form overhangs after hybridisation. The overhang from the LC3 oligonucleotide had to be complementary to the overhang produced by the used restriction enzyme in the LAM-PCR. Hybridisation of the linker cassette oligonucleotides LC1 and LC3 is shown for the enzymes *Tsp509I*, *MseI* and *HpyCH4IV* which generate an AATT, TA and GC overhang, respectively. Locations of the linker cassette primers LCI and LCII are indicated. These primers were used to amplify the vector-genome junctions. The LCIII primer sequence is the linker-specific part of the fusionprimer B-LK (454 pyrosequencing). LC, linker cassette.

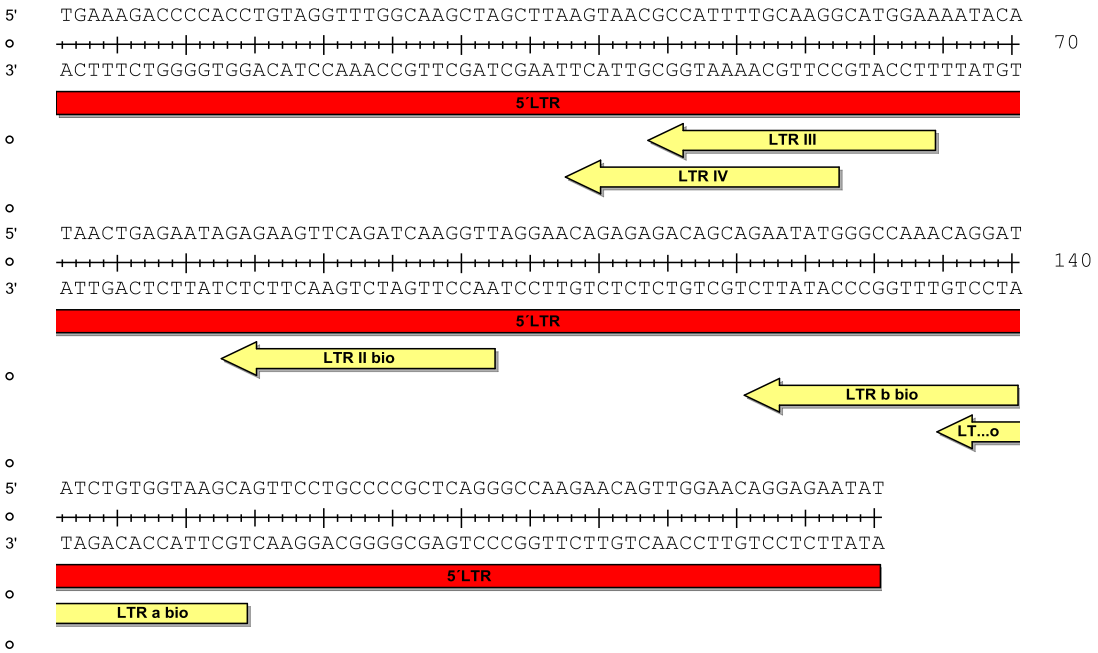

Figure 7-2 Location of 5' LAM-PCR primers for MLV based gamma-retroviral vectors. LTR, long terminal repeat; Bio, biotinylated

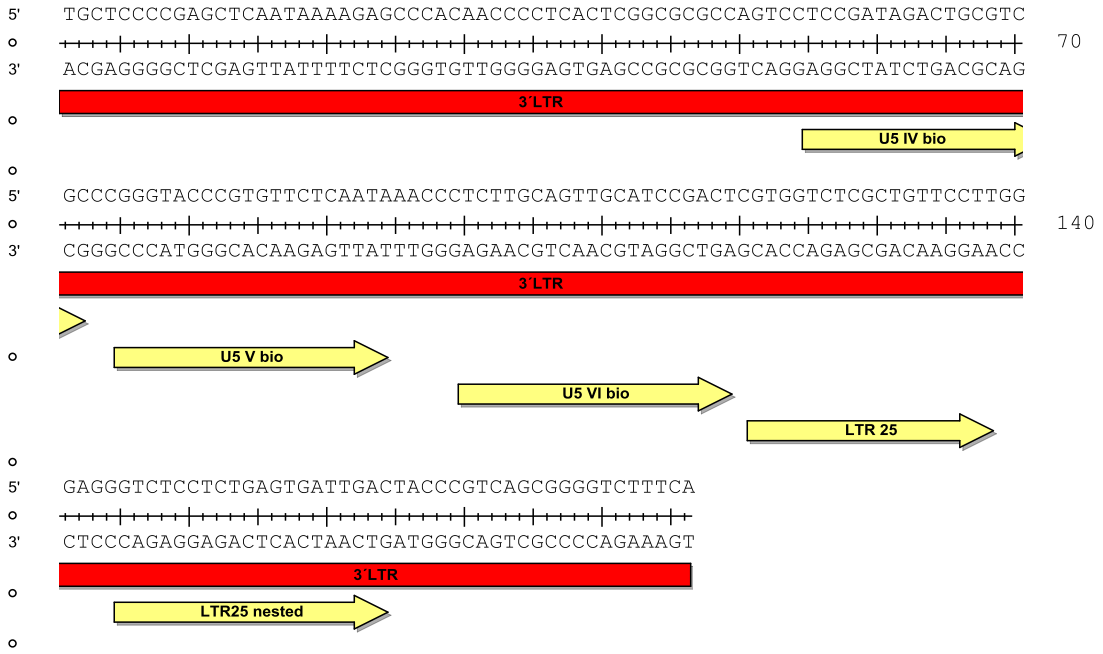

Figure 7-3 Location of 3' LAM-PCR primers for MLV based gamma-retroviral vectors. LTR, long terminal repeat; Bio, biotinylated

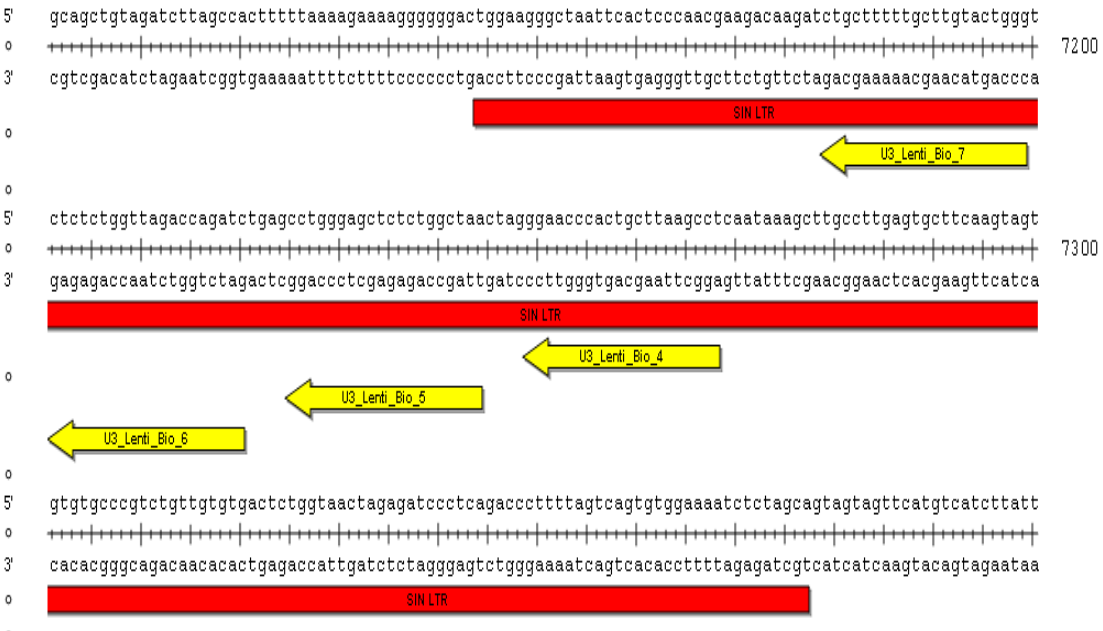

Figure 7-4 Locations of 5' LAM-PCR primers for the SIN lentiviral vectors. . LTR, long terminal repeat; Bio, biotinylated

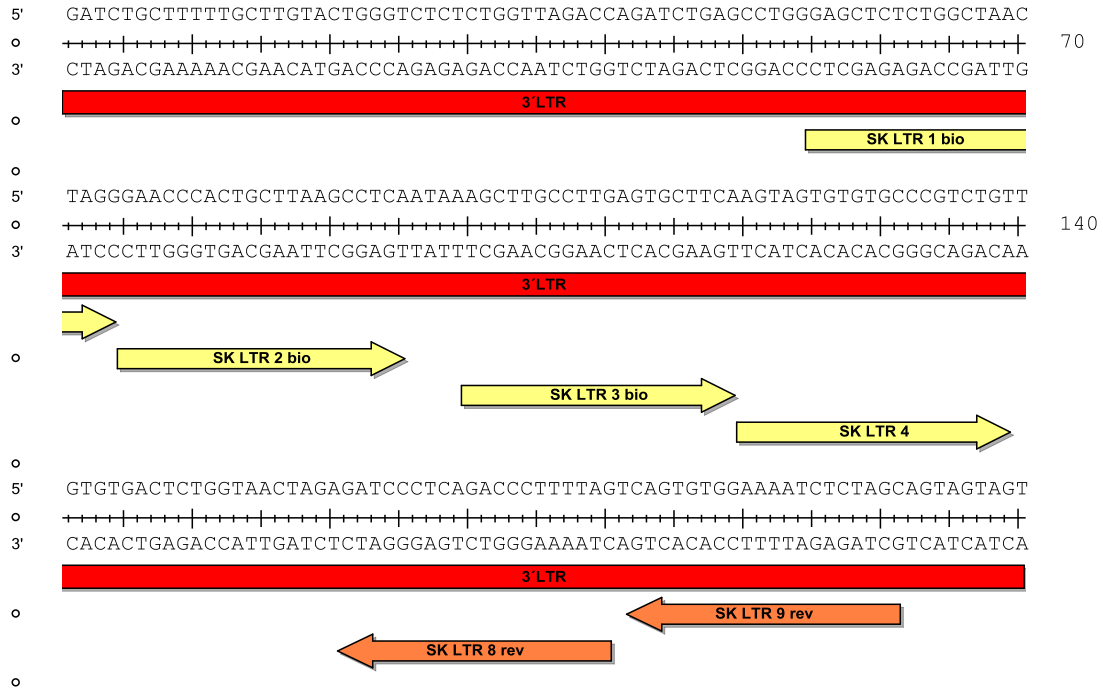

Figure 7-5 Locations of 3' LAM-PCR primers for the SIN lentiviral vectors. . LTR, long terminal repeat; Bio, biotinylated

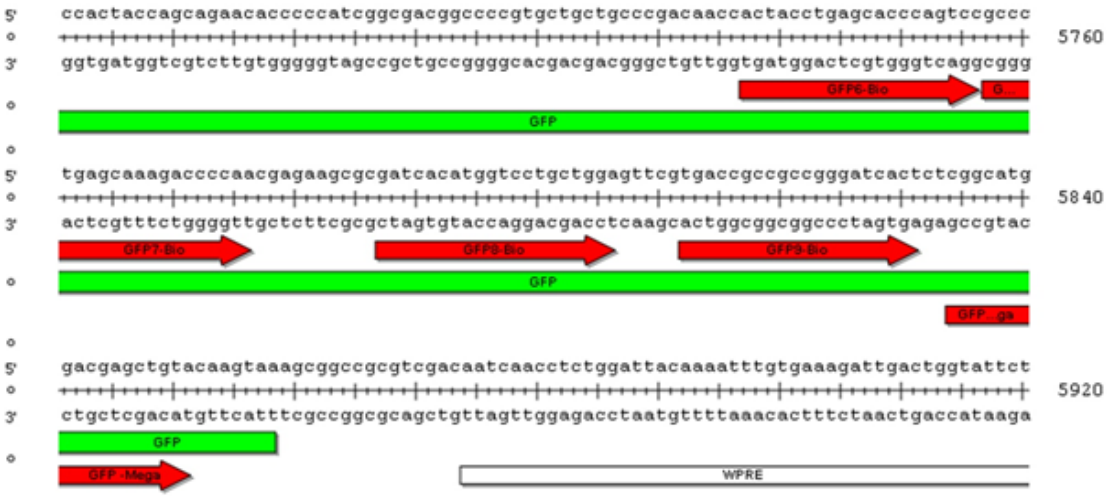

Figure 7-6 Locations of the eGFP-LAM-PCR primer in the LV/IDLV vector backbone. Bio, biotinylated