

Development and Safety Assessment of Lentiviral Vector Gene Therapy for SCID-X1

Ontwikkeling en veiligheidsevaluatie van lentivirale
gentherapie voor SCID-X1



ISBN: 978-90-8891-708-0

Cover: “Knowledge Tree” by Marshall Huston

Layout: Egied Simmons

Printing: Publisher BOXPress || Proefschriftmaken.nl

Copyright © M.W. Huston, Rotterdam 2013

All rights reserved. No part of this thesis may be reproduced or transmitted in any form or by any means without written permission of the copyright owner.

Acknowledgements

Funding for the studies detailed in this thesis was provided by program grants 43100016 and 43400010 of the Translational Gene Therapy Research Program of The Netherlands Health Research Organization ZonMW (www.zonmw.nl) and by large-scale collaborative projects of the Theme Health of the European Commission’s 6th and 7th Framework programs (<http://ec.europa.eu/research>) CONSERT (<http://www.consert.eu/>), PERSIST (<http://www.persist-project.eu/>) and CELL-PID (<http://www.cell-pid.eu/>).

Development and Safety Assessment of Lentiviral Vector Gene Therapy for SCID-X1

Ontwikkeling en veiligheidsevaluatie van lentivirale
gentherapie voor SCID-X1

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. H.G. Schmidt

and in accordance with the decision of the Doctorate Board

The public defense shall be held on
Tuesday 22 October 2013 at 13:30 hours
by

Marshall William Huston

born in Happy Camp, California, USA

Doctoral Committee

Promotor:

prof.dr. G. Wagemaker

Other members:

prof.dr. P.J. van der Spek

prof.dr. H.B. Gaspar

prof.dr. J.J. Cornelissen

Co-promotor:

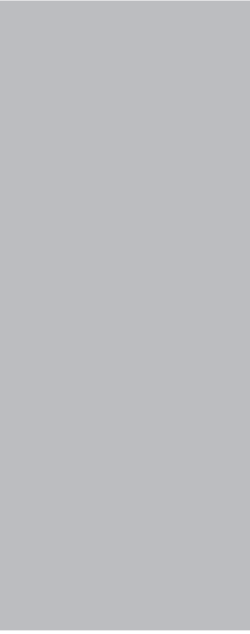
Dr. N.P. van Til

For my mother and father.

CONTENTS

Chapter 1:	General introduction	9
Chapter 2:	Correction of murine SCID-X1 by lentiviral gene therapy using a codon optimized IL2RG gene and minimal pretransplant conditioning.	35
Chapter 3:	Pre-transplant G-CSF mobilization improves lentiviral vector gene therapy in a SCID-X1 mouse model.	67
Chapter 4:	Low leukemia incidence in <i>Il2rg</i> ^{-/-} mice treated with lentiviral vector gene therapy.	87
Chapter 5:	Comprehensive investigation of the parameters of viral integration site analysis and their effects on the gene annotations produced.	109
Chapter 6:	Lentiviral integration profiles are stable across multiple species, phenotypes and tissue types.	141
Chapter 7:	General discussion	171
	Summary	183
	Samenvatting	187
	List of terms and abbreviations	191
	PhD portfolio	193
	List of publications	195
	Curriculum vitae	197
	Acknowledgements	199

Met Tyr Xle Ala Cys
 Met Xle Ala Cys
 Cys Xle Ala Cys
 Ala Lys Ser Cys
 Xle Ser Pyl Lys
 Arg Glu Asn Ser
 PylAsn Ser Thr Pyl
 Lys Leu SerThr LeuGlu Asn TrpIle LeuLeu Pyl His
 Cys Glu Ser GGAC GCC GGCATGCC TrpIle LeuLeu His Sec
 Ala Xle Thr AGCAGGTACGCGAGA His Ala Asn Ser
 Xle Thr AGGCTAGCTAG PylThr
 GluArg CAGGCTCGA Arg His
 CAGTGGGTAT PylAsn Arg AlaArg ArgIle
 GTGAGCTTGCTC Thr Ser Ile Asn Pyl Glu
 GATAGCGGGTIA Sec Thr Pyl Thr His
 TGCTAGCCCGCC His Thr His
 GGGAAATGCAGG SerSec
 TACGCTTGCTAGA
 CCGCAGGCATGC
 CTGTATGCTTGCTC
 01001101011000010111
 00100 1110011011010000
 01100 001011011000110
 1100 0010000001010111
 011 010010110110001101
 10 0011010010110000101
 101101001 00000010010
 00011101 01011100110111010
 0011011 11011011100010110000
 100000 01010
 00001 1010
 000 100100



General introduction

I. HEMATOPOIESIS

The average adult human body contains approximately 5 liters of blood, corresponding to roughly 8% of the body's total weight[1]. It is composed primarily of plasma (54%) and erythrocytes, commonly called red blood cells (45%). The remaining 1% is made up of leukocytes, often referred to as white blood cells.

90% of plasma is water, but the fluid also contains various enzymes, ions and proteins as well as diffused carbon dioxide gas. Glucose, mineral ions and hormones are distributed throughout the body via the blood plasma, while waste products carbon dioxide and dissolved proteins are deposited in the plasma for transport to disposal sites. Plasma also contains clotting factors to prevent excessive blood loss from damaged blood vessels.

The most important contributors to blood clotting are the platelets, which are produced by megakaryocytes. Megakaryocytes are large cells that remain in the bone marrow rather than circulating. They extend tendrils into the blood sinuses that release platelets into the blood via a budding process.

Erythrocytes, also known as red blood cells, are the most common cell type in the human body, making up 25% of the total cell population. 2-3 million are produced every second and released into circulation[1]. Erythrocytes are responsible for transporting oxygen to every tissue in the body. To this end, they contain an iron-rich molecule in their cytoplasm known as hemoglobin. In addition to binding oxygen, hemoglobin gives erythrocytes their red color. Erythrocytes lack a nucleus and most organelles, and only circulate for 100 to 120 days before breaking down.

Leukocytes, or white blood cells, make up a tiny fraction of the total blood volume, but are essential to the body's defense against pathological agents. Leukocytes are divided into several lineages based on their function and their appearance under a light microscope: granulocytes, monocytes, and lymphocytes.

Granulocytes are named for the numerous lysosomes and secretory vesicles (granules) they contain. There are three main sub-types of granulocytes. Neutrophils, the most common sub-type, are responsible for phagocytosis of microorganisms such as bacteria and play a key role in innate immunity. Basophils secrete histamine to help mediate inflammatory reactions, while eosinophils destroy parasites and modulate allergic inflammatory responses.

Monocytes mature into macrophages upon leaving the bloodstream and migrating into tissue. Macrophages and neutrophils are the primary phagocytes of the immune system, engulfing microorganisms and dead or damaged cells and depositing them in lysosomes to be digested. Monocytes can also differentiate into dendritic cells, which have some phagocytic activity but are more active at presenting foreign antigens to lymphocytes to trigger a specific immune response.

Lymphocytes, the primary cells of the adaptive immune system, are split into two main classes: B cells and T cells. Lymphocytes are able to recognize a broad range of antigens, create a specific immune response to the offending pathogen, and retain a memory of the triggering antigens so that further incursions by that pathogen will be met with a rapid

response. The development of monoclonal antibodies recognizing cell surface markers, as well as fluorescence activated cell sorting (FACS) to visualize these antibodies, has allowed lymphocytes to be categorized into specific fractions based on their role in adaptive immunity. The two main types of T cells are CD4⁺ ‘helper’ T cells and CD8⁺ cytotoxic T cells. CD4⁺ T cells, once activated by antigen presenting cells (APCs), rapidly divide and release cytokines to regulate a proper immune response. A subset of CD4⁺ T cells (CD4⁺CD25⁺FoxP3⁺ T cells) plays a crucial role in immune tolerance by suppressing T-cell mediated immunity and suppressing auto-reactive T cells. CD8⁺ T cells, sometimes called “killer” T cells, are activated by antigens bound to the MHC class I protein on the surface of distressed cells. CD8⁺ T cells destroy virus-infected cells and some tumor cells. Both types of T cells are positive for the CD3 marker.

B lymphocytes are positive for antibodies B220 and CD19, and express specialized antigen receptors, called immunoglobulins (Ig), with a single antigenic specificity. Immunoglobulins are often referred to as antibodies. Upon activation, B-cells mature into plasma cells and release large amounts of immunoglobulins into the bloodstream. These secreted forms of Ig are called antibodies, and are a key component of the humoral immune system. Immature B cells express IgM on their surface, while mature B cells express both IgM and IgD. Once a B cell is activated, it begins to secrete antibodies and may undergo isotype switching, allowing it to produce IgA, IgE or IgG.

A small fraction of lymphocytes, called natural killer (NK) cells, are involved in innate immunity and play a role in destroying virus-infected and tumor cells. Upon activation, NK cells release granules containing cytotoxic proteins to destroy the targeted cell. NK cells are positive for surface markers CD16 and CD56 in humans and NK1.1/NK1.2 in mice.

II. Hematopoietic stem cells and stem cell transplantation

The component cells of blood are all derived from the same source: the hematopoietic stem cell (HSC). As with other stem cells, HSC are capable of self-renewal, as well as differentiation into myeloid and lymphoid progenitor cells. These progenitor cells can in turn differentiate into mature blood cells of specific lineages[2], as shown in Figure 1. HSC are rare cells (estimated to be 1 in 100,000 hematopoietic cells) that reside in the bone marrow in specific niches, mainly near osteoblasts and adjacent to blood vessels. They can also be found in umbilical cord blood (UCB). HSC still cannot be identified based on morphological properties. Instead, they are recognized by their affinity with certain monoclonal antibodies. The primary marker used for human HSC is CD34, a transmembrane protein expressed on hematopoietic progenitor cells in peripheral blood, bone marrow and vascular endothelium[3]. Additional cell surface antigens are used to distinguish between immature HSC and committed lineage precursors. For instance, CD34⁺CD38⁻ cells are devoid of lineage markers and contain long-term repopulating stem cells, while CD34⁺CD38⁺ cells are only capable of short-term hematopoiesis[4]. In mice, lineage marker negative, Sca-1⁺ and c-Kit⁺ (LSK) cells are considered to be highly enriched for HSC. To compensate for the relatively short lifespan of many blood cell types, HSC produce vast numbers of new cells every day. HSC are capable of remarkable expansion in times of need: it is possible for a single HSC to repopulate an entire hematopoietic system[5]ZB.

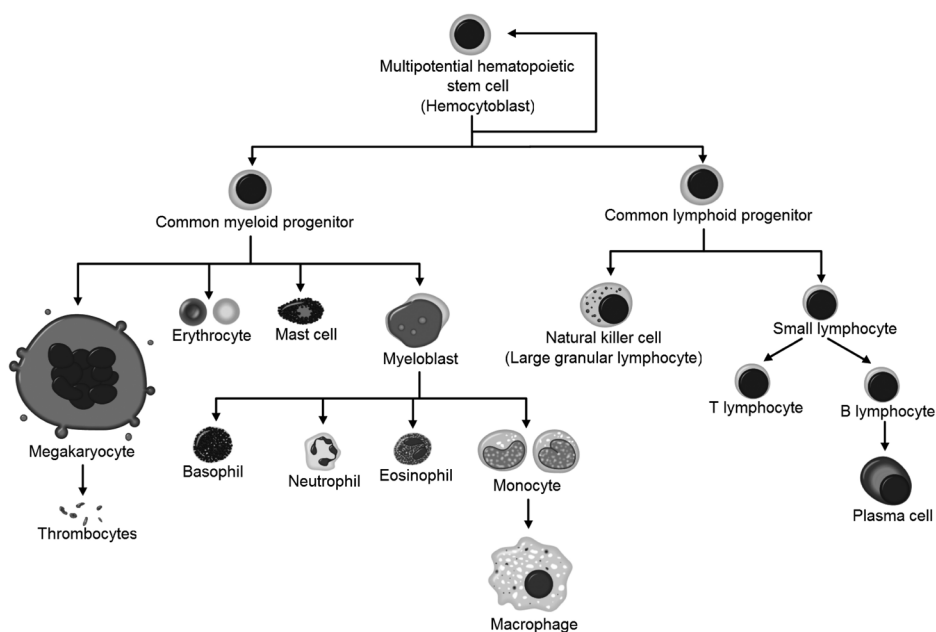


Figure 1: Hematopoiesis. Image © A. Rad / Wikimedia Commons / CC-BY-SA-3.0 / GFDL

Hematopoietic stem cell transplantation as therapy

Allogeneic bone marrow transplantation (BMT) has been used to treat a range of hematopoietic diseases such as leukemia, autoimmune disease, radiation poisoning and immunodeficiency diseases via introduction of healthy hematopoietic stem cells. BMT is also applied in some cancer patients to reduce mortality after intense chemotherapy, and can improve the survival rates in these patients by 20% or more[6].

A successful BMT results in the creation of a new, fully functional hematopoietic system in the recipient[7]. The procedure involves mobilization of HSC from the BM of an HLA-matched donor into the peripheral blood via granulate colony-stimulating factor (G-CSF) followed by purification of CD34⁺ cells. These cells are injected into the patient's bloodstream where they will migrate to the host's bone marrow and engraft.

It is critical that the donor and recipient have identical alleles of the major histocompatibility locus, called human leukocyte antigen (HLA), in order to prevent graft rejection and/or graft vs host disease (GVHD)[8]. BMT between HLA-matched donors and recipients has a very high survival rate (greater than 90%)[9]. Transplantation between non-HLA identical donors and recipients is possible, but carries a sharply increased risk of GVHD[10] and higher mortality (~40%). Pre-transplant depletion of T cells reduces the risk of GVHD in non-matched transplants. Novel molecular diagnostics technologies, such as next generation sequencing and microarray platforms, are applied to compare the histocompatibility loci and ensure the virus-free status of the to-be-transplanted bone marrow.

III. SCID

Human severe combined immunodeficiency (SCID) is a class of primary inherited immunodeficiencies caused by a mutation in one of the genes critical for the development of hematopoietic stem cells into mature T cell lineages. It was first reported 60 years ago in Switzerland, where doctors noticed that some infants were extremely lymphopenic (lacking in mature lymphocytes) and died within 1 or 2 years of birth[11]. Further research revealed that there was more than one cause of the SCID phenotype: in some families SCID was passed on via an X-linked recessive inheritance, while in other families SCID was linked to an autosomal recessive inheritance. It was this autosomal recessive form of SCID that was first linked to a molecular cause, adenosine deaminase deficiency, in 1972[12]. Two decades later the second, X-chromosome linked, cause of SCID was identified[13, 14]. Today SCID has been linked to mutations in 12 different genes, with more likely to be identified in the future[15-20].

Due to its recessive nature and extreme lethality, SCID is very rare in the general population. The prevalence of SCID varies depending on the genetic makeup of the population. The overall prevalence of SCID in the United States is estimated at 1 per 500,000 live births, while in Australia it has been estimated at 1 in 65,000[21]. The worldwide average is thought to be around 1 out of every 50,000 to 100,000 live births.

There are several forms of SCID, some of which affect B and NK cells in addition to T cells. The exact phenotype depends on which gene is defective[22]. Patients suffering from SCID develop symptoms within a few months after birth, including persistent diarrhea, recurring opportunistic infections and failure to thrive[10, 22, 23]. Causative organisms include *Candida albicans*, varicella, adenovirus, respiratory syncytial virus, cytomegalovirus, Epstein–Barr virus, and parainfluenza virus. Transplacentally acquired maternal cells may persist in these patients and cause a rash due to GVHD on the skin[24], or occasionally the liver[25]. As protective maternal antibodies wane after the first few months of life, affected children experience an increased frequency of infection. A fully developed SCID leads to overwhelming fungal, viral and bacterial infections and a failure to thrive, with death occurring in the first year of life unless treatment is obtained. The various types of SCID are discussed below.

T^B-NK⁻ SCIDs

The adenosine deaminase (*ADA*) gene was the first gene to be linked to SCID[12], and is responsible for approximately 15-20% of all SCID cases, about a third of all autosomal recessive SCIDs[26, 27]. ADA is an enzyme of the purine salvage pathway that catalyzes deamination of deoxyadenosine and adenosine. Patients with ADA deficiency have a marked reduction in thymopoiesis and therefore have few circulating T lymphocytes. Furthermore, those T lymphocytes that are produced have decreased *in vivo* survival time[28]. Besides the severe consequences for lymphocyte development, many other symptoms have been reported, such as deafness, neuro-developmental deficits and skeletal abnormalities[29-31]. In mice, ADA deficiency results in perinatal mortality[32, 33], which can

be prevented when expression of the ADA protein is restored in trophoblast cells[34]. In these rescued mice, severe lymphopenia and combined immunodeficiency that emulates the human ADA-SCID phenotype can be observed.

Reticular dysgenesis (RD) is a very rare disease caused by a defect in the adenylate kinase 2 gene (AK2)[35, 36]. Patients with RD, which are estimated to make up 2-4% of all SCID patients, lack B, T and NK cells. RD is characterized by complete failure of development of both myeloid and lymphoid cells. Hypogammaglobulinemia, lymphopenia, absent cell-mediated immunity, and neutropenia lead to severe and often fatal infections in the first year of life[37].

TB·NK⁺ SCIDs

Defects in the recombination activating genes *RAG1* or *RAG2* results in a TB·NK⁺ SCID phenotype[38, 39]. RAG proteins create the double strand DNA breaks that initiate V(D)J recombination, the process of immunoglobulin and T-cell receptor (TCR) gene rearrangement which lead to mature B and T cells[40, 41]. Mutations in the *RAG* genes results in severely reduced recombinase activity. The inability to rearrange Ig and TCR gene segments results in a developmental block in early stages of B and T cell development. NK cells, which do not undergo any recombination processes during maturation, are unaffected. *Rag1* and *Rag2* deficient mice have the same phenotype as human RAG-SCID patients[42, 43]. Some RAG-SCID patients retain a small population of autologous T cells, which can lead to the autoimmune disorder known as Omenn syndrome.

Omenn syndrome (OS) is a form of SCID that results from mutations one of several genes involved in V(D)J recombination, including *Artemis* and *IL7RA* but most often *RAG1* or *RAG2*[44]. OS patients suffer from autoimmune response leading to generalized erythrodermia, hepatosplenomegaly, lymphadenopathy and alopecia, as well as protracted diarrhoea and increased occurrence of life threatening infections[45]. In OS patients, B cells are mostly absent, while T cell counts are highly variable. T cells are active but oligoclonal, with a restricted TCR repertoire. Omenn syndrome patients often suffer from a graft-vs-host-like disease due to T lymphocyte infiltrates in the skin, gut, liver or spleen[39, 46, 47].

Other patients with T-B-NK⁺ SCID have normal *RAG* genes but were found to be sensitive to ionizing radiation, suggesting a defect in DNA double-strand break repair[48, 49]. This defect was later linked to the *Artemis* gene[17, 50, 51]. *Artemis* is involved in non-homologous end joining and functions in hairpin opening during V(D)J-recombination. The immunological phenotype of *Artemis* KO mice closely resembles that of *Artemis*-SCID patients, except that low levels of CD4⁺ T cells can be found in some mice[52].

A rare type of radiosensitive TB·NK⁺ SCID caused by DNA ligase IV (*LIG4*)-deficiency was recently indentified[53]. Ligase IV plays a role in non-homologous end joining during repair of double-strand breaks[54]. Mutations in the DNA ligase IV gene (*LIG4*) lead to a block in T- and B-cell development. DNA ligase IV-deficiency in mice causes late embryonic lethality, suggesting that *LIG4* has a key role in murine embryogenesis not shared by humans[55-57].

TB⁺NK⁺ SCIDs

Defects in the interleukin-7 receptor alpha (IL-7R α) chain result in a TB⁺NK⁺ form of SCID, and are responsible for approximately 10% of SCID cases[18, 58, 59]. IL-7R α deficiency is an autosomal recessive form of SCID. Deficiency in this receptor results in defective IL-7 signalling, resulting in a wide range of T cell abnormalities. IL-7R mice have a different phenotype than IL-7R α -SCID patients, as in the murine model B as well as T cell lineages are dependent on IL-7 signalling[60].

Deficiency of the hematopoietic-specific transmembrane protein tyrosine phosphatase CD45 also results in a TB⁺NK⁺ SCID phenotype[16, 61]. CD45 regulates the Src kinases necessary for efficient T and B cell antigen receptor signal transduction[62-64]. Patients with mutations in the *CD45* gene do have a small number of T cells, but they are not activated and do not respond to proliferation signals. As in other human SCIDs of this type, B cells are present but dysfunctional.

Isolated cases of SCID-causing mutations in genes coding for CD3 subunits have also been described[15, 65-67]. CD3 chains form a complex with the T-cell receptor heterodimer and play a crucial role in TCR signaling[68, 69]. All CD3 chains carry immunoreceptor tyrosine-based activation motifs (ITAMs)[70, 71]. Upon ligation of the TCR, ITAMs become phosphorylated and enable signalling events downstream of the TCR. Besides playing an important role in the function of mature T lymphocytes, CD3 subunits also play a crucial role during T-cell development. Each of the CD3 δ , CD3 ϵ and CD3 ζ subunits is required for assembly, surface expression and function of the TCR and their deficiency results in a block in T-development. Both CD3 δ and CD3 ϵ appear to be essential for intrathymic development of T cells whereas CD3 γ does not[72]. Some CD3 γ deficiency patients have a complete SCID phenotype while other individuals present with a milder immunodeficiency[73, 74]. Some, but not all, targeted *Cd3* gene disruptions in mice differ from human phenotype. The developmental block in *Cd3 γ* ^{-/-} mice is stronger than in that seen in humans [75], whereas the block in *Cd3 δ* ^{-/-} mice is weaker than human CD3 δ -deficiency[76, 77].

TB⁺NK⁻ SCIDs

The most common form of SCID, affecting 50-60% of patients, is X-linked SCID (SCID-X1). SCID-X1 is caused by mutation in the interleukin-2 receptor gamma gene (*IL2RG*) located on chromosome Xq13[13]. *IL2RG* produces the common γ (gamma) chain protein, an integral membrane protein required for receptor binding and internalization. The common γ chain (γ c) protein is a component of multiple interleukin receptor complexes including IL-2, IL-4, IL-7, IL-9, IL-15 and IL-21[78-82]. γ c-mediated activation of these complexes results in stimulation of the JAK-STAT pathway, a downstream signaling pathway composed of the Janus kinases (JAKs) and the signal transducers and activators of transcription (STATs), which act upon target genes[83, 84]. Without proper IL-7 signaling, the JAK-STAT pathway is disrupted and immature T cell precursors are unable to develop into fully functional T cell lineages[18]. Similarly, lack of IL-15 signal-

ing is thought to disrupt NK cell development[85]. B cells are present in SCID-X1 patients but dysfunctional due to the absence of T cell co-activation as well as loss of IL-4 and IL-21 signaling. Thus, despite having a normal or even elevated number of B cells, SCID-X1 patients have low serum immunoglobulin (Ig) levels and lack all necessary Ig isotypes. *Il2rg*^{-/-} mice are T^B NK⁻ and so have a different phenotype than the human disease. The lack of B cells in *Il2rg*^{-/-} mice is due to the importance of IL-7 signaling to B cell development in the murine, but not human, immune system[86].

In one interesting case, an X-SCID patient had a spontaneous reversion of a γ c mutation in a single T-cell precursor. This single corrected precursor was able to produce a diverse T-cell repertoire and partially ameliorate the X-SCID phenotype[87, 88].

Genetic defects in the Janus kinase 3 (*JAK3*) gene can cause an autosomal recessive form of SCID with a T^B NK⁻ phenotype[19, 89, 90]. JAK3, a tyrosine kinase that forms part of the aforementioned Janus family of kinases, associates with the γ c protein and plays a critical role in cytokine signaling[91, 92]. Upon ligand binding, cytokine receptor-associated Janus kinases are activated by transphosphorylation. Stat proteins bind the phosphorylated receptor chains and in turn are phosphorylated by JAK3. Phosphorylated Stat proteins translocate to the nucleus and accumulate, regulating gene expression. Because JAK3 and the common γ chain are directly associated, the phenotype of JAK3-deficient patients is identical to that of X-linked SCID patients and mice with a JAK3-deficiency mirror the T^B NK⁻ phenotype of *Il2rg*^{-/-} mice[93, 94].

Treatment of SCID

The preferred treatment for SCID is hematopoietic stem cell transplantation from a healthy donor. In a successful transplantation, the stem cells migrate to the patient's bone marrow and thymus and engraft. Because wild-type T cell progenitors have a selective advantage over cells containing debilitating mutations, in a successful transplant the SCID patient's immune system quickly reconstitutes, with the newfound ability to generate mature T, B and/or NK cell lineages. If diagnosis of SCID is made at birth or shortly after, HLA-identical or haploidentical bone marrow stem cell transplantation is highly successful in all forms of SCID[95]. Later diagnosis increases the risk of severe infections and reduces the effectiveness of bone marrow transplantation. The first bone marrow stem cell transplantation was successfully performed on a SCID-X1 patient in 1968 using a related, HLA-identical donor [96], and has since achieved a very high success rate of nearly 95% for patients with an HLA-matched donor[97-99]. Unfortunately, perfectly matched donors are available for less than a third of SCID patients[100, 101]. Transplantations from mismatched donors have improved in recent years, but still yield a significantly lower chance of survival (between 50% and 80%)[102]. BM transplantation from a non-haploidentical donor leads to an increased risk of graft vs host disease [103] and a poorer survival prognosis. Other potential complications of BMT for SCID include failure of engraftment, Epstein-Barr virus-associated lymphoproliferative disease, and a failure to reconstitute normal humoral immunity after successful functional T-cell engraftment[104]. Even

upon successful engraftment, some recipients of allogeneic bone marrow transplantation experience incomplete B or NK cell reconstitution due to the engraftment failure of donor progenitor cells of those types[105]. Lack of normal humoral immune response after successful T cell engraftment has been seen even in HLA-matched BMT and suggests that, unlike T cells, healthy donor B cells do not enjoy a strong selective advantage compared to the recipient's cells. Use of a conditioning regimen significantly increases the frequency of sustained engraftment and results in more frequent engraftment of donor B lymphocytes and myeloid cells[102]. Donor B-cell chimerism is strongly associated with the development of normal B-cell function. A fully myeloablative regimen can lead to full B-cell immunity development but is associated with a higher risk of toxicity, particularly in infected patients. For this reason, pre-transplant conditioning is not often used in SCID-X1 patients.

Prior to the advent of gene therapy, the only alternative to BMT was immunoglobulin therapy. Ig therapy can extend the lifespan of SCID patients, but it is expensive and non-curative and will eventually fail to prevent opportunistic infections. Thus, while HLA-identical BMT remains the preferred treatment for SCID, there is a need for alternative treatments that can benefit all patients. The still-developing technology of hematopoietic stem cell gene therapy has the potential to become a relatively quick and safe method of curing SCIDs, especially for patients lacking other treatment options.

Viral vectors for gene therapy

RNA viruses, specifically retroviruses, are the most commonly used method of integrating a gene into the target cell[106, 107]. An MLV-derived gammaretroviral vector was used in the first successful gene therapy trials for SCID-X1 and ADA-SCID[108-111], discussed below. Human immunodeficiency virus (HIV)-derived lentiviral vectors have been proposed as an alternative to gammaretroviral vectors. Lentiviral vectors (LV) are able to transduce quiescent cells and display a markedly different integration pattern than gammaretroviral vectors[112]. LV vectors show no preference towards integrating near transcription start sites and are more resistant to silencing[113]. Recent investigations into alpharetroviral vectors suggest that they may have an even more favorable integration pattern[114].

IV. Gene therapy for SCID: the first clinical trials

Gene therapy is a form of molecular medicine based on the introduction of a healthy copy of a mutated, disease-causing gene into the somatic cells of a patient in order to treat the underlying disease. The procedure is most effective when applied to two categories of cell: easily accessible stem cells, such as hematopoietic stem cells, or terminally differentiated, long-lived cells such as muscle fibers. The type of target cell also dictates what types of gene therapy vector should be used. Vectors able to deliver the therapeutic gene into the host genome are required for dividing cells, while non-integrating vectors can be used for postmitotic cells. Because of previously-mentioned the selective advantage that corrected

T lymphoid progenitors have compared to uncorrected cells in an immunodeficient host, the ability to correct even a few hematopoietic stem cells may have clinical benefit to SCID patients.

The concept of gene therapy was first proposed in the 1970s, when scientists proposed that ‘gene surgery’ could be used to cure inherited diseases caused by simple genetic mutations. Gene surgery would involve the removal of the defective gene, replacing it with a healthy copy. While recent developments with zinc finger nucleases and transposons are encouraging[115, 116], the ability to perform such feats of gene surgery in the clinic is still years away. Currently, scientists are focusing on the simpler task of gene therapy: inserting a healthy copy of the defective gene into the patient’s cells to ameliorate or correct the disease.

As the technology capable of manipulating nucleic acids and the scientific community’s understanding of genetics advanced throughout the 1980s, gene therapy was proposed as a potential cure for a diverse set of inherited diseases ranging from hepatic (Factor XIV deficiency) to hematopoietic (ADA-SCID and sickle-cell anemia) to neuronal (Lesch-Nyhan Syndrome) [117-119].

In 1990, the first gene therapy clinical trial was conducted on two ADA-SCID patients in the United States[120]. The doctors repeatedly transduced a set of the patient’s lymphocytes with a retroviral vector containing the ADA gene *ex vivo*, with the hope that the gene-corrected cells would produce a more complete immune system once reintroduced into the patient. Over the next few years a total of 19 ADA-SCID patients would be treated via gene transfer to lymphocyte or T cell progenitors[121]. Gene-modified T cells and ADA protein expression persisted for several years after treatment, some cellular and humoral immune responses improved, and no adverse effects were seen. However, the level of gene transfer and engraftment were low, and full reversal of the ADA-SCID phenotype was not seen in any patient[122]. Furthermore, much of the improved immune response observed in some patients was attributable to the enzyme replacement therapy conducted concurrently with the gene therapy procedure. Nevertheless, this trial demonstrated that safe and long-term expression of therapeutic transgenes was possible in the clinic.

During the 1990s, advancements continued in the field of cell transduction and bone marrow transplantation, including improved *ex vivo* gene transfer protocols to CD34⁺ hematopoietic stem cells and administration of busulfan chemotherapy pre-transplant to improve engraftment. Ten years after the initial ADA-SCID trial, new developments in vector and cell transplantation led to successful clinical trials for gene therapy of ADA-SCID and SCID-X1.

The 2000 ADA-SCID and SCID-X1 clinical trials

In 2000, Dr. Aiuti of Milan initiated an ADA-SCID gene therapy clinical trial involving two young patients (2 years old and 7 months old) who lacked an HLA-identical sibling and were not performing on enzyme replacement therapy [108]. CD34⁺ BM cells were harvested from these patients, transduced with a gammaretroviral vector containing the

healthy ADA transgene, and transplanted back into the patients after nonmyeloablative conditioning. Subsequently, a total of 15 patients were enrolled in this gene therapy trial and all are alive and well. The first 10 patients treated were reported on in 2009, with 9 demonstrating full T cell reconstitution, 8 remaining off of ADA enzyme replacement therapy and immune globulin replacement therapy discontinued in half of the patients [123]. Gene therapy also corrected some of the central and peripheral B cell tolerance defects found in ADA-SCID patients[124]. Complications due to the treatment included prolonged neutropenia in two patients and single cases of hypertension, Epstein-Barr virus reactivation and autoimmune hepatitis. Importantly, no adverse effects due to the viral vector have been reported. A similar trial was conducted in London and reported successful transduction and engraftment in 3 out of 5 patients, with immune reconstitution seen in 2 of the 3 successfully treated[110]. A trial in Los Angeles emphasized the importance of pre-transplant myelosuppressive conditioning to improve transduced stem cell engraftment. Out of 4 patients given gene therapy without conditioning, 2 had low levels of gene marking and no sustained immunologic improvement[125]. A second clinical trial including low-dose myelosuppression with busulfan yielded production of ADA by peripheral blood mononuclear cells at 30% to 100% of normal levels in 5 out of 6 patients. However, one patient experienced a prolonged cytopenia after busulfan conditioning due to a pre-existing cytogenetic abnormality, highlighting a potential limitation for patients subject to autologous gene transfer[126].

Since 2000, over 40 patients have been treated for ADA-SCID via gene therapy in Italy, the UK and the USA, with majority reporting substantial clinical benefit. To date, all of these patients are still alive and 29 of them no longer require ADA enzyme replacement therapy[123, 127]. No adverse events attributable to the retroviral vector have been reported in any of the ADA-SCID gene therapy clinical trials, an important distinction from the SCID-X1 trials.

The first successful gene therapy trials for SCID-X1 were initiated in Paris in 1999 and reported in 2000, with several follow-up reports in the years since[128-131]. Drs. Fischer and Cavazzana-Calvo isolated CD34⁺ BM cells from five boys suffering from SCID-X1 and transduced them *ex vivo* with a gammaretroviral vector carrying a healthy copy of the *IL2RG* gene. Four of the five patients treated responded positively to the treatment and showed improved immune function. Following this initial trial, five more patients were treated in Paris and ten in a similar trial by Drs. Gaspar and Thrasher in London[111, 130] for a total of 20 patients treated between 1999 and 2006, with the majority enjoying clinical benefit. T-cell counts, diversity and proliferative responses normalized, and infections were cleared. All patients had proviral integrations in nearly 100% of their T cells, while integrations in B cell populations were very low. In most patients, T and NK cells populations and serum Ig levels reached healthy proportions within a year after treatment and have been sustained for up to 12 years. Overall, hematopoietic stem cell gene therapy restored T cell populations in 18 out of 20 patients. Of the two patients that did not recover immune system function, one was a 10-month old boy with disseminated bacille Calmette-Guérin infection and splenomegaly. An attempt at allogeneic hematopoietic stem

cell transplantation was unsuccessful, and the patient died. At the time of the last report, 17 of the original 20 patients are still alive, a success rate comparable to HLA-identical bone marrow transplantation, and six no longer need intravenous immunoglobulin[132]. The 2000-2002 French and UK SCID-X1 clinical trials demonstrate clear efficacy of gene therapy for a human genetic disease. They also pointed out some of the limitations of the procedure, namely the failure of gene therapy in two older patients, aged 15 and 20 years[133]. More than ten years after the initial trials, the results indicate that IL2RG gene therapy is an effective method for treating SCID-X1 infants, if not more mature patients.

However, the success of these SCID-X1 clinical trials was tempered by the development of leukemia-like symptoms in two of the Paris patients within three years post-therapy[131]. Further investigation revealed that the uncontrolled monoclonal proliferation of T cells in these patients was due to the gammaretroviral vector integrating near the proto-oncogene LIM domain-only 2 (*LMO2*), resulting in vector-derived insertional mutagenesis. Since the initial leukemias were reported, a total of 4 patients in the Paris trial and 1 in the London trial have been diagnosed with a similar condition[109, 134-136]. *LMO2* was identified as the primary oncogene in the London and three of the Paris patients, with the other patient having an integration near proto-oncogene *CCND2*. One patient had an integration near proto-oncogene *BMI1* as well as *LMO2*. Other genetic abnormalities included chromosomal translocations, gain-of-function mutations activating *NOTCH1*, copy number changes, deletion of tumor suppressor gene *CDKN2A*, 6q interstitial losses, and SIL-TAL1 rearrangement[137]. Four of the five affected patients were successfully treated by chemotherapy without reducing the effectiveness of the gene therapy, while the one patient succumbed in spite of chemotherapy treatment[129, 137, 138].

In response to the leukemia cases, in 2003 the Federal Drug Administration (FDA) of the United States briefly placed a hold on “all active gene therapy trials using retroviral vectors to insert genes into blood stem cells.” Later in the same year the FDA chose to allow gene therapy clinical trials to continue on a case-by-case basis if no other treatment was available[139].

V. INTEGRATION ANALYSIS OF THE SCID GENE THERAPY TRIALS

The leukemia cases in the SCID-X1 clinical trials have underlined the importance of integration site analysis of gene therapy recipients. This analysis can reveal integration patterns of the viral vector, including preferences towards certain genomic features or potential integration ‘hot spots’. Ideally, integration analysis in pre-clinical trials on animal models will reveal the relative safety of new gene therapy vectors under investigation, allowing those vectors with potentially dangerous integration patterns to be excluded before they are used in human patients and adding a further argument in favor of using new vectors that are deemed to have a lower risk of triggering insertional mutagenesis.

Integration analysis is done on DNA purified from various hematopoietic tissues including peripheral blood, spleen, bone marrow, thymus and lymph nodes. Often cells are

sorted by phenotype (T cell, B cell, etc) to check for lineage-specific integration patterns. The viral insertions contain sequences known to the researcher, and determining their location in the host genome can be accomplished by several methods. The two most commonly used are ligation-mediated PCR[140] (LM-PCR) and linear amplification-mediated PCR[141-143] (LAM-PCR). In LM-PCR, DNA is chemically cleaved then amplified via primers complimentary to the end of the viral long terminal repeats (LTRs). Linker sequences are added to the ends of the primer-annealed DNA and these fragments are amplified via PCR. This amplifies the genomic DNA adjacent to the viral LTR, which can be indentified through sequencing. LAM-PCR is an expansion of LM-PCR in which the viral LTR-genome boundary is amplified by biotinylated primers. Biotinylated extension products are then purified, cut with an endonuclease and ligated to a linker cassette before a further round of amplification. LAM-PCR has greater sensitivity and specificity compared to LM-PCR and is able to reveal more integration sites. Recently, a new form of non-restrictive LAM-PCR (nrLAM-PCR)has been developed[144]. This method does not use a restriction digest step, and thus does not produce an amplicon bias based on the proximity of the integration site to nearest site(s) targeted by the restriction enzymes seen in LM-PCR and conventional LAM-PCR.

After sequencing is complete, the viral LTR is trimmed off and the remaining, genomic sequence is aligned to the host genome via BLAST, BLAT or a similar sequence-matching platform. Once the sequences are aligned to a unique genomic site, researchers search the area around the integration site via genome databases such as Ensembl or NCBI. Genes near the integration site are typically the focus of the investigation, although other genomic features such as CpG islands or miRNA binding sites may also be of interest.

Insertional mutagenesis and oncogenesis

The potential for viral integration to cause oncogenesis has been known for decades [145]. Prior to the initial gene therapy clinical trials it was considered to be a very minor risk. After the incidents of leukemia in the SCID-X1 clinical trial, the oncogenic potential of gammaretroviral vectors has been reexamined.

The SCID-X1 gene therapy leukemias have been closely studied by the whole gene therapy community. The gammaretroviral vectors used in the Paris and London clinical trials were, in retrospect, sub-optimal with regards to safety. Analysis of the integration sites of the French and UK trials revealed a preference for gammaretroviral vectors to integrate in gene-rich regions, near highly expressed genes. Many proto-oncogenes are highly expressed in stem cell and progenitor cells, increasing the risk of a vector integrating nearby. Furthermore, gammaretroviral vectors have been shown to preferentially integrate near transcription start sites at the 5' end of genes[135, 136], a locus that has an increased risk of upregulating adjacent genes downstream. The risk of nearby gene upregulation is enhanced if the therapeutic gene is under the control of a strong, constitutively active promoter, such as the viral LTR used in the London and Paris SCID-X1 trials. The vector used in the French trial preferentially integrated into common fragile sites. LMO2 is one

such fragile site, which may explain why this oncogene was so heavily targeted for integration[146, 147].

The frequency of a transgene integration triggering oncogenesis were substantially higher in the SCID-X1 clinical trial than the pre-trial estimates. The initial estimates, based on murine gene therapy experiments, were in the range of 2×10^{-7} . However, this assumed that the integrating transgene was only capable of affecting genes very close by, within 10kb or less. It has since been demonstrated that retroviral integrations are capable of affecting the transcription of genes at a distance of at least 90kb and possibly as much as 500kb[148, 149]. The risk of introducing a transgene within a dangerous distance of a proto-oncogene in the human genome has since been calculated to be roughly 10^{-3} to 10^{-2} [150], or roughly 1 per 100,000 integrations. Considering that the SCID-X1 clinical trials are estimated to have produced more than 10^8 random integrations, the potential for oncogene activation is clear.

While the consensus is that the gammaretroviral vector integrations near proto-oncogenes was responsible for the leukemia onset, the possible role of *IL2RG* as a potential oncogene remains controversial[151-153]. Recent evidence supports the notion that *IL2RG* and *LMO2* require additional cooperating mutations to trigger a leukemic event, and that γ c overexpression itself does not initiate leukemogenesis [154, 155] In addition, experiments on *Il2rg*^{-/-} mice suggest that the SCID-X1 phenotype itself causes cells to be more susceptible to insertional leukemogenesis by murine leukemia viruses or gammaretroviral vectors[156, 157]. There are several possible reasons for this. For one, immunodeficient organisms have defective antitumor surveillance machinery. It has also been hypothesized that the expanded pool of T cell progenitors present in the SCID-X1 phenotype increases the risk of oncogenesis, as most of the leukemias found in gene-therapy treated XSCID mice and all the leukemias found in the clinical patients were T cell derived. In one experiment, the frequency of T cell transformation was significantly increased when using donor cells from X-SCID mice compared to cells from immunocompetent mice [157]. There may also be a strong selective advantage for pre-thymic T cell progenitors triggered by renewed IL-7 signaling.

VI. RATIONALE FOR THE STUDIES PERFORMED IN THIS THESIS

Hematopoietic stem cell gene therapy has the potential to become a valuable tool for the treatment of inherited congenital diseases which are otherwise problematic to treat. Its efficacy has already been demonstrated in clinical trials for ADA-SCID and SCID-X1. However, the SCID-X1 trials also revealed the significant risk of gene therapy triggering adverse events and highlighted a need for improved gene therapy vectors and an increased understanding of the mechanisms and behavior of those vectors. It is to fulfill these urgent requirements that the experiments described in this thesis were performed.

The studies detailed in the following chapters were designed with three primary aims: developing new lentiviral gene therapy vectors for treatment of SCID-X1, improving the

efficacy of gene therapy in a murine model, and generating a lentiviral vector safety profile via analysis of integration patterns and leukemic events in experimental animals. To this end, new self-inactivating (SIN) lentiviral vectors were developed that used one of two physiological promoters, the cellular phosphoglycerate kinase (PGK) promoter or a 1.1kb region of the *IL2RG* gene promoter (γ Pr), to drive *IL2RG* expression. Physiological promoters are much weaker than the viral promoters used in the clinical trials and thought to be less of a risk to trigger oncogenesis. These new LV vectors tested *in vivo* via transduction and transplantation of Lin⁻ cells from *Il2rg*^{-/-} mice into *Il2rg*^{-/-} recipients and compared transplantation of wild type Lin⁻ cells or cells transduced by a LV vector containing a strong viral promoter. Mice were followed for up to 1 year and were monitored monthly for circulating T and B cell levels as well as adverse events. Leukemias which developed in the course of these experiments were confirmed via secondary transplantation and carefully analyzed for phenotype and vector integration sites which may have contributed to oncogenesis.

Mice treated with HSC gene therapy are typically subjected to total body irradiation prior to transplantation, in order to create open niches in the bone marrow for donor HSCs to engraft. However, this sort of cytoreductive conditioning is not plausible in severely immunocompromised SCID-X1 patients, who often suffer from multiple recurrent infections at the time gene therapy is initiated. In order to determine if a non-cytoreductive method of creating open bone marrow niches could improve engraftment of donor HSCs, a pre-transplantation protocol using granulocyte colony stimulating factor (G-CSF) was initiated. *Il2rg*^{-/-} mice were transplanted with either wild type or LV vector-treated Lin⁻ cells after a conditioning regimen consisting of either 2 Gy irradiation or G-CSF mobilization of recipient HSCs.

Because analyzing integration sites by hand is both time consuming and introduces the possibility of human error, the analysis process was automated. A web-based integration site analysis platform was developed, named Methods for Analyzing ViRal Integration Collections (MAVRIC). MAVRIC is capable of quickly annotating trimmed genomic sequences and returns data on distribution of distances from nearest 5' gene end, nearest gene expression level binning, and Ensembl data on any gene within a user-defined distance from the integration site. MAVRIC was used to annotate integration sites from *IL2RG* gene therapy experiments, as well integration sites generated in other disease models for which gene therapy is being evaluated, to generate a safety profile for the lentiviral vectors under development.

A large pool of lentiviral integrations from various *in vitro* and *in vivo* experiments covering several SCIDs as well as the lysosomal storage disorder Pompe's disease were annotated using MAVRIC. These integration sites were stratified based on species, disease phenotype, tissue, transgene and promoter. The resulting comparative analyses demonstrates the stability of lentiviral integration site patterns across species and disease phenotypes, while uncovering evidence of selective pressures acting on the LV integration profile *in vivo*.

Combined, the experiments described in this thesis demonstrate that lentiviral gene therapy vectors using a codon optimized IL2RG gene driven by select physiological promoters is effective for treatment of SCID-X1 when combined with some level of pre-transplant conditioning or HSC mobilization. Analysis of vector integration patterns and leukemias observed in experimental animals suggests that lentiviral HSC gene therapy has a low risk of triggering adverse events and is sufficiently safe for the treatment of life-threatening diseases.

REFERENCES

1. Alberts, B., *Molecular biology of the cell*. 5th ed. 2008, New York: Garland Science. xxxiii, 1601, [90].
2. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology*. *Cell*, 2008. **132**(4): p. 631-44.
3. Lanza, F., L. Healy, and D.R. Sutherland, *Structural and functional features of the CD34 antigen: an update*. *J Biol Regul Homeost Agents*, 2001. **15**(1): p. 1-13.
4. Civin, C.I., et al., *Sustained, retransplantable, multilineage engraftment of highly purified adult human bone marrow stem cells in vivo*. *Blood*, 1996. **88**(11): p. 4102-9.
5. Osawa, M., et al., *Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell*. *Science*, 1996. **273**(5272): p. 242-5.
6. Peters, W.P., et al., *High-dose chemotherapy and autologous bone marrow support as consolidation after standard-dose adjuvant therapy for high-risk primary breast cancer*. *J Clin Oncol*, 1993. **11**(6): p. 1132-43.
7. Thomas, E., et al., *Bone-marrow transplantation (first of two parts)*. *N Engl J Med*, 1975. **292**(16): p. 832-43.
8. Amos, D.B. and F.H. Bach, *Phenotypic expressions of the major histocompatibility locus in man (HL-A): leukocyte antigens and mixed leukocyte culture reactivity*. *J Exp Med*, 1968. **128**(4): p. 623-37.
9. Rocha, V., et al., *Graft-versus-host disease in children who have received a cord-blood or bone marrow transplant from an HLA-identical sibling*. *Eurocord and International Bone Marrow Transplant Registry Working Committee on Alternative Donor and Stem Cell Sources*. *N Engl J Med*, 2000. **342**(25): p. 1846-54.
10. Bortin, M.M. and A.A. Rimm, *Severe combined immunodeficiency disease. Characterization of the disease and results of transplantation*. *Jama*, 1977. **238**(7): p. 591-600.
11. Glanzmann, E. and P. Riniker, [*Essential lymphocytophthisis; new clinical aspect of infant pathology.*] *Essentielle Lymphocytophthise; ein neues Krankheitsbild aus der Sauglingspathologie*. *Ann Paediatr*, 1950. **175**(1-2): p. 1-32.
12. Giblett, E.R., et al., *Adenosine-deaminase deficiency in two patients with severely impaired cellular immunity*. *Lancet*, 1972. **2**(7786): p. 1067-9.
13. Noguchi, M., et al., *Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans*. *Cell*, 1993. **73**(1): p. 147-57.
14. Puck, J.M., et al., *The interleukin-2 receptor gamma chain maps to Xq13.1 and is mutated in X-linked severe combined immunodeficiency, SCIDX1*. *Hum Mol Genet*, 1993. **2**(8): p. 1099-104.
15. Dadi, H.K., A.J. Simon, and C.M. Roifman, *Effect of CD3delta deficiency on maturation of alpha/beta and gamma/delta T-cell lineages in severe combined immunodeficiency*. *N Engl J Med*, 2003. **349**(19): p. 1821-8.
16. Kung, C., et al., *Mutations in the tyrosine phosphatase CD45 gene in a child with severe combined immunodeficiency disease*. *Nat Med*, 2000. **6**(3): p. 343-5.
17. Moshous, D., et al., *Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency*. *Cell*, 2001. **105**(2): p. 177-86.
18. Puel, A., et al., *Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency*. *Nat Genet*, 1998. **20**(4): p. 394-7.
19. Russell, S.M., et al., *Mutation of Jak3 in a patient with SCID: essential role of Jak3 in lymphoid development*. *Science*, 1995. **270**(5237): p. 797-800.
20. Schwarz, K., et al., *RAG mutations in human B cell-negative SCID*. *Science*, 1996. **274**(5284): p. 97-9.
21. Yee, A., et al., *Severe combined immunodeficiency: a national surveillance study*. *Pediatr Allergy Immunol*, 2008. **19**(4): p. 298-302.
22. Buckley, R.H., et al., *Human severe combined immunodeficiency: genetic, phenotypic, and functional diversity in one hundred eight infants*. *J Pediatr*, 1997. **130**(3): p. 378-87.
23. Stephan, J.L., et al., *A retrospective single-center study of clinical presentation and outcome in 117 patients with severe combined immunodeficiency*. *Immunodeficiency*, 1993. **4**(1-4): p. 87-8.

24. Denianke, K.S., et al., *Cutaneous manifestations of maternal engraftment in patients with severe combined immunodeficiency: a clinicopathologic study*. Bone Marrow Transplant, 2001. **28**(3): p. 227-33.
25. Muller, S.M., et al., *Transplacentally acquired maternal T lymphocytes in severe combined immunodeficiency: a study of 121 patients*. Blood, 2001. **98**(6): p. 1847-51.
26. Hirschhorn, R., *Overview of biochemical abnormalities and molecular genetics of adenosine deaminase deficiency*. Pediatr Res, 1993. **33**(1 Suppl): p. S35-41.
27. Hershfield, M.S., *Genotype is an important determinant of phenotype in adenosine deaminase deficiency*. Curr Opin Immunol, 2003. **15**(5): p. 571-7.
28. Parkman, R., et al., *Gene therapy for adenosine deaminase deficiency*. Annu Rev Med, 2000. **51**: p. 33-47.
29. Albuquerque, W. and H.B. Gaspar, *Bilateral sensorineural deafness in adenosine deaminase-deficient severe combined immunodeficiency*. J Pediatr, 2004. **144**(2): p. 278-80.
30. Cederbaum, S.D., et al., *The chondro-osseous dysplasia of adenosine deaminase deficiency with severe combined immunodeficiency*. J Pediatr, 1976. **89**(5): p. 737-42.
31. Rogers, M.H., et al., *Cognitive and behavioral abnormalities in adenosine deaminase deficient severe combined immunodeficiency*. J Pediatr, 2001. **139**(1): p. 44-50.
32. Migchielsen, A.A., et al., *Adenosine-deaminase-deficient mice die perinatally and exhibit liver-cell degeneration, atelectasis and small intestinal cell death*. Nat Genet, 1995. **10**(3): p. 279-87.
33. Wakamiya, M., et al., *Disruption of the adenosine deaminase gene causes hepatocellular impairment and perinatal lethality in mice*. Proc Natl Acad Sci U S A, 1995. **92**(9): p. 3673-7.
34. Blackburn, M.R., S.K. Datta, and R.E. Kellems, *Adenosine deaminase-deficient mice generated using a two-stage genetic engineering strategy exhibit a combined immunodeficiency*. J Biol Chem, 1998. **273**(9): p. 5093-100.
35. Barjaktarevic, I., et al., *Altered functional balance of Gfi-1 and Gfi-1b as an alternative cause of reticular dysgenesis?* Med Hypotheses, 2009.
36. Small, T.N., et al., *Association of reticular dysgenesis (thymic aplasia and congenital aleukocytosis) with bilateral sensorineural deafness*. J Pediatr, 1999. **135**(3): p. 387-9.
37. de, V.O. and V. Seynhaeve, *Reticular dysgenesis*. Lancet, 1959. **2**(7112): p. 1123-5.
38. Schwarz, K., et al., *Severe combined immunodeficiency (SCID) in man: B cell-negative (B-) SCID patients exhibit an irregular recombination pattern at the JH locus*. J Exp Med, 1991. **174**(5): p. 1039-48.
39. Villa, A., et al., *V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations*. Blood, 2001. **97**(1): p. 81-8.
40. Lewis, S.M., *The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses*. Adv Immunol, 1994. **56**: p. 27-150.
41. Tonegawa, S., *Somatic generation of antibody diversity*. Nature, 1983. **302**(5909): p. 575-81.
42. Mombaerts, P., et al., *RAG-1-deficient mice have no mature B and T lymphocytes*. Cell, 1992. **68**(5): p. 869-77.
43. Shinkai, Y., et al., *RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement*. Cell, 1992. **68**(5): p. 855-67.
44. Kato, M., et al., *Omenn syndrome--review of several phenotypes of Omenn syndrome and RAG1/RAG2 mutations in Japan*. Allergol Int, 2006. **55**(2): p. 115-9.
45. Omenn, G.S., *Familial Reticuloendotheliosis with Eosinophilia*. N Engl J Med, 1965. **273**: p. 427-32.
46. Aleman, K., et al., *Reviewing Omenn syndrome*. Eur J Pediatr, 2001. **160**(12): p. 718-25.
47. Corneo, B., et al., *Identical mutations in RAG1 or RAG2 genes leading to defective V(D)J recombinase activity can cause either T-B-severe combined immune deficiency or Omenn syndrome*. Blood, 2001. **97**(9): p. 2772-6.
48. Nicolas, N., et al., *A human severe combined immunodeficiency (SCID) condition with increased sensitivity to ionizing radiations and impaired V(D)J rearrangements defines a new DNA recombination/repair deficiency*. J Exp Med, 1998. **188**(4): p. 627-34.
49. Cavazzana-Calvo, M., et al., *Increased radiosensitivity of granulocyte macrophage colony-forming units and skin fibroblasts in human autosomal recessive severe combined immunodeficiency*. J Clin Invest, 1993. **91**(3): p. 1214-8.

50. Kobayashi, N., et al., *Novel Artemis gene mutations of radiosensitive severe combined immunodeficiency in Japanese families*. Hum Genet, 2003. **112**(4): p. 348-52.
51. Noordzij, J.G., et al., *Radiosensitive SCID patients with Artemis gene mutations show a complete B-cell differentiation arrest at the pre-B-cell receptor checkpoint in bone marrow*. Blood, 2003. **101**(4): p. 1446-52.
52. Rooney, S., et al., *Leaky Scid phenotype associated with defective V(D)J coding end processing in Artemis-deficient mice*. Mol Cell, 2002. **10**(6): p. 1379-90.
53. van der Burg, M., et al., *A new type of radiosensitive T-B-NK⁺ severe combined immunodeficiency caused by a LIG4 mutation*. J Clin Invest, 2006. **116**(1): p. 137-45.
54. van Gent, D.C., J.H. Hoeijmakers, and R. Kanaar, *Chromosomal stability and the DNA double-stranded break connection*. Nat Rev Genet, 2001. **2**(3): p. 196-206.
55. Barnes, D.E., et al., *Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice*. Curr Biol, 1998. **8**(25): p. 1395-8.
56. Frank, K.M., et al., *Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV*. Nature, 1998. **396**(6707): p. 173-7.
57. Gao, Y., et al., *A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis*. Cell, 1998. **95**(7): p. 891-902.
58. Buckley, R.H., *The multiple causes of human SCID*. J Clin Invest, 2004. **114**(10): p. 1409-11.
59. Roifman, C.M., et al., *A partial deficiency of interleukin-7R alpha is sufficient to abrogate T-cell development and cause severe combined immunodeficiency*. Blood, 2000. **96**(8): p. 2803-7.
60. Peschon, J.J., et al., *Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice*. J Exp Med, 1994. **180**(5): p. 1955-60.
61. Tchilian, E.Z., et al., *A deletion in the gene encoding the CD45 antigen in a patient with SCID*. J Immunol, 2001. **166**(2): p. 1308-13.
62. Trowbridge, I.S. and M.L. Thomas, *CD45: an emerging role as a protein tyrosine phosphatase required for lymphocyte activation and development*. Annu Rev Immunol, 1994. **12**: p. 85-116.
63. Weiss, A. and D.R. Littman, *Signal transduction by lymphocyte antigen receptors*. Cell, 1994. **76**(2): p. 263-74.
64. Thomas, M.L., *The leukocyte common antigen family*. Annu Rev Immunol, 1989. **7**: p. 339-69.
65. de Saint Basile, G., et al., *Severe combined immunodeficiency caused by deficiency in either the delta or the epsilon subunit of CD3*. J Clin Invest, 2004. **114**(10): p. 1512-7.
66. Recio, M.J., et al., *Differential biological role of CD3 chains revealed by human immunodeficiencies*. J Immunol, 2007. **178**(4): p. 2556-64.
67. Roberts, J.L., et al., *T-B+NK⁺ severe combined immunodeficiency caused by complete deficiency of the CD3zeta subunit of the T-cell antigen receptor complex*. Blood, 2007. **109**(8): p. 3198-206.
68. Call, M.E. and K.W. Wucherpfennig, *The T cell receptor: critical role of the membrane environment in receptor assembly and function*. Annu Rev Immunol, 2005. **23**: p. 101-25.
69. Clevers, H., et al., *The T cell receptor/CD3 complex: a dynamic protein ensemble*. Annu Rev Immunol, 1988. **6**: p. 629-62.
70. Irving, B.A., A.C. Chan, and A. Weiss, *Functional characterization of a signal transducing motif present in the T cell antigen receptor zeta chain*. J Exp Med, 1993. **177**(4): p. 1093-103.
71. Samelson, L.E. and R.D. Klausner, *Tyrosine kinases and tyrosine-based activation motifs. Current research on activation via the T cell antigen receptor*. J Biol Chem, 1992. **267**(35): p. 24913-6.
72. Arnaiz-Villena, A., et al., *T lymphocyte signalling defects and immunodeficiency due to the lack of CD3 gamma*. Immunodeficiency, 1993. **4**(1-4): p. 121-9.
73. Arnaiz-Villena, A., et al., *Brief report: primary immunodeficiency caused by mutations in the gene encoding the CD3-gamma subunit of the T-lymphocyte receptor*. N Engl J Med, 1992. **327**(8): p. 529-33.
74. Regueiro, J.R., et al., *Familial defect of CD3 (T3) expression by T cells associated with rare gut epithelial cell autoantibodies*. Lancet, 1986. **1**(8492): p. 1274-5.
75. Haks, M.C., et al., *The CD3gamma chain is essential for development of both the TCRalpha and TCRgamma lineages*. Embo J, 1998. **17**(7): p. 1871-82.

76. Berger, M.A., et al., *Subunit composition of pre-T cell receptor complexes expressed by primary thymocytes: CD3 delta is physically associated but not functionally required.* J Exp Med, 1997. **186**(9): p. 1461-7.
77. Dave, V.P., et al., *CD3 delta deficiency arrests development of the alpha beta but not the gamma delta T cell lineage.* Embo J, 1997. **16**(6): p. 1360-70.
78. Asao, H., et al., *Cutting edge: the common gamma-chain is an indispensable subunit of the IL-21 receptor complex.* J Immunol, 2001. **167**(1): p. 1-5.
79. Giri, J.G., et al., *Utilization of the beta and gamma chains of the IL-2 receptor by the novel cytokine IL-15.* Embo J, 1994. **13**(12): p. 2822-30.
80. Kimura, Y., et al., *Sharing of the IL-2 receptor gamma chain with the functional IL-9 receptor complex.* Int Immunol, 1995. **7**(1): p. 115-20.
81. Kondo, M., et al., *Sharing of the interleukin-2 (IL-2) receptor gamma chain between receptors for IL-2 and IL-4.* Science, 1993. **262**(5141): p. 1874-7.
82. Noguchi, M., et al., *Interleukin-2 receptor gamma chain: a functional component of the interleukin-7 receptor.* Science, 1993. **262**(5141): p. 1877-80.
83. Leonard, W.J., *The molecular basis of X-linked severe combined immunodeficiency: defective cytokine receptor signaling.* Annu Rev Med, 1996. **47**: p. 229-39.
84. Leonard, W.J. and J.J. O'Shea, *Jaks and STATs: biological implications.* Annu Rev Immunol, 1998. **16**: p. 293-322.
85. Kennedy, M.K., et al., *Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice.* J Exp Med, 2000. **191**(5): p. 771-80.
86. DiSanto, J.P., et al., *Lymphoid development in mice with a targeted deletion of the interleukin 2 receptor gamma chain.* Proc Natl Acad Sci U S A, 1995. **92**(2): p. 377-81.
87. Bousso, P., et al., *Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor.* Proc Natl Acad Sci U S A, 2000. **97**(1): p. 274-8.
88. Stephan, V., et al., *Atypical X-linked severe combined immunodeficiency due to possible spontaneous reversion of the genetic defect in T cells.* N Engl J Med, 1996. **335**(21): p. 1563-7.
89. Macchi, P., et al., *Mutations of Jak-3 gene in patients with autosomal severe combined immune deficiency (SCID).* Nature, 1995. **377**(6544): p. 65-8.
90. Pesu, M., et al., *Jak3, severe combined immunodeficiency, and a new class of immunosuppressive drugs.* Immunol Rev, 2005. **203**: p. 127-42.
91. Miyazaki, T., et al., *Functional activation of Jak1 and Jak3 by selective association with IL-2 receptor subunits.* Science, 1994. **266**(5187): p. 1045-7.
92. Russell, S.M., et al., *Interaction of IL-2R beta and gamma c chains with Jak1 and Jak3: implications for XSCID and XCID.* Science, 1994. **266**(5187): p. 1042-5.
93. Park, S.Y., et al., *Developmental defects of lymphoid cells in Jak3 kinase-deficient mice.* Immunity, 1995. **3**(6): p. 771-82.
94. Thomis, D.C., et al., *Defects in B lymphocyte maturation and T lymphocyte activation in mice lacking Jak3.* Science, 1995. **270**(5237): p. 794-7.
95. Buckley, R.H., *Molecular defects in human severe combined immunodeficiency and approaches to immune reconstitution.* Annu Rev Immunol, 2004. **22**: p. 625-55.
96. Gatti, R.A., et al., *Immunological reconstitution of sex-linked lymphopenic immunological deficiency.* Lancet, 1968. **2**(7583): p. 1366-9.
97. Buckley, R.H., et al., *Hematopoietic stem-cell transplantation for the treatment of severe combined immunodeficiency.* N Engl J Med, 1999. **340**(7): p. 508-16.
98. Haddad, E., et al., *Long-term immune reconstitution and outcome after HLA-nonidentical T-cell-depleted bone marrow transplantation for severe combined immunodeficiency: a European retrospective study of 116 patients.* Blood, 1998. **91**(10): p. 3646-53.
99. Fischer, A., *Thirty years of bone marrow transplantation for severe combined immunodeficiency.* N Engl J Med, 1999. **340**(7): p. 559-61.

100. Antoine, C., et al., *Long-term survival and transplantation of haemopoietic stem cells for immunodeficiencies: report of the European experience 1968-99*. Lancet, 2003. **361**(9357): p. 553-60.
101. Cavazzana-Calvo, M., A. Thrasher, and F. Mavilio, *The future of gene therapy*. Nature, 2004. **427**(6977): p. 779-81.
102. Fischer, A., et al., *European experience of bone-marrow transplantation for severe combined immunodeficiency*. Lancet, 1990. **336**(8719): p. 850-4.
103. Beatty, P.G., et al., *Marrow transplantation from related donors other than HLA-identical siblings*. N Engl J Med, 1985. **313**(13): p. 765-71.
104. Haddad, E., et al., *Long-term chimerism and B-cell function after bone marrow transplantation in patients with severe combined immunodeficiency with B cells: A single-center study of 22 patients*. Blood, 1999. **94**(8): p. 2923-30.
105. Bertrand, Y., et al., *Influence of severe combined immunodeficiency phenotype on the outcome of HLA non-identical, T-cell-depleted bone marrow transplantation: a retrospective European survey from the European group for bone marrow transplantation and the European society for immunodeficiency*. J Pediatr, 1999. **134**(6): p. 740-8.
106. Kay, M.A., J.C. Glorioso, and L. Naldini, *Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics*. Nat Med, 2001. **7**(1): p. 33-40.
107. Verma, I.M. and M.D. Weitzman, *Gene therapy: twenty-first century medicine*. Annu Rev Biochem, 2005. **74**: p. 711-38.
108. Aiuti, A., et al., *Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning*. Science, 2002. **296**(5577): p. 2410-3.
109. Cavazzana-Calvo, M., et al., *Gene therapy for severe combined immunodeficiency*. Annu Rev Med, 2005. **56**: p. 585-602.
110. Gaspar, H.B., et al., *Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning*. Mol Ther, 2006. **14**(4): p. 505-13.
111. Gaspar, H.B., et al., *Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector*. Lancet, 2004. **364**(9452): p. 2181-7.
112. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration*. Nat Biotechnol, 2006. **24**(6): p. 687-96.
113. Pfeifer, A., et al., *Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos*. Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2140-5.
114. Suerth, J.D., et al., *Alpharetroviral self-inactivating vectors: long-term transgene expression in murine hematopoietic cells and low genotoxicity*. Mol Ther, 2012. **20**(5): p. 1022-32.
115. Urnov, F.D., et al., *Highly efficient endogenous human gene correction using designed zinc-finger nucleases*. Nature, 2005. **435**(7042): p. 646-51.
116. Zayed, H., et al., *Development of hyperactive sleeping beauty transposon vectors by mutational analysis*. Mol Ther, 2004. **9**(2): p. 292-304.
117. Miller, J., *Toward Gene Therapy: Lesch-Nyhan Syndrome*. Science News, 1983. **124**(6): p. 90-91.
118. Anderson, W.F., *Prospects for human gene therapy*. Science, 1984. **226**(4673): p. 401-9.
119. Ledley, F.D., et al., *Retroviral gene transfer into primary hepatocytes: implications for genetic therapy of liver-specific functions*. Proc Natl Acad Sci U S A, 1987. **84**(15): p. 5335-9.
120. Culver, K.W., W.F. Anderson, and R.M. Blaese, *Lymphocyte gene therapy*. Hum Gene Ther, 1991. **2**(2): p. 107-9.
121. Bordignon, C., et al., *Gene therapy in peripheral blood lymphocytes and bone marrow for ADA- immunodeficient patients*. Science, 1995. **270**(5235): p. 470-5.
122. Blaese, R.M., et al., *T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years*. Science, 1995. **270**(5235): p. 475-80.
123. Ai uti, A., et al., *Gene therapy for immunodeficiency due to adenosine deaminase deficiency*. N Engl J Med, 2009. **360**(5): p. 447-58.

124. Sauer, A.V., et al., *Defective B cell tolerance in adenosine deaminase deficiency is corrected by gene therapy*. J Clin Invest, 2012. **122**(6): p. 2141-52.
125. Gaspar, H.B., et al., *How I treat ADA deficiency*. Blood, 2009. **114**(17): p. 3524-32.
126. Engel, B.C., et al., *Prolonged pancytopenia in a gene therapy patient with ADA-deficient SCID and trisomy 8 mosaicism: a case report*. Blood, 2007. **109**(2): p. 503-6.
127. Gaspar, H.B., et al., *Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction*. Sci Transl Med, 2011. **3**(97): p. 97ra80.
128. Cavazzana-Calvo, M., et al., *Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease*. Science, 2000. **288**(5466): p. 669-72.
129. Fischer, A. and M. Cavazzana-Calvo, *Gene therapy of inherited diseases*. Lancet, 2008. **371**(9629): p. 2044-7.
130. Hacein-Bey-Abina, S., et al., *Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy*. N Engl J Med, 2002. **346**(16): p. 1185-93.
131. Hacein-Bey-Abina, S., et al., *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1*. Science, 2003. **302**(5644): p. 415-9.
132. Cavazzana-Calvo, M., et al., *Gene therapy for primary immunodeficiencies: part I*. Curr Opin Immunol, 2012. **24**(5): p. 580-4.
133. Thrasher, A.J., et al., *Failure of SCID-X1 gene therapy in older patients*. Blood, 2005. **105**(11): p. 4255-7.
134. Bushman, F.D., *Retroviral integration and human gene therapy*. J Clin Invest, 2007. **117**(8): p. 2083-6.
135. Deichmann, A., et al., *Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy*. J Clin Invest, 2007. **117**(8): p. 2225-32.
136. Schwarzwaelder, K., et al., *Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo*. J Clin Invest, 2007. **117**(8): p. 2241-9.
137. Hacein-Bey-Abina, S., et al., *Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1*. J Clin Invest, 2008. **118**(9): p. 3132-42.
138. Howe, S.J., et al., *Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients*. J Clin Invest, 2008. **118**(9): p. 3143-50.
139. Kaiser, J., *Gene therapy. Panel urges limits on X-SCID trials*. Science, 2005. **307**(5715): p. 1544-5.
140. Pfeifer, G.P., et al., *Genomic sequencing and methylation analysis by ligation mediated PCR*. Science, 1989. **246**(4931): p. 810-3.
141. Schmidt, M., et al., *High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR)*. Nat Methods, 2007. **4**(12): p. 1051-7.
142. Schmidt, M., et al., *Clonality analysis after retroviral-mediated gene transfer to CD34+ cells from the cord blood of ADA-deficient SCID neonates*. Nat Med, 2003. **9**(4): p. 463-8.
143. Schmidt, M., et al., *Polyclonal long-term repopulating stem cell clones in a primate model*. Blood, 2002. **100**(8): p. 2737-43.
144. Paruzynski, A., et al., *Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing*. Nat Protoc, 2010. **5**(8): p. 1379-95.
145. Trentin, J.J., Y. Yabe, and G. Taylor, *The quest for human cancer viruses*. Science, 1962. **137**: p. 835-41.
146. Wu, X., et al., *Transcription start regions in the human genome are favored targets for MLV integration*. Science, 2003. **300**(5626): p. 1749-51.
147. Bester, A.C., et al., *Fragile sites are preferential targets for integrations of MLV vectors in gene therapy*. Gene Ther, 2006. **13**(13): p. 1057-9.
148. Bartholomew, C. and J.N. Ihle, *Retroviral insertions 90 kilobases proximal to the Evi-1 myeloid transforming gene activate transcription from the normal promoter*. Mol Cell Biol, 1991. **11**(4): p. 1820-8.
149. Kustikova, O., et al., *Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking*. Science, 2005. **308**(5725): p. 1171-4.
150. Baum, C., et al., *Side effects of retroviral gene transfer into hematopoietic stem cells*. Blood, 2003. **101**(6): p. 2099-114.

151. Pike-Overzet, K., et al., *Gene therapy: is IL2RG oncogenic in T-cell development?* Nature, 2006. **443**(7109): p. E5; discussion E6-7.
152. Thrasher, A.J., et al., *Gene therapy: X-SCID transgene leukaemogenicity.* Nature, 2006. **443**(7109): p. E5-6; discussion E6-7.
153. Woods, N.B., et al., *Gene therapy: therapeutic gene causing lymphoma.* Nature, 2006. **440**(7088): p. 1123.
154. Dave, U.P., et al., *Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy.* PLoS Genet, 2009. **5**(5): p. e1000491.
155. Ginn, S.L., et al., *Lymphomagenesis in SCID-X1 mice following lentivirus-mediated phenotype correction independent of insertional mutagenesis and gammac overexpression.* Mol Ther. **18**(5): p. 965-76.
156. Scobie, L., et al., *A novel model of SCID-X1 reconstitution reveals predisposition to retrovirus-induced lymphoma but no evidence of gammaC gene oncogenicity.* Mol Ther, 2009. **17**(6): p. 1031-8.
157. Shou, Y., et al., *Unique risk factors for insertional mutagenesis in a mouse model of XSCID gene therapy.* Proc Natl Acad Sci U S A, 2006. **103**(31): p. 11730-5.

Met TyrXle Ala Cys
 Met Xle Ala Cys
 Cys Lys Ser Cys Asn Pyl Thr
 Ala Lys Ser Cys Lys Pyl Thr
 Xle Ser Pyl Asn Sar SecSer
 PylAsn Glu Asn Sar Pyl AsaAsnHis IleAsaAsn
 Ser Leu SerThr LeuGlu Asn TrpIle LeuLeu His Sec
 Lys Cys Glu Ser GGAC GCC GGCATGCC TrpIle Asn Ser
 Ala Xle Thr AGCAGGTACGCGAGA His Ala PylThr
 GluArg AGGCTAGCTAG Arg His PylThr
 CAGGCTCGA PylAsn Arg AlaArg ArgIle
 CAGTGGGTAT Thr Ser Ile Asn Glu Thr
 GTGAGCTTGCTC Ser Glu Pyl Thr His
 GATAGCGGGTA Sec Thr Thr His
 TGCTAGCCCGCC His SerSec
 GGGAAATGCAGG
 TACGCTTGCTAGA
 CCGCAGGCATGC
 CTGTATGCTTGCTC
 01001101011000010111
 00100 1110011011010000
 01100 001011011000110
 1100 0010000001010111
 011 010010110110001101
 10 0011010010110000101
 101101001 00000010010
 00011101 01011100110111010
 0011011 11011011100010110000
 100000 01010
 00001 1010
 000 100100



Correction of murine SCID-X1 by lentiviral gene therapy using a codon optimized IL2RG gene and minimal pre-transplant conditioning

Marshall W. Huston¹, Niek P. van Til¹, Trudi P. Visser¹, Shazia Arshad¹, Martijn H. Brugman^{1,2}, Claudia Cattoglio³, Ali Nowrouzi⁴, Yuedan Li¹, Axel Schambach², Manfred Schmidt⁴, Christopher Baum², Christof von Kalle⁴, Fulvio Mavilio^{3,5}, Fang Zhang⁶, Mike P. Blundell⁶, Adrian J Thrasher^{6,7}, Monique M.A. Verstegen¹, and Gerard Wagemaker¹

¹Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands

²Department of Experimental Hematology, Hannover Medical School, Hannover, Germany

³Istituto Scientifico H. San Raffaele, Milan, Italy

⁴Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center (DKFZ), Heidelberg, Germany

⁵Department of Biomedical Sciences, University of Modena and Reggio Emilia, Modena, Italy

⁶Centre for Immunodeficiency, Molecular Immunology Unit, Institute of Child Health, University College London, London, UK

⁷Department of Clinical Immunology, Great Ormond Street Hospital NHS Trust, London, UK

ABSTRACT

Clinical trials have demonstrated the potential of *ex vivo* hematopoietic stem cell gene therapy to treat X-linked severe combined immunodeficiency (SCID-X1) using gammaretroviral vectors, leading to immune system functionality in the majority of treated patients without pre-transplant conditioning. The success was tempered by insertional oncogenesis in a proportion of the patients. To reduce the genotoxicity risk, a self-inactivating (SIN) lentiviral vector with improved expression of a codon optimized human IL-2 receptor gamma gene (*IL2RG*) cDNA (co γ c), regulated by its 1.1kb promoter region (γ cPr), was compared in efficacy to the viral spleen focus forming virus (SF) and the cellular phosphoglycerate kinase (PGK) promoters. Pre-transplant conditioning of *IL2rg*^{-/-} mice resulted in long-term reconstitution of T and B lymphocytes, normalized natural antibody titers, humoral immune responses, ConA/IL-2 stimulated spleen cell proliferation, and polyclonal T cell receptor gene rearrangements with a clear integration preference of the SF vector for proto-oncogenes, contrary to the PGK and γ cPr vectors. We conclude that SIN lentiviral gene therapy using co γ c driven by the γ cPr or PGK promoter corrects the SCID phenotype, potentially with an improved safety profile, and that low dose conditioning proved essential for immune competence, allowing for a reduced threshold of cell numbers required.

INTRODUCTION

X-linked severe combined immunodeficiency disease (SCID-X1) is a rare disorder caused by mutations in the IL-2 receptor gamma gene (*IL2RG*) leading to a non-functional common gamma chain protein (γ_c). The γ_c protein is a shared subunit of IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 signaling, critical for growth and maturation of lymphocyte populations¹. Patients suffer from a complete absence of mature T and natural killer (NK) cells. B cells are present but dysfunctional in SCID-X1 humans and absent in *Il2rg*^{-/-} mice. Without treatment, patients succumb to severe, recurrent infections within the first year of life. Bone marrow (BM) transplantation from an HLA-identical donor is life-saving with a high success rate.² However, HLA-identical donors are only available to one out of every three patients. Non-HLA-identical BM transplantation has been attempted but carries a vastly increased risk of complications and mortality, making SCID-X1 an excellent candidate for clinical gene therapy trials for those patients who lack such a donor.

The efficacy of gammaretroviral (RV) gene therapy to successfully treat SCID-X1 was demonstrated in seminal clinical trials.³⁻⁴ However, some patients required immunoglobulin-replacement therapy. Additionally, 5 out of 20 patients developed leukemia several years post-therapy due to RV-induced insertional mutagenesis, which in most cases was successfully treated without reducing the effectiveness of the gene therapy.⁵⁻⁷

To increase safety of future gene therapy treatment, the use of self-inactivating (SIN) RV and lentiviral vectors (LV) incorporating internal promoters has been proposed, which should reduce the risk of undesirable activation of genes adjacent to the integration site⁸⁻⁹ and the bias towards integrating near transcription start sites.¹⁰⁻¹² LV are less vulnerable to silencing than RV¹³ and do not require *ex vivo* growth factor stimulation, thereby preventing loss of stem cell repopulating capacity.

We generated a series of third generation SIN lentiviral vectors with a codon optimized human *IL2RG* open reading frame (hereafter designated *coyc*) that markedly improved mRNA transcription and translation.¹⁴⁻¹⁵ Transgene expression was driven by the strong spleen focus forming virus (SF) promoter to determine the potential genotoxicity of the procedure. For clinical application, the human phosphoglycerate kinase¹⁶ (PGK) promoter and a ~1.1kb section of the native *IL2RG* promoter sequence¹⁷ (γ_c Pr) were tested in *Il2rg*^{-/-} mice, since cellular/physiological promoters have been suggested to further improve vector safety.¹⁸ We found that the clinical lentiviral vectors provided safe and functional correction of the phenotype of SCID-X1 mice, and that the level of pre-transplant conditioning was a critical determinant for successful B cell reconstitution and immunoglobulin responses.

RESULTS

Lentiviral vector construction for SCID-X1 gene therapy

A SIN LV¹⁹⁻²⁰ incorporating part of the native human *IL2RG* complementary DNA (cDNA) driven by the viral SF promoter²¹ was improved for efficacy by replacement of the native *IL2RG* by the *coyc* sequence, and resulting in an average 8.4-fold increase in viral titer, a 3.6-fold increase in transgene mRNA expression per integration, a 33-fold increase in IL2RG protein expression corresponding to a 3-fold increase in IL2RG protein membrane expression (Tables 1 and 2 and Figure 1). In addition, the cellular PGK or γ Pr promoter was used to control *coyc* expression, the latter shown capable of driving transgene expression in lymphoid cell lines.¹⁷

Table 1. Titer comparison of SF- γ c and SF-co γ c lentiviral (LV) vectors.

LV vector	Viral Titer (TU/ml) $\times 10^6$	mRNA expression per integration (picograms)	Fold increase in γ c protein expression
SF-co γ c	1.21	1.77	33
SF- γ c	0.14	0.45	1

Titers, IL2RG mRNA levels and relative γ c protein levels were determined in HeLa cells for both SF- γ c and SF-co γ c lentiviral vector ($n = 2$). After 7 days the number of integrations per cell and WPRE mRNA expression per integration were measured by qPCR of DNA and cDNA, respectively, and γ c protein levels were measured via ELISA assay.

Table 2. Comparison of surface γ c expression between SF- γ c and SF-co γ c LV.

LV vector	Percent γ c positive	Mean Fluorescence Intensity (MFI)	LV integrations per cell
SF-co γ c MOI 100	52.7	2890	5.43
SF-co γ c MOI 10	31.3	1856	1.33
SF- γ c MOI 100	15	1151	5.76
SF- γ c MOI 10	5.3	1538	0.51
negative	4.0	2057	0.00

Lineage negative *Il2rg*^{-/-} BM cells were transduced with SF- γ c or SF-co γ c at MOI 100 or 10, cultured for 4 days, and then measured for surface γ c expression via flow cytometry. Integrations per cell were determined via qPCR.

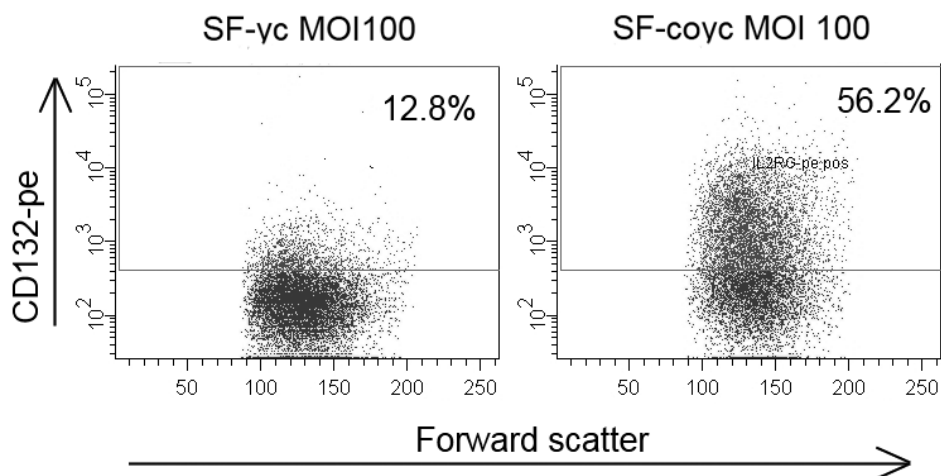


Figure 1. Comparison of codon-optimized γc vector vs non-optimized *in vitro*. Flow cytometry images of Lin⁻ cells transduced with either LV-SF- γc or LV-SF-co γc at a Hela MOI of 100. Surface expression of CD132 is much higher in the cells transduced with the codon optimized vector, with a similar number of integrations per cell found in both vectors (5.4 for co γc and 5.8 for γc).

Wild type Lin⁻ BM cells were transduced with a non-saturating but fixed MOI of γc PROM-GFP (BM copy number per cell: 4.9), PGK-GFP (BM copy number per cell: not determined) or SF-GFP LV vector (BM copy number per cell: 3.2) and transplanted in 6 Gy sublethal TBI conditioned mice to compare promoter specificity and activity (Figure 2). In the γc Pr-GFP mice, GFP expression in peripheral blood 3 months post-transplantation was present on average in 50% of CD3⁺ cells, 92% of IgM/B220⁺ cells and 65% of CD11b⁺ cells, whereas expression in erythrocytes and platelets was negligible. Similar percentages of GFP expression were observed in the constitutively active SF-GFP group, with detectable GFP levels in erythrocytes and thrombocytes (36-42%), similar to the constitutively active PGK promoter. Fluorescence intensity in CD3⁺ cells was similar across all 3 promoters, but GFP intensity was for the SF-GFP group 5-fold higher in B cells and 4-fold higher in CD11b⁺ cells compared to the physiological promoters (Figure 2b). GFP expression in γc Pr-GFP-transduced cells was stable for eight months. We concluded that the γc promoter element displayed improved stable selectivity with a lower expression level compared to the SF promoter.

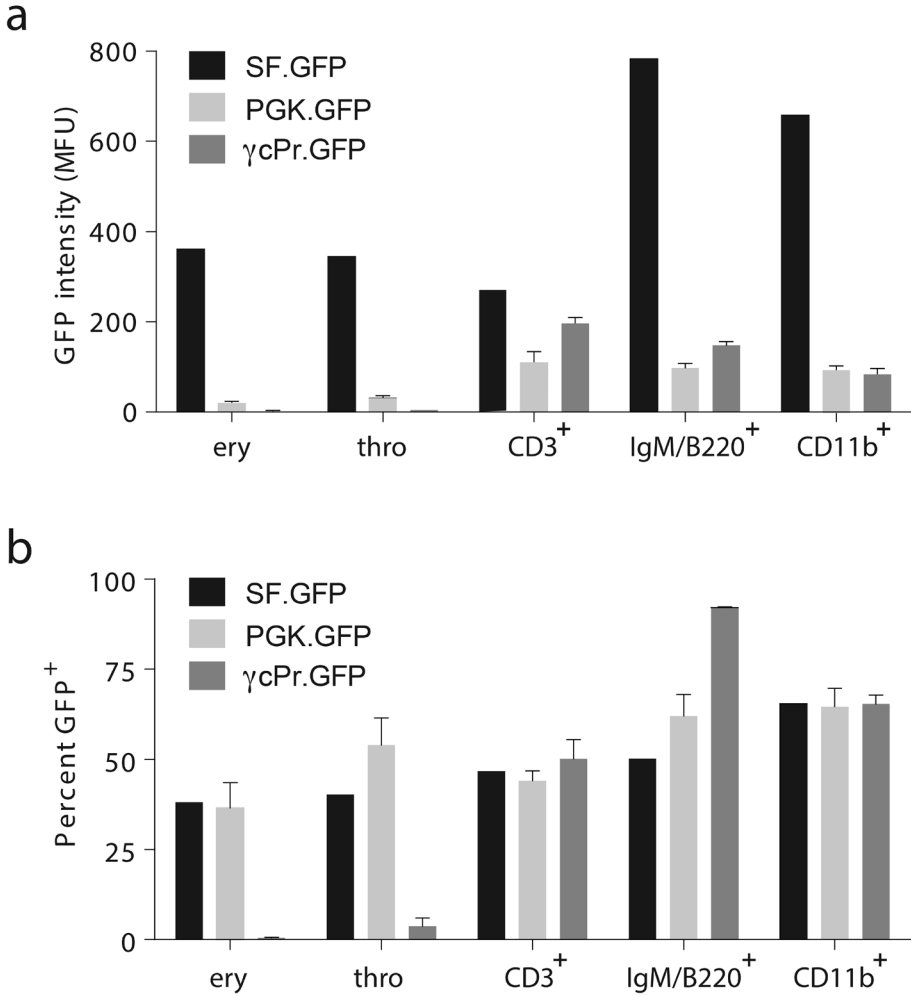


Figure 2. Comparison of SF, PGK and γ cPr promoter expression *in vivo*. (a) the proportion of GFP positive erythrocytes (ery), thrombocytes (thro), T lymphocytes (CD3⁺), B lymphocytes (IgM/B220⁺) and myeloid cells (CD11b⁺) in *Il2rg*^{-/-} mice three months after transplantation of wild type Lin⁻ cells transduced with SF-GFP (dark grey bars, n=2), PGK-GFP (light grey bars, n=3) or γ cPr-GFP (black bars, n=3). (b) shows the fluorescence intensity (mean fluorescent units or MFU) of GFP expression.

T and B cell reconstitution following lentiviral gene therapy for murine SCID-X1

4×10^5 SF-coyc, γ cPr-coyc (both MOI 9-10) or PGK-coyc (MOI 2, due to low vector titer) transduced male *Il2rg*^{-/-} Lin⁻ bone marrow cells were transplanted into 6 Gy irradiated female *Il2rg*^{-/-} recipients (n = 5 for PGK and n = 4 for SF and γ cPr). As controls, *Il2rg*^{-/-} and wild type Lin⁻ BM cells were transduced with γ cPr-GFP and transplanted into *Il2rg*^{-/-} mice (n = 4 and n = 3, respectively). The number of integrations per BM cell corrected for donor/recipient chimerism was proportional to the MOI, reaching 2.1, 0.9 and 2.0 for the groups SF-coyc, PGK-coyc and γ cPr-coyc respectively (Supplementary Table S1a). The control groups containing the γ cPr-GFP vector on average had 4.9 integrations for wild type transplanted cells and 4.0 integrations for *Il2rg*^{-/-} cells. Polyclonal integration patterns were confirmed for all vectors by LAM-PCR of BM DNA at 8 months post-transplant (Figure S1). An initial, limited analysis revealed the LV characteristic preferential integration in active transcription units, with no obvious preference for transcription start sites. Of note, potential oncogenes, listed in the mouse retrovirus tagged cancer gene database (RTCGD).²² were over-represented among integrations of vectors containing the SF promoter. Contrary to the vectors with cellular promoters, almost 15% of SF integrations were found within 100kb of a gene listed in the RTCGD (Table 3). Even in this limited data set, this difference is statistically significant (P=0.02).

Peripheral white blood cell counts increased in all gene therapy treated mice and were stable for 7 months post-therapy. T and B lymphocyte reconstitution is shown in Figure 3a. Animals transplanted with *Il2rg*^{-/-} cells treated with γ cPr-GFP had 7-11% peripheral blood CD3⁺CD4⁺ cells, depending on the time interval after transplantation, and were negative for CD8⁺, B220⁺IgM⁺, IgD⁺ and CD19⁺ cells. Animals transplanted with *Il2rg*^{-/-} cells treated with SF-coyc, PGK-coyc or γ cPr-coyc began to recover T and B cell populations two months post-transplant, as did mice given wild type cells. Reconstitution of CD4⁺ and CD8⁺ cells were similar for the coyc-treated groups and for the wild type group, and T cell populations remained stable. Two-sided ANOVA comparing to the GFP group returned P values of <0.01 for both T and B lymphocytes for wild type and all coyc-treated groups at months 5-7 after transplantation. B220⁺, IgM⁺ and IgD⁺ cell counts were similar for the wild type and SF-coyc groups, whereas the PGK-coyc and γ cPr-coyc mice had B cell counts 2 to 4-fold lower, but prominently increased relative to the GFP control (P<0.001 for wild type and coyc groups).

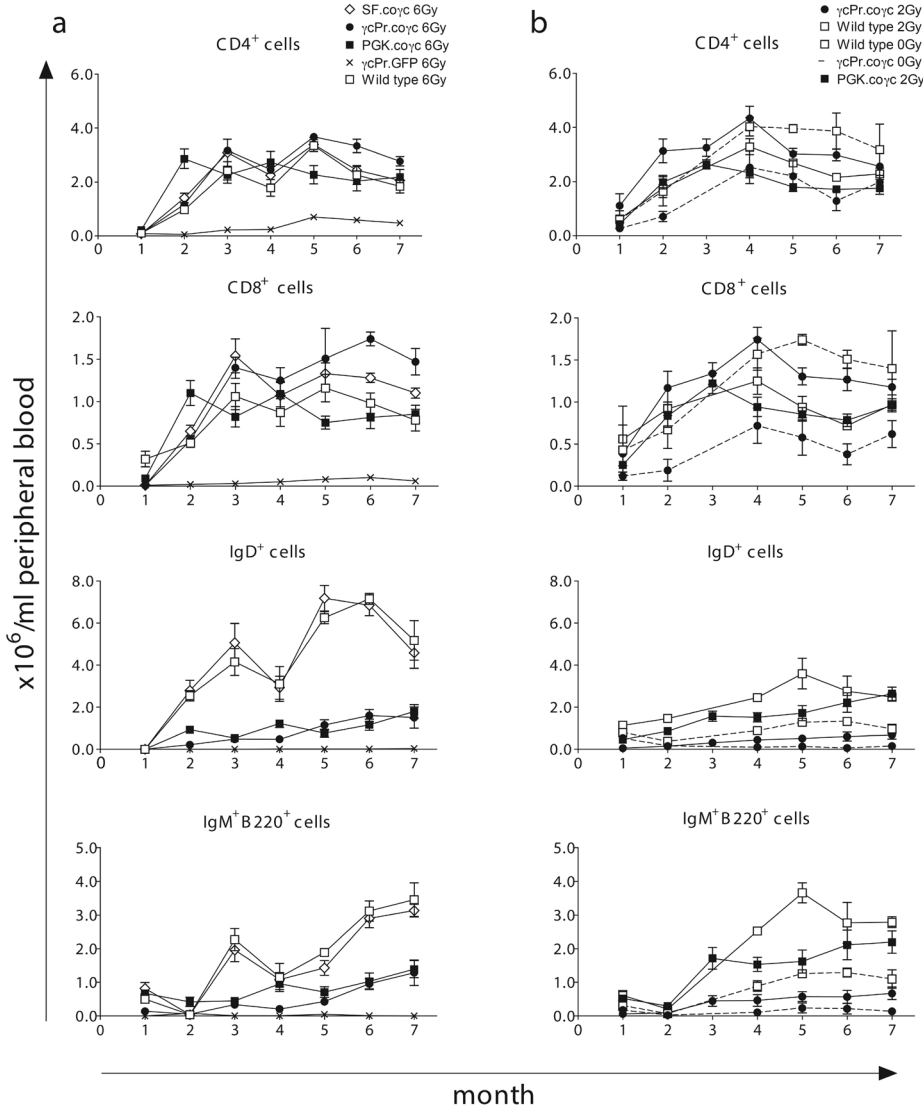


Figure 3. T and B lymphocyte reconstitution in peripheral blood. *Il2rg*^{-/-} mice were transplanted with lentiviral vector transduced wild type or *Il2rg*^{-/-} Lin- cells. Absolute lymphocyte cell numbers in peripheral blood of gene therapy treated mice were determined monthly by flow cytometry. A comparison of all the lentiviral vectors tested is shown in (a): *Il2rg*^{-/-} Lin- cells with γ cPr-GFP, SF-co γ c, γ cPr-co γ c (n = 4 per group) or PGK-co γ c (n = 5), and wild type Lin- cells with γ cPr-GFP (n = 3). Mice were subjected to 6 Gy irradiation and received 4×10^5 cells. T and B lymphocytes are presented as CD4⁺, CD8⁺ or IgD⁺, IgM/B220⁺ cells respectively. (b) represents a comparison of a titration in irradiation conditioning: *Il2rg*^{-/-} mice were subjected to 2Gy irradiation and transplanted with 5×10^5 cells, either wild type Lin- cells (n = 3) or *Il2rg*^{-/-} Lin- cells transduced with γ cPr-co γ c (n = 9) or PGK-co γ c (n = 7). A further five *Il2rg*^{-/-} mice were transplanted with *Il2rg*^{-/-} Lin- cells transduced with γ cPr-co γ c (n = 3) or wild type Lin- cells (n = 2), without pre-transplant irradiation.

To further study the influence of conditioning intensity and cell number on engraftment, 30 *Il2rg*^{-/-} mice were treated with LV- γ Pr-co γ c or LV-PGK-co γ c transduced *Il2rg*^{-/-} cells or wild type cells with the pre-transplant conditioning reduced to 2 Gy (n=13 for γ cPr-co γ c, n=7 for PGK-co γ c and n=3 for wild type) or no conditioning radiation (“0 Gy”) (γ cPr-co γ c n=5 and wild type n=2). Four γ cPr-co γ c treated mice in the 2 Gy group received 10⁵ transduced cells compared to the 5x10⁵ cells given to the other mice. Reconstitution of T and B cells is shown in Figure 3b. Reduction of pre-transplant conditioning from 6 Gy to 2 Gy had little or no effect on the reconstitution of T and B cells in either co γ c treated mice or the wild type controls. Eliminating pre-transplant conditioning altogether resulted in reduced T and B cell counts for γ cPr-co γ c treated mice (P<0.01 for B cells compared to the 2 Gy group), whereas increased T cell counts and reduced B cells seen in wild type 0 Gy controls were lower, but not statistically significant relative to the 2 Gy wild type group. Mice injected with 10⁵ γ cPr-co γ c transduced cells had approximately 25% lower T cell counts relative to mice given 5x10⁵ cells, however, this difference was not statistically significant either (data not shown). Mice were sacrificed at 9 months post-therapy and T and B lymphocyte markers were determined in BM and spleen (Figure 4). Spleen T cell populations were similar for mice treated with co γ c vector transduced cells and recipients of wild type cells. Mice treated with γ cPr-GFP lacked CD8⁺ cells and had reduced levels of CD4⁺ cells. BM of conditioned mice treated with co γ c vectors or wild type cells had similar levels of B lymphocytes, whereas GFP-treated *Il2rg*^{-/-} mice had virtually no B lymphocytes (Figure 4a). Similar results were found in the spleens (Figure 4b).

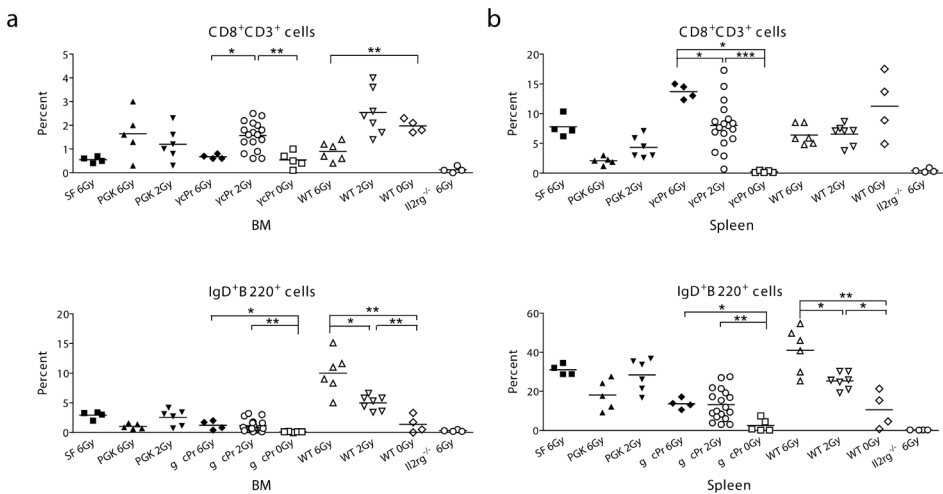


Figure 4. Percentages of CD8⁺ and IgD⁺ cells in spleen and BM. Percentages of CD3/CD8⁺ and IgD/B220⁺ cells in BM (a) and spleens (b) of experimental mice. Mice were subject to 6Gy, 2Gy or 0Gy irradiation and transplanted with 5x10⁵ wild type Lin- cells (n= 17) or 5x10⁵ *Il2rg*^{-/-} Lin- cells transduced with SF-co γ c (n=4), PGK-co γ c (n=11), γ cPr-co γ c (n= 27) or γ cPr-GFP (n=4). All groups consist of mice from 2 separate experiments, excepting the PGK and SF groups. Significant differences are seen in the percentages of IgD/B220⁺ cells between mice given pre-transplant conditioning and those who did not (*P<0.05, **P<0.01).

Mice given wild type or γ cPr-co γ c-treated cells with no pre-transplant irradiation conditioning had significantly lower percentages of IgD/B220⁺ cells in their spleens and BM than mice given similar cells with conditioning. The average numbers of integrations per BM cell corrected for donor/recipient chimerism were 2.2 for the PGK-co γ c mice and 3.5, 3.2 and 2.4 for the γ cPr-co γ c mice given 0 Gy TBI/ 5×10^5 cells, 2 Gy TBI/ 5×10^5 cells and 2 Gy TBI/ 10^5 cells, respectively (Supplementary Table S1b).

To confirm that the γ cPr-co γ c vector could effectively restore T and B cell populations even at a low MOI, another 19 *Il2rg*^{-/-} mice were administered 2 Gy irradiation and treated with 5×10^5 *Il2rg*^{-/-} LV- γ cPr-co γ c transduced cells (n=15) or wild type cells (n=4). Cells transduced with LV- γ cPr-co γ c using a HTU MOI of 10, 3 or 1 were transplanted into *Il2rg*^{-/-} mice (n=5 per group). Reducing the MOI of the γ cPr-co γ c from 10 to 3 had no effect on the efficacy of leukocyte reconstitution, whereas reducing the MOI to 1 resulted in a more protracted course of T cell reconstitution to reach levels of T and B cells similar to the MOI 10 and 3 γ cPr-co γ c groups at 5 months post-therapy (Figure S2). Mice were sacrificed at 8 months post-therapy and the average vector copy per BM cell, corrected for donor/recipient chimerism, was 1.9, 1.0 and 0.2 for the mice given cells transduced with a HTU MOI of 10, 3 and 1, respectively (Supplementary Table S1b).

Humoral immune responses

At 15 weeks post-transplant, peripheral blood plasma was tested for basal immunoglobulin levels. Plasma concentrations of IgM and IgG1 were similar among the wild type group and the two co γ c treated groups (Figure 5a) and clearly distinct from the absent antibodies in GFP-treated *Il2rg*^{-/-} mice ($P < 0.05$). Natural antibodies specific to TNP-KLH²³ were absent in GFP-treated *Il2rg*^{-/-} mice but present in PGK-co γ c mice as well as wild type controls. We conclude that natural antibodies normalized as well in successfully treated mice.

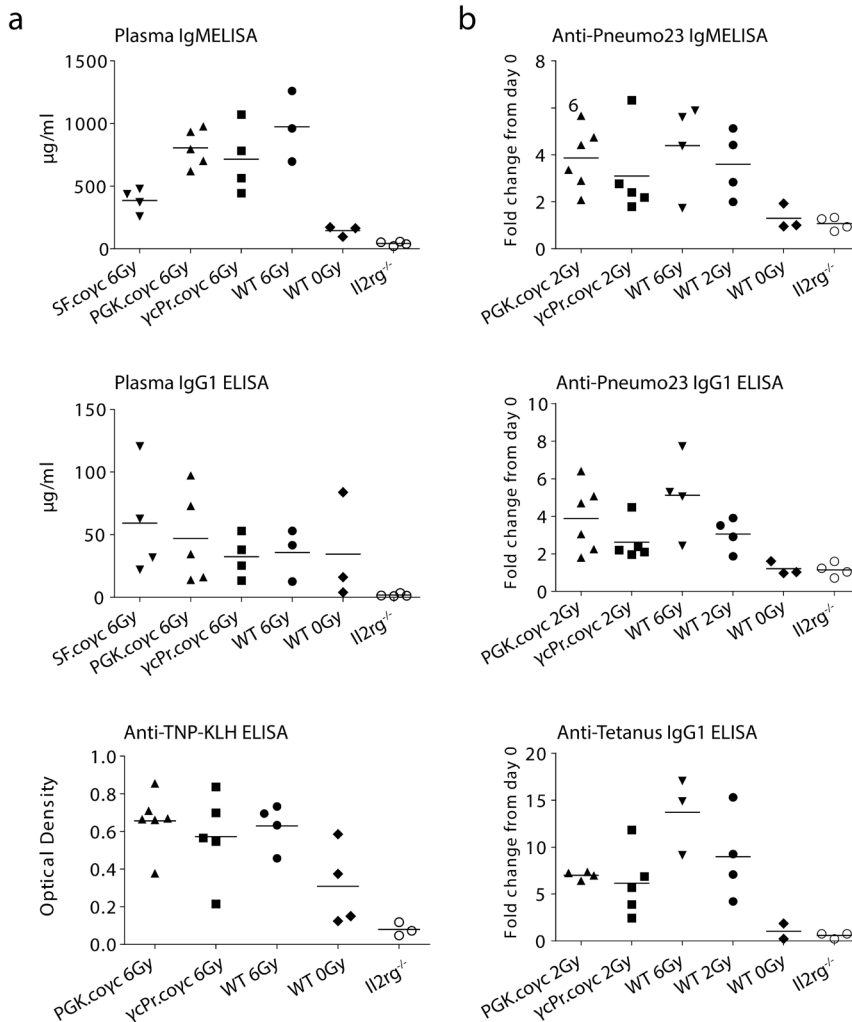


Figure 5. Response of plasma immunoglobulin levels to immunization. (a) plasma immunoglobulin levels quantified for IgM, IgG1 or TNP-KLH. *Il2rg*^{-/-} mice were transplanted with wild type Lin- cells (n=3) or *Il2rg*^{-/-} Lin- cells transduced with γ cPr-coyc (n=4), SF-coyc (n=4), PGK-coyc (n=5) or γ cPr-GFP (n=4, group designated *Il2rg*^{-/-}). Plasma was collected at 15 weeks after transplantation. All coyc-treated groups are significantly different than the GFP control group for levels of plasma IgM, IgG1 and anti-TNP-KLH antibodies ($P < 0.05$). (b) Quantification of Pneu23 and tetanus specific immune responses. *Il2rg*^{-/-} mice transplanted with 5×10^5 *Il2rg*^{-/-} Lin-cells or wild type cells were subjected to a Pneu23 immunization scheme 6 months after transplantation, with plasma collected 10 days later. Shown are fold changes in Pneu23-specific IgM and IgG1 antibody levels before and after immunization. Mice had been transduced with γ cPr-coyc (n=5) or PGK-coyc (n=6) after 2Gy irradiation, or given *Il2rg*^{-/-} Lin- cells with no vector (n=4), or given wild type cells after 6Gy, 2Gy or 0Gy irradiation (n=4 for 6 and 2Gy and n=3 for 0Gy group). All groups except for the WT 0Gy group are significantly different than the *Il2rg*^{-/-} control group ($P < 0.05$). One month later, the surviving mice were subjected to a tetanus toxoid immunization scheme. Tetanus injections were repeated 2 times at 2 week intervals. Shown are plasma IgG1 antibody levels 2 weeks after the last boost. No antibodies to either Pneu23 or tetanus were observed prior to immunization.

To confirm that a specific T cell dependent antibody response was restored in γ Pr-coyc and PGK-coyc mice, a tetanus toxoid immunization scheme was started at four months after transplantation in wild type and γ Pr-coyc recipient mice in the 2 Gy radiation groups (Figure 5b). Mice were injected three times with tetanus toxoid at two-week intervals. Tetanus-specific IgG antibody levels increased over time in all groups with similar antibody responses between wild type and coyc treated mice. The antibody titers were also similar among γ Pr-coyc treated mice receiving 5×10^5 or 1×10^5 Lin⁻ cells. The specific antibody response was absent in plasma of *Il2rg*^{-/-} mice.

A T cell independent antibody test was performed in mice that received wild type, PGK-coyc or γ Pr-coyc-treated cells in the 2 and 0 Gy radiation groups (Figure 5b). Mice were injected with Pneumo23 and plasma was collected after 10 days had passed. IgM and IgG antibody levels in coyc and wild type cell treated mice were clearly normalized relative to the non-responding GFP treated mice. Of note, in the 0 Gy TBI recipients of γ Pr-coyc transduced cells, 3 out of 4 mice immunized failed to produce antibodies, significantly different from all other treated groups ($p=0.001$) and consistent with the severely reduced B cell reconstitution in these mice.

To assess T lymphocyte response to cytokine signaling, the proliferative potential of spleen cells was determined in response to ConA and IL-2. As shown in Figure 6, cells of mice treated with SF, PGK or γ Pr containing coyc vectors showed similar levels of cellular expansion in response to ConA and IL-2. A response was not obtained in spleen cells of *Il2rg*^{-/-} mice treated with GFP ($P<0.01$).

TCR repertoire

Functional T cell receptor (TCR) rearrangement was determined from spleen cDNA with primers specific to the Constant (C) part of the T cell antigen receptor β chain and one of the Variable (V) gene segment-specific primers.²⁴⁻²⁵ A total of 21 V gene segment primers were used to create PCR products, which were then diluted and analyzed on an ABI 3130xl sequencer. Peak patterns for mice treated with coyc vectors or wild type cells were diverse, indicating the existence of a polyclonal T cell population that is able to create a diverse, functional repertoire of antigen receptors (Figure 7). Peak patterns for *Il2rg*^{-/-} mice treated with the GFP control vector tended to have only a few peaks, indicating that the antigen receptor repertoires in T cells in these mice had no or limited functionality.

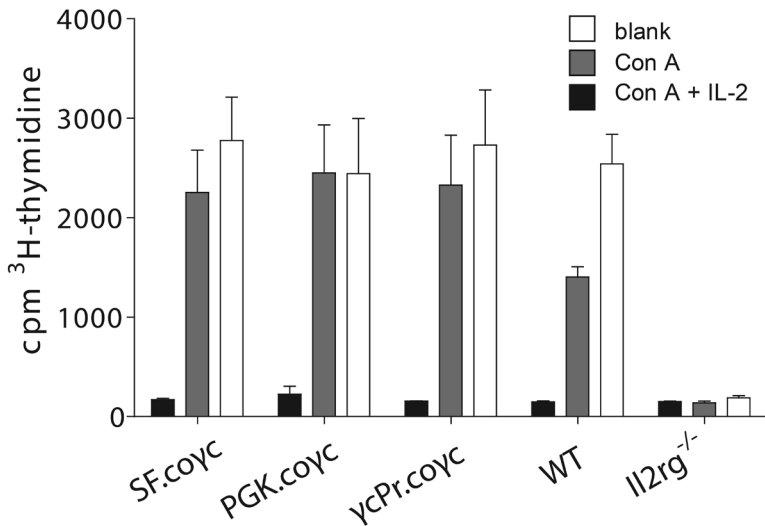


Figure 6. Spleen cell responses to stimulation with ConA/IL-2. Spleen cells were collected from wild type mice, *Il2rg*^{-/-} mice, and *Il2rg*^{-/-} mice transplanted with *Il2rg*^{-/-} Lin⁻ cells transduced with γ cPr-co γ c, SF-co γ c or PGK-co γ c (N = 3 for all groups). Measurements of co γ c groups were adjusted for levels of spleen cell chimerism (Table 2). Wild type and co γ c-treated groups were significantly different compared to *Il2rg*^{-/-} group (**P<0.01) for both Con A and Con A + IL-2 stimulation, whereas the gene therapy treated mice did not significantly differ from the wild type mice.

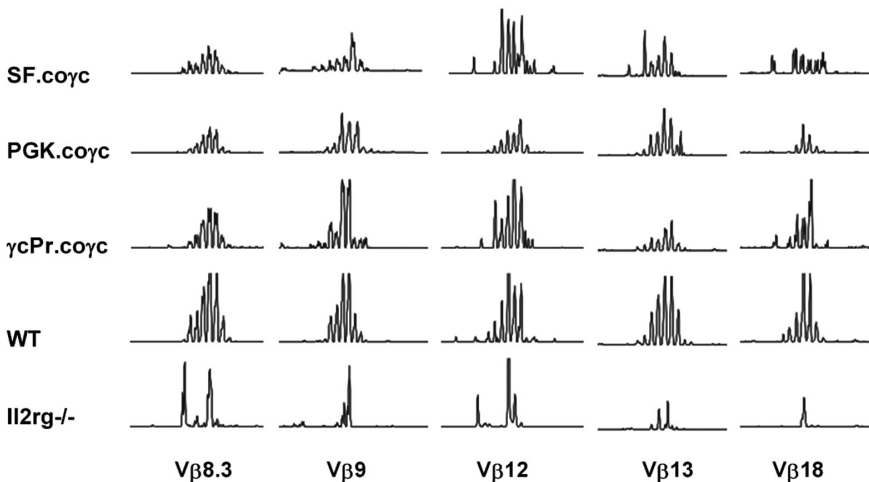


Figure 7. Genescan analysis of murine T cell receptor β (TCR β) repertoire. PCR was performed on cDNA synthesized from total spleen RNA, using a specific primer for the Constant part of the TCR β chain in combination with 21 Variable (V) gene segment-specific primers. Of the 21 V gene segments tested, the most representative samples are shown here. Diverse peak patterns indicate the rearranged TCR β genes are derived from a polyclonal cell population, whereas samples with only a few peaks reflect inability to form a broad T cell receptor repertoire.

Secondary transplants

All mice in these studies were sacrificed 8 to 9 months after transplantation and BM cells were re-transplanted into one or two *Il2rg*^{-/-} mice following 6 Gy irradiation (2x10⁶ total BM cells per mouse). The secondary transplants were monitored monthly by blood counts and flow cytometry as described for up to 8 months. Reconstitution was universally inferior to the primary transplants, with reduced circulating B cell counts (for the γ cPr-co γ c secondary recipients group $0.1 \pm 0.0 \times 10^9$ vs $1.2 \pm 0.2 \times 10^9$ IgM⁺B220⁺ B cells/L and for wild type secondary recipients, $1.9 \pm 0.2 \times 10^9$ vs $4.7 \pm 0.5 \times 10^9$ IgM⁺B220⁺ B cells/L) and poor antibody responses to Pneumo23 and tetanus seen in mice given co γ c-treated or wild type cells (data not shown). Some secondary transplants had undetectable levels of B cells, with this effect most often seen when the primary transplants were treated with physiological promoters or received no pre-transplant conditioning.

NK cell reconstitution

A pilot experiment to obtain insight into NK cell reconstitution following transplantation of cells transduced with the therapeutic γ c gene was performed in *Rag2*^{-/-}/*Il2rg*^{-/-} **C57BL/6 mice, also using the opportunity to compare the LV- γ cPr-co γ c vector to the LV-UCOE-IL2RG construct published earlier.**²⁶ *Il2rg*^{-/-} lin⁻ cells were transduced using an MOI of 15 and of 10 for the γ cPr and UCOE vectors, respectively, and transplanted into *Rag2*^{-/-}/*Il2rg*^{-/-} recipients. Wild type **C57BL/6** mice and untreated *Rag2*^{-/-}/*Il2rg*^{-/-} **C57BL/6** mice were used as controls. Percentages of CD3⁺, B220⁺ and NK1.1⁺ cells were similar for the two vectors, and also similar to those obtained with the the SF and PGK vectors in recipients at a BALB/c background, although the mice treated with the γ cPr vector had a higher number of vector copies per cell (Supplementary Table S2), again stressing that vector copy number/cell has no major impact above a certain threshold level.

DISCUSSION

To address the risk of gammaretroviral vector-mediated insertional mutagenesis in gene therapy of SCID-X1, we generated self-inactivating SIN HIV-1 derived lentiviral vectors¹² containing the bPRE4* element described previously.²⁰ These vectors integrate with high efficiency in hematopoietic stem cells by an overnight transduction without the requirement of pre-stimulation with growth factors and have a reduced likelihood of integrating near transcription start sites, as well as improved resistance to silencing.^{10, 13} Vectors for transfer of the therapeutic transgene were driven by the SF, PGK or γ cPr promoters. Lineage negative *Il2rg*^{-/-} cells transduced with these vectors and transplanted in clinically feasible numbers into *Il2rg*^{-/-} recipient mice proved effective in restoring immune competence, as measured by T cell numbers, polyclonal TCR β gene rearrangements, ConA/IL-2 stimulated T cell proliferation, T cell dependent and independent B cell antibody responses, and by restored levels of natural antibodies. All three vectors proved effective at restoring

T cell populations and immune system functionality to wild type levels, provided a minimum level of pre-transplant conditioning was applied. A series of GFP vectors served as controls, and additionally demonstrated improved selectivity of the γ CPr promoter for expression in white blood cells.

We compared the efficacy of the SF-co γ C, γ CPr-co γ C and PGK-co γ C vectors at reconstituting T and B cell populations in *Il2rg*^{-/-} mice. The levels of T and B cell reconstitution in SF-co γ C vectors, which approach recovery rates of wild type cells, are similar to the initial results of retroviral gene therapy for murine SCID-X1.²⁷ Likewise, the lymphocyte numbers and the results of a spleen cell proliferation assay compare favorably with a more recent approach to murine SCID-X1 gene therapy with lentiviral vectors using a similar MOI.²⁶ The γ CPr-co γ C and PGK-co γ C vectors fully restored T lymphocyte populations in the periphery, BM and spleen, but were less effective in restoring B cell populations. Mice treated with co γ C vectors regained immune function as shown by spleen cell reactivity to proliferation signals and displayed a repertoire of TCR β gene rearrangements. Although the B cell populations were reduced, PGK and γ CPr-co γ C vector treated mice had plasma levels of IgM and IgG1 similar to mice given wild type cells with a comparable pre-transplant conditioning regimen, as well as fully normalized antibody responses to Pneumo23 and tetanus toxoid. The reduced number of circulating B cells in γ CPr-co γ C treated animals apparently does not negatively influence specific antibody production resulting from T cell dependent or from T cell independent responses.

Due to a low vector titer, the MOI of the PGK-co γ C vector was 5-fold lower than the other vectors with a corresponding reduction in the number of integrations per cells. The integrations per cell were also dissimilar for the comparison study between the UCOE and γ CPr vectors, which aimed at measuring NK cell levels. However, we demonstrated that reducing the γ CPr MOI to 0.2 integrations per cell had no significant effect on reconstitution, thus allowing comparisons at different numbers of integrations per cell.

Although full functional restoration was accomplished, the reduced numbers of circulating B lymphocytes in γ CPr-co γ C and PGK-co γ C treated mice compared to those from SF-co γ C and wild type groups is noteworthy. This might be the result of species-specific differences in lymphocyte development: humans with SCID-X1 are able to produce high levels of B cells through non- γ C-dependent pathways,^{1N} whereas *Il2rg*^{-/-} mice do not have this ability. Another possibility is that the lower levels of γ C protein expression driven by γ CPr and PGK relative to that by the SF promoter reduces the selective advantage of B cell precursor cells to repopulate bone marrow, leading to lower circulating levels and/or proliferation. This may also underline the importance of pre-transplant conditioning.

Reduction of the pre-transplant conditioning of recipient mice from 6 Gy to 2 Gy had no significant effect on the efficacy of the gene therapy treatment, consistent with the principles of conditioning for autologous stem cell engraftment earlier established in thalassemic mice.²⁸ However, also in line with the latter observations, eliminating the pre-transplant conditioning altogether severely reduced the efficacy of the treatment in

mice transplanted with either γ CPr-co γ C treated or wild type cells, with a corresponding drop in donor cell BM chimerism, percentage of lymphocytes present in spleen and BM, and failures to produce antibodies upon immunization. *Il2rg*^{-/-} mice that received wild type cells but were not pre-conditioned resulted in low to no detectable B cells levels in blood BM and spleen, and many of these mice had poor recovery of plasma antibody levels and specific immune responses. This is similar to conditions applied in the first gammaretroviral vector SCID-X1 trials,⁴ resulting in some of the patients requiring immunoglobulin-replacement therapy. To our knowledge, this is the first paper underlining the importance of pre-transplant conditioning for murine SCID-X1. Extrapolation of this observation to human SCID-X1 gene therapy predicts improved immune reconstitution by applying mild conditioning and a reduced threshold of cell numbers required, thereby improving both efficacy and safety.

Titration of the LV- γ CPr-co γ C vector demonstrated convincingly that lowering the vector dose to reach the preferred 1 copy per cell for clinical application had no significant effect on the reconstitution of T and B cell lineages, in contrast to results previously described by Zhou *et al.*²⁹ Using an HTU MOI of 1 resulted in on average 0.2 vector copies per BM cell *in vivo*, compared to 1.9 copies/BM cell in a mouse given an MOI of 10, with no significant difference in the T and B cell counts seen in these animals at 6 months post-therapy. Likewise, reduction of the Lin- cell number transplanted to 10⁵/mouse (4x10⁶/kg, equivalent to 4x10⁷/kg unfractionated BM cells or roughly 10⁶/kg CD34⁺ cells in humans) did not noticeably influence immune reconstitution.

In this manuscript, we also provide evidence that the codon optimization proved essential for improved viral vector titers, transgene transcript levels and protein production as has been shown before in other vectors,¹⁵ allowing immune reconstitution at a level of one transgene copy per cell. This is in contrast to a recent report in which low copy numbers per cell of a vector with a 233-bp human EF1 α promoter resulted in significantly reduced B and NK cell reconstitution, and IgG plasma levels.²⁹

The use of strong retroviral enhancer/promoter sequences is not desirable for clinical application since the leukemia cases that occurred in the initial gene therapy trials were caused by integrations in vulnerable regions near proto-oncogenes resulting in subsequent upregulation and constitutive expression.⁶⁻⁷ Furthermore, viral promoters are at risk of being silenced by methylation in mammalian cells.³⁰⁻³³ The SF promoter in particular has been shown to modify gene expression in regions greater than 100 kb,³⁴ reducing the potential of activating neighboring genes by cellular promoters¹⁸ may be beneficial for clinical application. In a recent study, Zhou *et al.*²⁹ tried to address whether a 233-bp human EF1 α promoter was able to upregulate *LMO2* with a *LMO2* activation assay in human leukemic Jurkat T-cells when inserted in the *LMO2* locus, and compared to a γ C genomic sequence inserted in the reverse orientation. However, lentiviral vectors integrate throughout genes, not at fixed locations or orientations, and have shown no preference to integrate near oncogenes or near transcription start sites. In the present study,

we encountered a significant number of integrations of the SF vector in the immediate vicinity of RTCGD listed proto-oncogenes, but not with the PGK or γ cPr vectors (Supplementary Table 3).

Constitutive *IL2RG* expression has been postulated to pose an increased risk of oncogenesis.³⁵ However, this implication of *IL2RG* acting as an oncogene under certain circumstances is controversial,³⁶⁻³⁷ since aberrant constitutive expression of *LMO2* results in a differentiation block of developing T cell precursors under proliferative pressure,³⁸ which by itself would be sufficient to explain the onset of leukemia after a secondary oncogenic event. In this scenario, *IL2RG* expression is permissive, not causative. More recent evidence supports the notion that *IL2RG* and *LMO2* require additional cooperating mutations.³⁹ In addition, a novel model of SCID-X1 reconstitution suggests that the SCID-X1 phenotype itself causes cells to be more susceptible to insertional leukemogenesis by murine leukemia viruses or gammaretroviral vectors.⁴⁰ Of note, no insertions were detected in the *Lmo2* locus after thorough integration analysis of the leukemic clones obtained in the course of this study.

The study involved 70 primary recipients, of which bone marrow cells have been retransplanted eight months after transplantation into secondary recipients to increase the proliferative pressure on the transduced cell population. Secondary recipients had poor lymphocyte reconstitution, particularly in the B cell compartments. Current observation times of primary and secondary recipients are at least 15 months after transplantation. While we have detected leukemias and T cell malignancies in some mice, none of these adverse events were shown to be vector-derived after qPCR and LAM-PCR analysis and as such were concluded to be background malignancies inherent to the radiation exposure, mouse strain or other factors in the procedure.⁴¹ The mice and their retransplants will be monitored life-long in the context of a larger study directed at potential adverse effects and at a full integration analysis, to be reported separately. Some mice treated with similar vectors in experiments not included in this manuscript have developed leukemias that are likely vector derived, which will be reported on separately.

We conclude that the γ cPr-co γ c vector, as well as the PGK-co γ c vector, provides an attractive, required alternative to replace constitutive viral promoters, allowing the development of efficacious therapeutic lentiviral vector mediated gene transfer for the correction of SCID-X1 with a potentially improved safety profile both by the use of the SIN lentiviral backbone⁴²⁻⁴³ and as judged from an initial integration analysis. Additionally, our study underscores the importance of pre-transplant conditioning for consistent successful correction, which should have an impact on current clinical protocol development.

METHODS

Mice

BALB/c and *Il2rg*^{-/-} mice were bred in the Experimental Animal Center of Erasmus MC. *Il2rg*^{-/-} single KO mice with a targeted γ c deletion were derived from a 10th generation backcross of BALB/c *Rag2*^{-/-}/*γc*^{-/-} mice, kindly provided by Dr. H Spits,⁴⁴ with syngenic BALB/c wild type mice. All mice were used at 6 to 10 weeks of age and were maintained in specified pathogen free conditions. Experiments were approved by the institutional Animal Ethical Committee of Erasmus MC in accordance with legislation in the Netherlands. Comparison studies between the LV- γ cPr-co γ c vector described here and a LV-UCOE vector were performed using *Il2rg*^{-/-} donor mice and *Rag2*^{-/-}/*Il2rg*^{-/-} recipients, plus wild type controls (**C57BL/6** background). These experiments were conducted at University College, London, according to their protocols,²⁶ with approval in accordance with legislation in the UK.

Lentiviral vector plasmids

Third generation self-inactivating (SIN) LV incorporating the γ c cDNA were constructed using the HIV-1 vector backbone⁴²⁻⁴³ with SF promoter and a modified woodchuck posttranslational regulatory element (bPRE4*) as described previously,²⁰ but now incorporating the γ c cDNA. The native human *IL2RG* cDNA was cloned from a LZRS-IL-2RG-IRES-EGFP retroviral plasmid,⁴⁴ kindly provided by Dr. Karin Pike-Overzet, Dept of Immunology, Erasmus MC, Rotterdam, the Netherlands, and cloned into pRRL.PPT.SF.EGFP.WPRE4*.SIN (LV-SF-GFP) to create LV-SF- γ c. Codon-optimized human *IL2RG* sequence (co γ c) predicted by GeneOptimizer[®] software (GeneArt AG, Regensburg, Germany) was cloned into LV-SF- γ c to generate the LV-SF-co γ c vector.

The human *IL2RG* promoter fragment (γ cPr) was obtained by PCR amplification of a ~1.1kb human DNA region¹⁷ using the following primers: 5'-CTCGAGAGGATGTCTTGTTGGTCT-3', and 5'-GGATCCCGCTTGCTCTTCATTCCC-3' with restriction sites underlined (Eurogentec, the Netherlands). Post-PCR, the fragment was ligated into **pCR-TOPO-TA** vector (Invitrogen, Carlsbad, CA) and verified by restriction digest and sequencing. The γ cPr was subsequently ligated into a GFP LV in place of the PGK promoter by digestion with XhoI and BamHI to create LV- γ cPr-GFP. Similarly, γ cPr and PGK were cloned into SF-co γ c to produce LV- γ cPr-co γ c and LV-PGK-co γ c.

Production of lentiviral particles and titration

LVs were produced by standard calcium phosphate transfection of HEK 293T cells¹⁹ with the packaging plasmids pMDL-g/pRRE, pMD2-VSVg and pRSV-Rev. LV were concentrated by ultracentrifugation at 20,000 r.p.m. for 2 hours at 4°C and stored at -80°C. Viral vector titration was performed on HeLa cells by serial dilutions to determine transducing units (HeLa transducing units, HTU) per mL. The transduced cells were harvested on day

11, DNA and total RNA were purified and the copy number per cell was determined by real-time quantitative polymerase chain reaction (qPCR).

Real-time quantitative polymerase chain reactions

qPCR to quantify the integrated proviral copy number⁴⁵ was performed on the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) with 100 ng genomic DNA template and SYBR Green PCR Master Mix (Applied Biosystems) with HIV primer, 5'-CTGGAAGGGCTAATTCACCTC-3', and 5'-GGTTTCCCTTTC-GCTTTCAG-3' resulting in a 274 bp fragment of the proviral DNA. The PCR conditions were 95 °C for 10 min and 45 cycles as follows: 95°C for 15 s and 62°C for 1 min. A standard calibration curve was determined from flow sorted mouse 3T3 cells transduced with LV-SF-GFP at a multiplicity of infection (MOI) of 0.06 containing one integration per genome. All PCR reactions were performed in triplicate and analyzed with SDS2.2.2 software.

Determination of WPRE mRNA levels

Total RNA of transduced HeLa cells was converted to cDNA via the QuantiTect Reverse Transcription kit (Qiagen, Venlo, the Netherlands). qPCR reactions were performed using 5ng cDNA and primer 5'-ATTGCCACCACCTGTCAACT-3' and 5'-CCGACAACAC-CACGGAATTA-3' to amplify a fragment of the WPRE. qPCR reactions were performed on the ABI Prism 7900HT using a serial dilution of SF-IL2RG plasmid as the standard curve and a 55 °C annealing temperature.

Chimerism assay

qPCR reactions were performed on spleen and BM DNA from *Il2rg*^{-/-} mice transplanted with male donor cells transduced with LV vectors to amplify a region of the Y chromosome as described by Pujal et al⁴⁶, using 100ng of genomic DNA and primers 5'-TCATC-GGAGGGCTAAAGTGTCAC-3' and 5'-TGGCATGTGGGTTCTGTCC-3'. A standard curve was generated using spleen and BM DNA from male BALB/c mice.

LM-PCR and LAM-PCR

High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR) and ligation-mediated PCR (LM-PCR) was performed on BM, peripheral blood or spleen DNA from *Il2rg*^{-/-} mice transduced with LV vectors. For LAM-PCR⁴⁷ restriction enzymes Tsp509I or MseI were used with the lentiviral (HIV) primer set. PCR products were run on high resolution polyacrylamide (Spreadex) gel. To the nested primers of the LM-PCR⁴⁸⁻⁴⁹ protocol, the following primers were added to create the sequencing tags: 3'-GCCTCCCTCGCGCCATCAG-5'+ MID (multiplex identifier) to the primer annealing to the viral LTR, and 3'-GCCTTGCCAGCCCGCTCAG-5'+ MID to the primer annealing to the linker.

Transduction and transplantation of lineage negative BM cells

BM cells from male *Il2rg*^{-/-} and congenic wild type mice were purified by lineage depletion (Lin⁻) (BD Biosciences, Santa Clara, CA). Lin⁻ BM cells were transduced overnight at 10⁶ cells/ml in serum free modified Dulbecco's medium with supplements⁵⁰ in the presence of murine stem cell factor (mSCF) 100 ng/mL, human FMS-like tyrosine kinase 3-ligand (hFlt3-L) 50ng/ml and murine thrombopoietin (mTPO) 10 ng/ml at HTU MOI 9, except for the PGK-coy γ c vector (MOI 2) and a vector dose titration of γ cPr-coy γ c from MOI 10 to 3 to 1. Subsequently, 5 \times 10⁵, 4 \times 10⁵ or 10⁵ cells were injected into the tail vein of 6 Gy or 2 Gy irradiated female *Il2rg*^{-/-} recipients or without conditioning.

Immunophenotyping by flow cytometry

Flow cytometric analyses were performed on cells obtained from blood, bone marrow and spleen. Peripheral blood was collected monthly in EDTA tubes by retro-orbital puncture under isoflurane anesthesia. Complete blood cell counts were measured using a Vet ABC hematology analyzer (Scil animal care company GmbH, Germany). Blood was lysed and leukocytes were washed three times with Hank's balanced salt solution (HBSS, Invitrogen) containing 0.5% (wt/vol) bovine serum albumin and 0.05% (wt/vol) sodium azide (HBN). Cells were incubated for 30 minutes at 4°C in HBN containing 2% heat-inactivated normal mouse serum and antibodies against CD3, CD4, CD8, B220, CD19, IgM, IgD, CD11b Sca-1, C-kit and T cell receptor β (TCR β) directly conjugated to R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP) or allophycocyanin (APC; all antibodies BD Biosciences). Subsequently, cells were washed and measured on a FACSCalibur (Becton Dickinson). BM and spleen cells were evaluated similarly. Additionally, GFP expression was measured in mice treated with the LV-SF-GFP or LV- γ cPr-GFP vectors.

Determination of surface γ c expression

Lineage negative *Il2rg*^{-/-} BM cells were purified as described above and transduced with LV-SF- γ c or LV-SF-coy γ c vectors at a HeLa MOI of 100, 10 or 1. Transduced cells were cultured for 4 days, washed and then surface γ c expression was determined via flow cytometry. A CD132-pe antibody (BD Biosciences) was used to determine γ c expression.

Immunizations and ELISAs

HeLa cells were transduced with a fixed MOI for either LV-SF- γ c or LV-SF-coy γ c and kept in culture for 7 days. Copy number per cell was confirmed by qPCR. An ELISA was performed on extracts from these transduced cells coated on the plates, which were normalized for protein content. Biotinylated goat anti-human common gamma chain (BAF284, R&D Systems) was applied to these plates, and signal was detected with streptavidine-peroxidase. The relative fold increase in γ c expression was determined, adjusted for copy number per cell.

An ELISA was done on peripheral blood plasma collected 15 weeks after transplantation

to determine basal IgM and IgG1 plasma levels. Briefly, Covalink 96-well plates (Nunc A/S, Roskilde, Denmark) were coated first with DSP (dithiobis[succinimidyl propionate], Pierce Biotechnology, Rockford, IL) in methanol and then with unlabeled (capture) anti-mouse IgM (1B4B1) or IgG1 (H143.225.8) antibodies (SouthernBiotech, Birmingham, AL) for two hours. The wells were blocked overnight at 4°C with 1% bovine serum albumin in PBS. The next day the plates were washed six times with PBS/0.05% Tween 20 and incubated with serially diluted serum for 1 hour at room temperature (RT). Plates were washed and incubated with a goat anti-mouse horseradish peroxidase-conjugated antibody at RT for 1 hour. After subsequent washing the wells were stained with TMB substrate (Kirkegaard and Perry Laboratories, Gaithersburg, MD) and the reaction was stopped after 10 minutes by adding 1M phosphoric acid. The absorbance of each plate was then read at 405nm using a FLUOstar Optima plate reader. Antibody concentrations were calculated by using purified IgM (M5909) and IgG (I5381) antibody standards (Sigma, Saint Louis, MO).

An ELISA for natural antibodies to trinitrophenol keyhole limpet hemocyanin (TNP-KLH) was done on peripheral blood taken from PGK-coyc-treated mice as well as mice given WT and *Il2rg*^{-/-} cells. The ELISA was performed as described, using TNP(17)-KLH (T-5060-5, Biosearch Technologies, Novato, CA) and anti-KLH (12B4.G3.A8) antibody (AbCam, Cambridge, United Kingdom). Optical density was used to measure the antibodies.

T cell dependent specific antibody responses were determined by intraperitoneal immunization with 16 IU of tetanus toxoid. This process was repeated twice at two week intervals, and plasma was collected two weeks after the last immunization. An ELISA was performed on plates coated with tetanus toxoid, and HRP goat-anti-mouse IgG1 (InVitrogen) was used to measure signal. Anti-tetanus antibody TetE3 was used to make a standard line (AbCam).

T cell independent responses were obtained by subcutaneous immunization with 0.5µg of each 23 purified pneumococcal capsular antigens (PNEUMO 23, Sanofi Pasteur MSD, Brussels, Belgium), and plasma was collected 10 days later. ELISA plates were coated with Pneumo23 and goat-anti-mouse IgM or goat-anti-mouse IgG1 were used to measure signal (Invitrogen). Optical density was measured of pre-immunization and post-immunization plasma to determine the antibody response.

In vitro spleen cell proliferation assay

Spleen cells were cultured at 1×10^6 cells/mL in DMEM supplemented with 10% FCS and antibiotics (pen/strep). Concanavalin A (Con A, 2.5µg/ml, Sigma) was added with or without human interleukin-2 (IL-2, 500 units/ml, Biogen Idec, Badhoevedorp, The Netherlands). On day 3, 0.5µCi ³H-thymidine was added to each well. On day 4 cells were washed and ³H-thymidine incorporation was measured on a TopCount microplate scintillation counter (GMI, Ramsey, MN).

Genescan

Genescan analysis of murine T cell receptor β (TCR β) repertoire was performed cDNA derived from on 2 μ g total RNA purified from fresh spleen cells using the Allprep DNA/RNA kit (Qiagen). cDNA was synthesized using the QuantiTect Reverse Transcription kit (Qiagen). PCR was performed on the cDNA using a specific primer for the Constant (C) part of the T cell antigen receptor β chain (5'-FAM-CTTGGGTGGAGTCA-CATTCTC-3') in combination with 21 Variable (V) gene segment-specific primers to form PCR products between 140 and 240bp as described.²⁵ All PCR products were analyzed on an ABI 3100 or 3130xl sequencer.

Statistics

Statistical analysis was performed using SPSS 11 and Graphpad Prism. Significance of difference was determined by a Mann-Whitney test, for categorical data by Fisher's Exact test. Data are presented \pm SEM where applicable.

ACKNOWLEDGMENTS

The authors thank Luigi Naldini for the third generation self-inactivating lentiviral vector, and recognize the assistance of Karin Pike-Overzet, Mark Rodijk and Frank J. T. Staal under the guidance of Jacques J.M. van Dongen in implementing the Gene Scan analysis as described by Pannetier et al.²⁴⁻²⁵ and in providing the LZRS-IL2RG-IRES-EGFP retroviral plasmid, and are grateful to Roya Sarwari for technical assistance and Elnaz Farahbakhshian for providing the Y chromosome primers. Funding was provided by the European Commission (5th, 6th and 7th Framework Programs, Contracts QLK3-CT-2001-00427-INHERINET, LSHB-CT-2004-005242-CONCERT, and LSHB-CT-2006-19038-Magselec-tofection), The Netherlands Health Research and Development Organization ZonMw (Translational Gene Therapy Research program, project 43100016), the Wellcome Trust (A.J.T, MB) and the UK Biotechnological and Biological Sciences Research Council (FZ).

Authorship contributions

Contribution: M.W.H., N.P.T., M.H.B., S.A., C.C., Y.L., K.K. F.Z. M.P.B., and T.P.V. performed experiments and analysis; A.S. contributed to vector design; M.W.H. analyzed results and made the figures; M.W.H., N.P.T., M.M.A.V., and G.W. designed the research; M.S., C.K., C.B., A.J.T. and F.M. supervised analyses, reviewed the manuscript and contributed to the text; and M.W.H., N.P.T., and G.W. wrote the paper.

Conflict-of-interest disclosure

The authors declare no competing financial interests.

REFERENCES

1. Noguchi M, Yi H, Rosenblatt HM, Filipovich AH, Adelstein S, Modi WS, *et al.* (1993). Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* **73**: 147-157.
2. Fischer A, Griscelli C, Friedrich W, Kubanek B, Levinsky R, Morgan G, *et al.* (1986). Bone-marrow transplantation for immunodeficiencies and osteopetrosis: European survey, 1968-1985. *Lancet* **2**: 1080-1084.
3. Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J, *et al.* (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* **364**: 2181-2187.
4. Hacein-Bey-Abina S, Hauer J, Lim A, Picard C, Wang GP, Berry CC, *et al.* (2010). Efficacy of gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **363**: 355-364.
5. Bushman FD. (2007). Retroviral integration and human gene therapy. *J Clin Invest* **117**: 2083-2086.
6. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, *et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**: 3132-3142.
7. Howe SJ, Mansour MR, Schwarzwaelder K, Bartholomae C, Hubank M, Kempinski H, *et al.* (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest* **118**: 3143-3150.
8. Bartholomew C, Ihle JN. (1991). Retroviral insertions 90 kilobases proximal to the Evi-1 myeloid transforming gene activate transcription from the normal promoter. *Mol Cell Biol* **11**: 1820-1828.
9. Modlich U, Bohne J, Schmidt M, von Kalle C, Knoss S, Schambach A, *et al.* (2006). Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity. *Blood* **108**: 2545-2553.
10. Montini E, Cesana D, Schmidt M, Sanvito F, Ponzoni M, Bartholomae 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.
11. Schambach A, Bohne J, Chandra S, Will E, Margison GP, Williams DA, *et al.* (2006). Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells. *Mol Ther* **13**: 391-400.
12. Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, Naldini L, *et al.* (1998). Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol* **72**: 9873-9880.
13. Pfeifer A, Ikawa M, Dayn Y, Verma IM. (2002). Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos. *Proc Natl Acad Sci U S A* **99**: 2140-2145.
14. Deml L, Bojak A, Steck S, Graf M, Wild J, Schirmbeck R, *et al.* (2001). Multiple effects of codon usage optimization on expression and immunogenicity of DNA candidate vaccines encoding the human immunodeficiency virus type 1 Gag protein. *J Virol* **75**: 10991-11001.
15. Moreno-Carranza B, Gentsch M, Stein S, Schambach A, Santilli G, Rudolf E, *et al.* (2009). Transgene optimization significantly improves SIN vector titers, gp91phox expression and reconstitution of superoxide production in X-CGD cells. *Gene Ther* **16**: 111-118.
16. Hannan GN, Lehnert SA, MacAvoy ES, Jennings PA, Molloy PL. (1993). An engineered PGK promoter and lac operator-repressor system for the regulation of gene expression in mammalian cells. *Gene* **130**: 233-239.
17. Markiewicz S, Bosselut R, Le Deist F, de Villartay JP, Hivroz C, Ghysdael J, *et al.* (1996). Tissue-specific activity of the gammac chain gene promoter depends upon an Ets binding site and is regulated by GA-binding protein. *J Biol Chem* **271**: 14849-14855.
18. Zychlinski D, Schambach A, Modlich U, Maetzig T, Meyer J, Grassman E, *et al.* (2008). Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Mol Ther* **16**: 718-725.
19. Follenzi A, Naldini L. (2002). HIV-based vectors. Preparation and use. *Methods Mol Med* **69**: 259-274.

20. van Til NP, Stok M, Aerts Kaya FS, de Waard MC, Farahbakhshian E, Visser TP, *et al.* (2010). Lentiviral gene therapy of murine hematopoietic stem cells ameliorates the Pompe disease phenotype. *Blood* **115**: 5329-5337.
21. Baum C, Hegewisch-Becker S, Eckert HG, Stocking C, Ostertag W. (1995). Novel retroviral vectors for efficient expression of the multidrug resistance (mdr-1) gene in early hematopoietic cells. *J Virol* **69**: 7541-7547.
22. Akagi K, Suzuki T, Stephens RM, Jenkins NA, Copeland NG. (2004). RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res* **32**: D523-527.
23. Korver K, Zeijlemaker WP, Schellekens PT, Vossen JM. (1984). Measurement of primary in vivo IgM- and IgG-antibody response to KLH in humans: implications of pre-immune IgM binding in antigen-specific ELISA. *J Immunol Methods* **74**: 241-251.
24. Pannetier C, Cochet M, Darche S, Kourilsky P, inventors. Process for determining the quantity of a DNA fragment of interest by a method of enzymatic amplification of DNA. US patent 5635354, 1998.
25. Pannetier C, Cochet M, Darche S, Casrouge A, Zoller M, Kourilsky P. (1993). The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc Natl Acad Sci U S A* **90**: 4319-4323.
26. Zhang F, Thornhill SI, Howe SJ, Ulaganathan M, Schambach A, Sinclair J, *et al.* (2007). Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. *Blood* **110**: 1448-1457.
27. Lo M, Bloom ML, Imada K, Berg M, Bollenbacher JM, Bloom ET, *et al.* (1999). Restoration of lymphoid populations in a murine model of X-linked severe combined immunodeficiency by a gene-therapy approach. *Blood* **94**: 3027-3036.
28. Wagemaker G, Visser TP, van Bekkum DW. (1986). Cure of murine thalassemia by bone marrow transplantation without eradication of endogenous stem cells. *Transplantation* **42**: 248-251.
29. Zhou S, Mody D, DeRavin SS, Hauer J, Lu T, Ma Z, *et al.* (2010). A self-inactivating lentiviral vector for SCID-X1 gene therapy that does not activate LMO2 expression in human T cells. *Blood* **116**: 900-908.
30. Challita PM, Kohn DB. (1994). Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. *Proc Natl Acad Sci U S A* **91**: 2567-2571.
31. Hoeben RC, Migchielsen AA, van der Jagt RC, van Ormondt H, van der Eb AJ. (1991). Inactivation of the Moloney murine leukemia virus long terminal repeat in murine fibroblast cell lines is associated with methylation and dependent on its chromosomal position. *J Virol* **65**: 904-912.
32. Klug CA, Cheshier S, Weissman IL. (2000). Inactivation of a GFP retrovirus occurs at multiple levels in long-term repopulating stem cells and their differentiated progeny. *Blood* **96**: 894-901.
33. Pikaart MJ, Recillas-Targa F, Felsenfeld G. (1998). Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* **12**: 2852-2862.
34. Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, *et al.* (2005). Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking. *Science* **308**: 1171-1174.
35. Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM. (2006). Gene therapy: therapeutic gene causing lymphoma. *Nature* **440**: 1123.
36. Pike-Overzet K, de Ridder D, Weerkamp F, Baert MR, Verstegen MM, Brugman MH, *et al.* (2006). Gene therapy: is IL2RG oncogenic in T-cell development? *Nature* **443**: E5; discussion E6-7.
37. Thrasher AJ, Gaspar HB, Baum C, Modlich U, Schambach A, Candotti F, *et al.* (2006). Gene therapy: X-SCID transgene leukaemogenicity. *Nature* **443**: E5-6; discussion E6-7.
38. Pike-Overzet K, de Ridder D, Weerkamp F, Baert MR, Verstegen MM, Brugman MH, *et al.* (2007). Ectopic retroviral expression of LMO2, but not IL2Rgamma, blocks human T-cell development from CD34+ cells: implications for leukemogenesis in gene therapy. *Leukemia* **21**: 754-763.
39. Dave UP, Akagi K, Tripathi R, Cleveland SM, Thompson MA, Yi M, *et al.* (2009). Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy. *PLoS Genet* **5**: e1000491.

40. Scobie L, Hector RD, Grant L, Bell M, Nielsen AA, Meikle S, *et al.* (2009). A novel model of SCID-X1 reconstitution reveals predisposition to retrovirus-induced lymphoma but no evidence of gammaC gene oncogenicity. *Mol Ther* **17**: 1031-1038.
41. Ginn SL, Liao SH, Dane AP, Hu M, Hyman J, Finnie JW, *et al.* (2010). Lymphomagenesis in SCID-X1 mice following lentivirus-mediated phenotype correction independent of insertional mutagenesis and gammac overexpression. *Mol Ther* **18**: 965-976.
42. Ailles LE, Naldini L. (2002). HIV-1-derived lentiviral vectors. *Curr Top Microbiol Immunol* **261**: 31-52.
43. Follenzi A, Naldini L. (2002). Generation of HIV-1 derived lentiviral vectors. *Methods Enzymol* **346**: 454-465.
44. Gimeno R, Weijer K, Voordouw A, Uittenbogaart CH, Legrand N, Alves NL, *et al.* (2004). Monitoring the effect of gene silencing by RNA interference in human CD34+ cells injected into newborn RAG2-/- gammac-/- mice: functional inactivation of p53 in developing T cells. *Blood* **104**: 3886-3893.
45. van Til NP, Markusic DM, van der Rijt R, Kunne C, Hiralall JK, Vreeling H, *et al.* (2005). Kupffer cells and not liver sinusoidal endothelial cells prevent lentiviral transduction of hepatocytes. *Mol Ther* **11**: 26-34.
46. Pujal JM, Gallardo D. (2008). PCR-based methodology for molecular microchimerism detection and quantification. *Exp Biol Med (Maywood)* **233**: 1161-1170.
47. Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K, Ball C, Pilz I, *et al.* (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* **4**: 1051-1057.
48. Kustikova OS, Modlich U, Fehse B. (2009). Retroviral insertion site analysis in dominant haematopoietic clones. *Methods Mol Biol* **506**: 373-390.
49. Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD. (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. *Science* **246**: 810-813.
50. Wognum AW, Visser TP, Peters K, Bierhuizen MF, Wagemaker G. (2000). Stimulation of mouse bone marrow cells with kit ligand, FLT3 ligand, and thrombopoietin leads to efficient retrovirus-mediated gene transfer to stem cells, whereas interleukin 3 and interleukin 11 reduce transduction of short- and long-term repopulating cells. *Hum Gene Ther* **11**: 2129-2141.

SUPPLEMENTARY MATERIAL

Table S1A. Averages of chimerism and proviral copies per BM or spleen cell.

Vector	HeLa MOI	Spleen cell chimerism (%)	BM cell chimerism (%)	Avg integrations per donor BM cell
SF-coγc (n=4)	9	41.2 ± 4.4	58.6 ± 9.6	2.1 ± 0.3
γcPr-coγc (n=4)	9	27.8 ± 8.0	65.2 ± 2.9	2.0 ± 0.4
PGK-coγc (n=5)	2	46.9 ± 15.8	26.6 ± 2.3	0.9 ± 0.3
γcPr-GFP (n=4)	9	30.1 ± 18.4	44.2 ± 1.7	4.0 ± 2.4
γcPr-GFP (wild type, n=3)	9	83.7 ± 4.2	75.3 ± 4.3	4.9 ± 0.8

Table S1B. Averages of chimerism and proviral copies per BM cell in *Il2rg*^{-/-} mice treated with γcPr-coγc.

n	HeLa MOI	Radiation dose	Lin- cell dose	BM cell chimerism (%)	Integrations per donor BM cell	BM cells (x 10 ⁶)	Spleen cells (x 10 ⁶)
5	9	0Gy	5*10 ⁵	2.5 ± 0.9	3.5 ± 0.8	29.6 ± 3.1	186.5 ± 32.6
9	9	2Gy	5*10 ⁵	27.3 ± 8.2	3.2 ± 0.6	19.8 ± 3.8	136.2 ± 8.8
4	9	2Gy	1*10 ⁵	18.3 ± 7.1	2.4 ± 0.7	22.2 ± 3.0	177.2 ± 38.8
5	10	2Gy	5*10 ⁵	32.1 ± 12.2	1.9 ± 0.7	24.7 ± 3.6	174 ± 52.3
5	3	2Gy	5*10 ⁵	43.2 ± 22.4	1.0 ± 0.3	25.6 ± 2.7	145.9 ± 44.0
5	1	2Gy	5*10 ⁵	27.2 ± 11.9	0.2 ± 0.1	20.8 ± 3.3	99.4 ± 17.4
					WT	30.3 ± 3.4	202.9 ± 25.6
					<i>Il2rg</i> ^{-/-}	9.4 ± 0.9	57.0 ± 14.4

Percentage of male donor cells in BM and spleen of female recipient mice (chimerism) are shown, as determined by qPCR amplification of Y-chromosome DNA template. Also shown is the number of vector copies per *Il2rg*^{-/-} donor cell at 8 months post-transplant. 2A shows average copy numbers per donor cell were calculated by normalizing the LV copy number of each individual mouse by its chimerism. Averages are shown ± SEM. All mice received 6Gy irradiation and 4x10⁵ transduced Lin- cells. 2B shows chimerism, number and average copy number per donor BM cell, as well as spleen cell counts, of mice treated with γcPr-coγc at different vector, radiation and cell doses. Cell counts from mice given untreated *Il2rg*^{-/-} or WT Lin- cells are also included for comparison.

Table S2. T, B and NK cell percentages in spleen.

Mouse	CD3 ⁺	IgM ⁺ B220 ⁺	CD3 ⁺ NK1.1 ⁺	Copies/cell	Background
Wild type (n=2)	28.5	52.0	2.2		C57BL/6
<i>Rag2</i> ^{-/-} / <i>Il2rg</i> ^{-/-} (n=2)	0	0	0		
γ CPr-co γ C (n=4)	26.0 \pm 4.7	32.3 \pm 5.3	0.7 \pm 0.2	3.7 \pm 0.2	
UCOE-co γ C (n=3)	31.5 \pm 3.5	50.7 \pm 1.5	0.5 \pm 0.0	0.5 \pm 0.1	
Wild type (n=4)	26.8 \pm 1.0	45.7 \pm 4.1	n/a		BALB/c
SF-co γ C (n=4)	29.4 \pm 0.7	31.1 \pm 1.4	n/a	1.9 \pm 0.2	
PGK-co γ C (n=5)	26.7 \pm 1.9	23.8 \pm 3.3	n/a	2.8 \pm 0.7	

Averages of flow cytometry percentages for CD3⁺ and CD3⁺NK1.1⁺ cells in the spleens of *Il2rg*^{-/-} **C57BL/6** mice treated with LV- γ CPr-co γ C (n=4) or untreated (n=2) and wild type **C57BL/6** mice (n=2). γ CPr-co γ C and UCOE averages are shown \pm SEM. Measurements were made at 6 months post-transduction. Measurements from mice treated with SF-co γ C and PGK-co γ C in a similar experiment using *Il2rg*^{-/-} BALB/c mice are included for reference.

In order to measure NK1.1 positive cells in spleen, *Il2rg*^{-/-} mice were subjected to a LV- γ CPr-co γ C transduction/transplantation scheme as described in the Methods section, with the following exceptions: the HeLa MOI on *Il2rg*^{-/-} Lin⁻ cells was 15, recipients were *Rag2*^{-/-}/*Il2rg*^{-/-}, and all mice were of a **C57BL/6** background. Wild type **C57BL/6** mice and untreated *Rag2*^{-/-}/*Il2rg*^{-/-} **C57BL/6** mice were used as controls. These experiments were conducted at University College, London, according to their protocols²⁶, and were all approved in accordance with legislation in the UK.

Table S3. Viral integrations near RTCGD genes, by promoter.

	genes in RTCGD	genes not in RTCGD
SF promoter	7 (14.3%)	42 (85.7%)
Cellular promoters	0 (0%)	33 (100%)

Number of viral integration sites found within 100kb of a gene listed in the Retrovirus Tagged Cancer Gene Database in vectors using either the SF promoter or a physiological promoter. The difference is statistically significant (p=0.017). Integration sites were derived from a total of 23 mice.

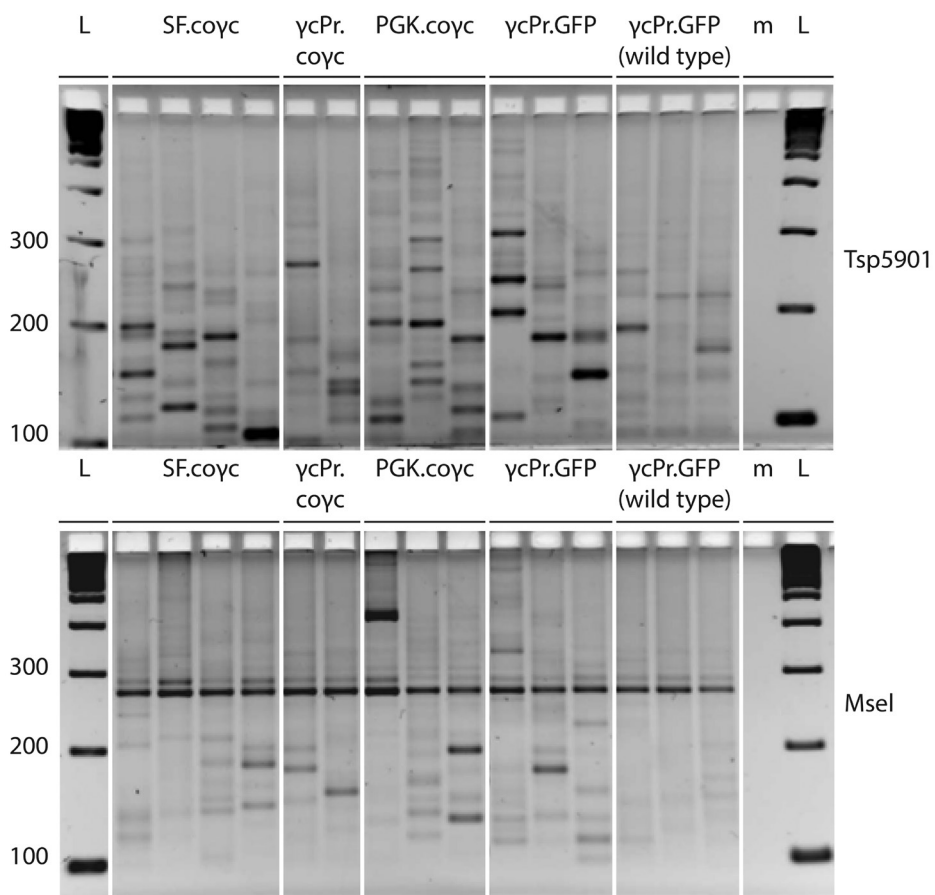


Figure S1. LAM-PCR of bone marrow DNA in gene therapy mice.

LAM-PCR integration site analysis of selected gene therapy-treated mice. High-resolution gel images LAM-PCR amplicons derived from BM DNA after digestion with Tsp590I (A) or MseI (B) restriction enzymes. Lanes designated 'L' are ladder lanes, and lanes designated 'm' contain genomic DNA from untreated mice.

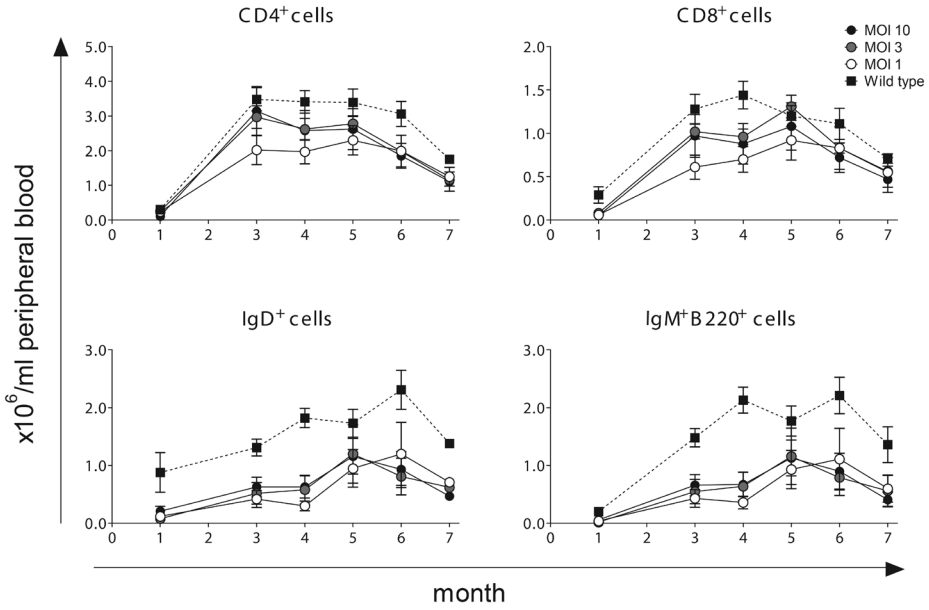


Figure S2. LV- γ cPr-coyc titration FACS data.

Figure displays the results of the LV- γ cPr-coyc vector titration: *Il2rg^{-/-}* mice were transplanted with *Il2rg^{-/-}* Lin-cells transduced with γ cPr-coyc at an MOI of 10, 3 or 1 HTU ($n = 5$ per group) or wild-type Lin-cells ($n = 4$). Shown are the group averages for CD4⁺CD3⁺, CD8⁺CD3⁺, IgD⁺ and IgM⁺B220⁺ cells in peripheral blood as measured monthly by flow cytometry. qPCR of BM DNA found the average integrations per cell to be equal to 1.9, 1.0, and 0.2 for the MOI 10, 3 and 1 groups, respectively. Mice were subjected to 2 Gy irradiation and received 5×10^5 cells.



Met Tyr Xle Ala Cys Asn Pyl Thr
Ala Xle Ala Cys Ser Pyl Thr
Cys Xle Ala Lys Ser Pyl Thr
Ala Xle Ala Lys Ser Pyl Thr
PylAsn Glu Asn Ser SecSer
Ser Lys Leu SerThr LeuGlu Asn TrpIle LeuLeu IleAsaAsn His
Lys Cys Glu Ser GGAC GCC GGCATGCC TrpIle LeuLeu Pyl His Sec
Ala Xle Thr AGCAGGTACGCGAGA His Ala Asn Ser
GluArg Thr AGGCTAGCTAG PylThr
CAGGCTCGA Arg His
CAGTGGGTAT PylAsn Arg AlaArg ArgIle
GTGAGCTTGCTC Thr Ile Glu Asn Pyl Thr
GATAGCGGGTA Ser Sec Thr Pyl Thr His
TGCTAGCCGGC His SerSec
GGGAATGCAGG
TACGCTTGCTAGA
CCGCAGGCATGC
CTGTATGCTTGCTC
01001101011000010111
00100 111001101101000
01100 001011011000110
1100 0010000001010111
011 010010110110001101
10 0011010010110000101
101101001 00000010010
00011101 01011100110111010
0011011 11011011100010110000
100000 01010
00001 1010

Pre-transplant mobilization with granulocyte colony stimulating factor improves lentiviral vector gene therapy in a SCID-X1 mouse model

Marshall W. Huston, Adriaan R. A. Riegman, Yvette van Helsdingen, Niek P. van Til
and Gerard Wagemaker.

Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands

Manuscript in preparation

ABSTRACT

Hematopoietic stem cell (HSC) viral vector gene therapy has proven to be an effective treatment of severe combined immunodeficiencies (SCIDs). It has been hypothesized that creating open bone marrow niches in HSC transplant recipients improves the chances of donor HSCs to successfully engraft and accelerates hematopoietic reconstitution, which could be very helpful for transplantation of SCID patients. However, the poor health of most of these patients at diagnosis precludes the myelosuppressive methods used in standard bone marrow transplantation protocols. To investigate the feasibility of hematopoietic stem cell transplantation in a non-cytoreductive manner, HSCs of *Il2rg*^{-/-} mice were mobilized with the growth factor granulocyte colony stimulating factor (G-CSF) and subsequently transplanted with wild type lineage negative (Lin⁻) cells or *Il2rg*^{-/-} Lin⁻ cells transduced with codon optimized *IL2RG* lentiviral vectors. Mice were monitored for the reconstitution of lymphocyte populations, level of donor cell chimerism and immune system responses and the results compared to mice given wild type or gene therapy treated Lin⁻ cells after 2 Gy irradiation. The results demonstrate that G-CSF pre-transplant treatment of immunocompromised recipients results in levels of lymphocyte reconstitution similar to mild irradiation and superior to no pre-transplant conditioning.

INTRODUCTION

The most common form of severe combined immunodeficiency (SCID), affecting 50-60% of patients, is X-linked SCID (SCID-X1). It is caused by a mutation in the IL-2 receptor gamma gene (*IL2RG*) leading to a non-functional common gamma chain protein [1]. While bone marrow transplantation from healthy HLA-identical or haploidentical donors has proven highly effective at treating SCID-X1 [2], HLA-matched sibling donors are available to only ~10% of patients [3]. Transplantation from mismatched donors has a significantly higher risk of morbidity and mortality [4]. In order to treat SCID-X1 patients lacking a suitable HLA-matched donor, alternative methods such as gene therapy are being developed.

The efficacy of gammaretroviral gene therapy to successfully treat SCID-X1 was demonstrated in seminal clinical trials in London and Paris [5, 6]. CD34⁺ bone marrow (BM) cells were *ex vivo* transduced with a gammaretroviral vector carrying a correct IL-2RG cDNA copy. This approach restored T cell populations in 17 out of 20 patients, a success rate comparable to HLA-identical bone marrow transplantation [7]. T and NK cell counts and serum Ig levels have been sustained in some patients for up to twelve years, and at the time of the last report 6 of the 17 patients no longer need intravenous immunoglobulin[8]. All patients had detectable proviral integrations in their T cells, whereas gene-marking in B cell populations was very low (<1%) as were functional B cells levels. More recently initiated gene therapy trials to treat SCID-X1 with self-inactivating gamma-retroviral vectors adhere to similar protocols without pre-conditioning.

A clinical trial for ADA-SCID gammaretroviral gene therapy, initiated in Milan in 2000, ran concurrently to the SCID-X1 trial. A report on the first 10 patients treated stated that 9 of these patients demonstrated full T cell reconstitution, 8 remained off ADA enzyme replacement therapy and immune globulin replacement therapy had been discontinued in 5 [9]. There were far more vector-marked B and NK cells in the peripheral blood (52% and 59%, respectively) compared to what was seen in the SCID-X1 trial. While direct comparison of the results of the Milan ADA-SCID and SCID-X1 clinical trials is impractical, one notable difference between the trials was that patients in the ADA trial were given low intensity conditioning with Busulfan prior to transplantation, which may have promoted engraftment of multilineage stem/progenitor cells. A London ADA-SCID trial which also used a mild pre-transplant conditioning regimen reported high levels of gene marking in T, B and NK cells[10].

The beneficial role of pre-transplant conditioning in improving engraftment of gene-therapy treated donor hematopoietic stem cells (HSCs) has been previously demonstrated in a mouse model of SCID-X1[11]. However, current methods of emptying bone marrow niches, including irradiation, myelosuppressive conditioning and chemotherapeutic agents, have significant toxicity and can lead to substantial transplant related morbidity and mortality. In order to test whether creating sufficient open bone marrow niches

for donor HSC engraftment can be accomplished using a potentially less invasive and clinically applicable method, we initiated a gene therapy treatment in *Il2rg*^{-/-} mice using the previously described self-inactivating lentiviral (LV) vectors[11] after HSC mobilization via granulocyte colony stimulating factor (G-CSF)[12, 13]. G-CSF is used in the clinic to mobilize HSCs and progenitor cells into the peripheral blood for use as donor cells in bone marrow transplantation protocols[14, 15]. We hypothesized that mobilization of recipient HSCs into the bloodstream would improve the engraftment of gene therapy treated donor HSCs.

To test this hypothesis, similar to the clinical gene therapy protocols for treating SCID-X1, we tested lentiviral gene therapy in *Il2rg*^{-/-} mice with the addition of G-CSF pre-transplant and compared the subsequent lymphocyte reconstitution to mice given 2Gy irradiation conditioning.

RESULTS

Evaluation of mobilization factor G-CSF

In order to assess the efficacy of G-CSF to mobilize HSCs, *Il2rg*^{-/-} mice were treated with four injections of G-CSF, and another group was given four injections of saline solution as a control (n=5 for both groups). Peripheral blood was collected from these mice one day prior to starting the mobilization treatments and five hours after the last G-CSF injection. Absolute number of white blood cells and percentage of CD11b⁺, Sca-1⁺ and c-Kit⁺ cells were measured for both time points and the ratio of the absolute number of each cell type before and after mobilization was calculated.

These mobilization protocols resulted in a large increase in the number of circulating white blood cells (WBC), composed mainly of CD11b⁺ myeloid cells in *Il2rg*^{-/-} mice (Figure 1a). Of the progenitor/stem cell markers measured, the most dramatic mobilization effect was seen in the c-Kit⁺Sca-1⁻ subtype, which had a 36-fold increase. The more stem cell enriched Sca-1⁺c-Kit⁺ population had a fold increase of 6.1. In control mice injected with saline solution no significant changes were observed. The absolute numbers of these cell types post-mobilization is shown in Figure 1b.

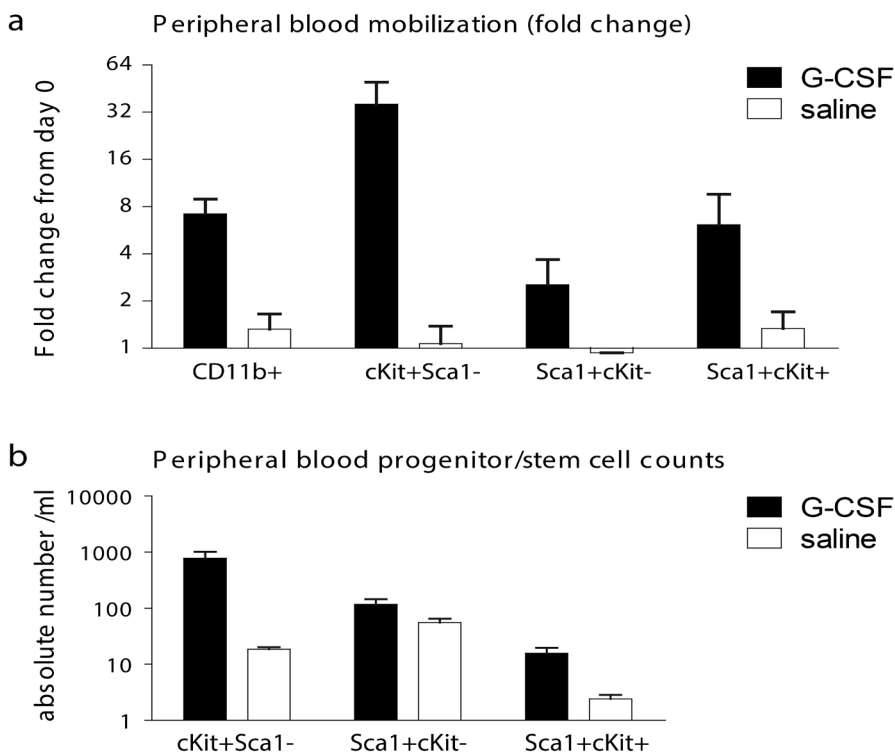


Figure 1: Effect of G-CSF mobilization protocols on circulating stem and progenitor cells in peripheral blood of *Il2rg*^{-/-} mice. (a) shows the fold change in the absolute numbers of myeloid and stem cell marker positive cells measured in peripheral blood after treatment with mobilization factor G-CSF compared to the numbers prior to mobilization. (b) shows the absolute numbers of c-Kit⁺, Sca-1⁺ and c-Kit⁺Sca-1⁺ cells after mobilization. Figures contain the combined results of two experiments.

Wild type Lin⁻ cell transplantation after stem cell mobilization in *Il2rg*^{-/-} mice

Female *Il2rg*^{-/-} mice were subjected to the G-CSF mobilization protocol described above and subsequently transplanted with 3×10^5 male lineage negative (Lin⁻) wild-type (WT) BALB/c cells (n = 6 for G-CSF). Other groups were transplanted with 3×10^5 WT Lin⁻ cells after either 2 Gy irradiation (n = 5) or saline injections (n = 3). Mice were monitored for 6 months with monthly analysis of peripheral blood lymphocyte reconstitution and subsequently sacrificed for analysis of spleen and bone marrow lymphocyte populations. Absolute lymphocyte numbers in peripheral blood were calculated via WBC counts and FACS analysis and compared between the various groups (Figure 2a-c). Mature T cells were identified via flow cytometric analysis of CD4⁺CD3⁺ and CD8⁺CD3⁺ cells, and mature B cells for B220⁺IgM⁺IgD⁺ cells (Figure 2 d-f) and these were similar between G-CSF and 2Gy irradiation groups in both tissues. Similar results were found in the bone marrow (data not shown). Overall, *Il2rg*^{-/-} mice transplanted with wild type Lin⁻ cells after

G-CSF or 2 Gy total body irradiation (TBI) conditioning had lymphocyte levels similar to untreated age-matched BALB/c mice. Mice transplanted with wild type cells after saline injections tended to have reduced B cell counts. Levels of Y-chromosome chimerism (i.e. donor cell chimerism) were determined in bone marrow and spleen cells. Y-chimerism was highest in the 2Gy group and lowest in the saline group (Table 1).

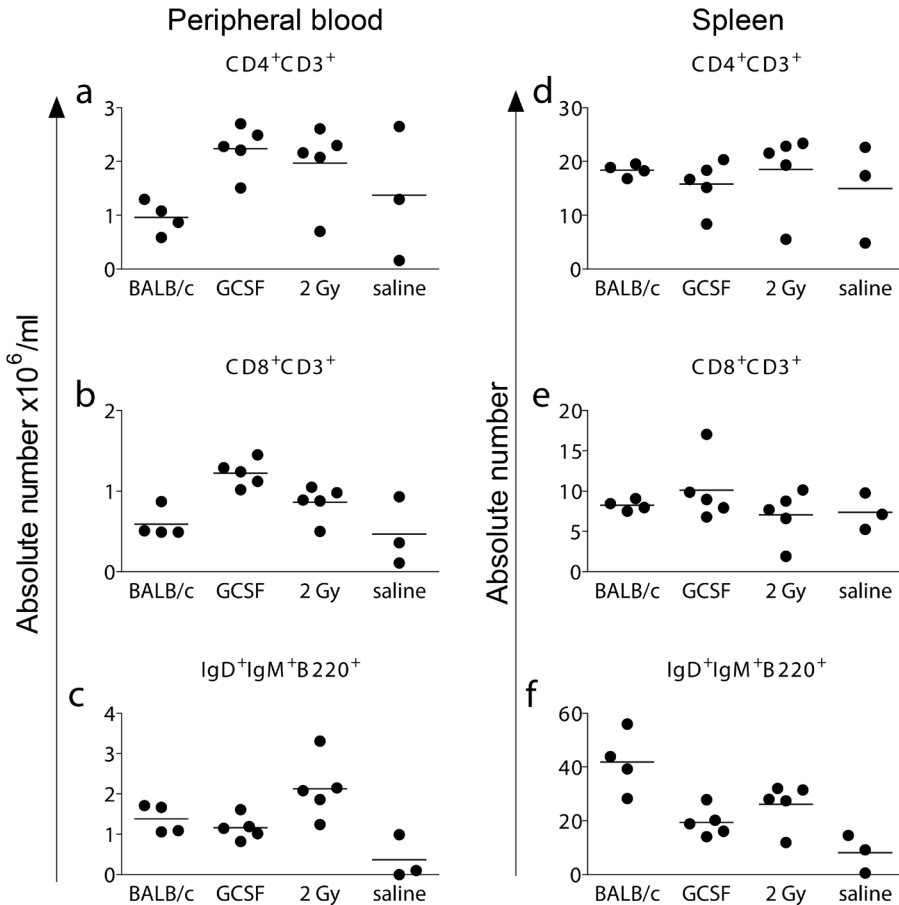


Figure 2: Absolute lymphocyte numbers in *Il2rg*^{-/-} mice transplanted with wild type Lin⁻ cells after conditioning regimens. Flow cytometry profiles and cell counts were used to calculate mature lymphocyte numbers in peripheral blood (a-c, absolute number per ml) and in spleen (d-f, absolute number) of *Il2rg*^{-/-} mice transplanted with 3×10^5 wild type Lin⁻ BM cells after G-CSF mobilization or total body irradiation. 2Gy = 2 Gy irradiation conditioning. Measurements were taken six months post-transplant. Untreated age-matched wild type (BALB/c) mice are included for comparison.

Table 1: Spleen and BM Y-chimerism levels in *Il2rg*^{-/-} mice transplanted with wild type BALB/c Lin- cells after various conditioning regimens.

Group	Spleen cell chimerism (%)	BM cell chimerism (%)
G-CSF (n=5)	59.7 ± 17.4	33.3 ± 14.6
2 Gy irradiation (n=5)	73.4 ± 14.4	57.0 ± 16.3
Saline (n=3)	36.1 ± 10.0	25.2 ± 13.4

Data are shown as mean ± SEM

Lentiviral vector gene therapy in *Il2rg*^{-/-} mice treated with mobilization factors

In order to determine the viability of mobilization factors as a pre-transplant regimen prior to lentiviral gene therapy, female *Il2rg*^{-/-} mice were subjected to G-CSF mobilization (n = 5) or 2 Gy irradiation (n = 4) and subsequently transplanted with 3×10^5 Lin- *Il2rg*^{-/-} cells taken from a pool of cells transduced overnight with self-inactivating (SIN) lentiviral vectors containing codon optimized human *IL2RG* gene (*coIL2RG*) previously described[11]. Transgene expression was driven by the spleen focus forming virus promoter (SF, HeLa multiplicity of infection (MOI) = 3).

One mouse in the group given 2 Gy irradiation conditioning developed a large tumor in its left hindleg and had to be sacrificed at 6 months post-transplant. Analysis of the spleen, bone marrow and peripheral blood revealed no hematopoietic abnormalities in this mouse. One mouse in the G-CSF group developed an extremely high WBC count (258×10^6 cells per ml) 6 months post-transplant and had to be sacrificed. Analysis of this mouse revealed an enlarged spleen (980mg), a high percentage of CD4⁺CD8⁺ cells in the bone marrow (60.9%) and 2.4 vector integrations per BM cell. Integration analysis by LAM-PCR is pending and will be included in a future publication: for the purposes of this study, this leukemic mouse was excluded from group comparison analysis. The remaining mice were sacrificed at seven months post-transplant for analysis of spleen and bone marrow lymphocyte populations.

Absolute lymphocyte numbers in peripheral blood and in spleen cells were compared to values taken from untreated *Il2rg*^{-/-} mice and wild type BALB/c mice (Figure 3). Significant differences ($p < 0.05$) compared to untreated *Il2rg*^{-/-} mice were found in *SF.coIL2RG*-treated mice in the CD8⁺ and IgD⁺IgM⁺B220⁺ peripheral blood lymphocyte populations regardless of whether the mice were conditioned with G-CSF or 2 Gy irradiation. The same significant differences were seen in untreated BALB/c controls (Figure 2). Spleen vector copy per cell taken from *SF.coIL2RG*-treated mice revealed average integrations per cell of 1.1 and 0.9 for the G-CSF and 2 Gy conditioned groups, respectively (Table 2). Spleen and BM Y-chimerism levels were higher, but not significantly higher, in the 2 Gy group than in the G-CSF group.

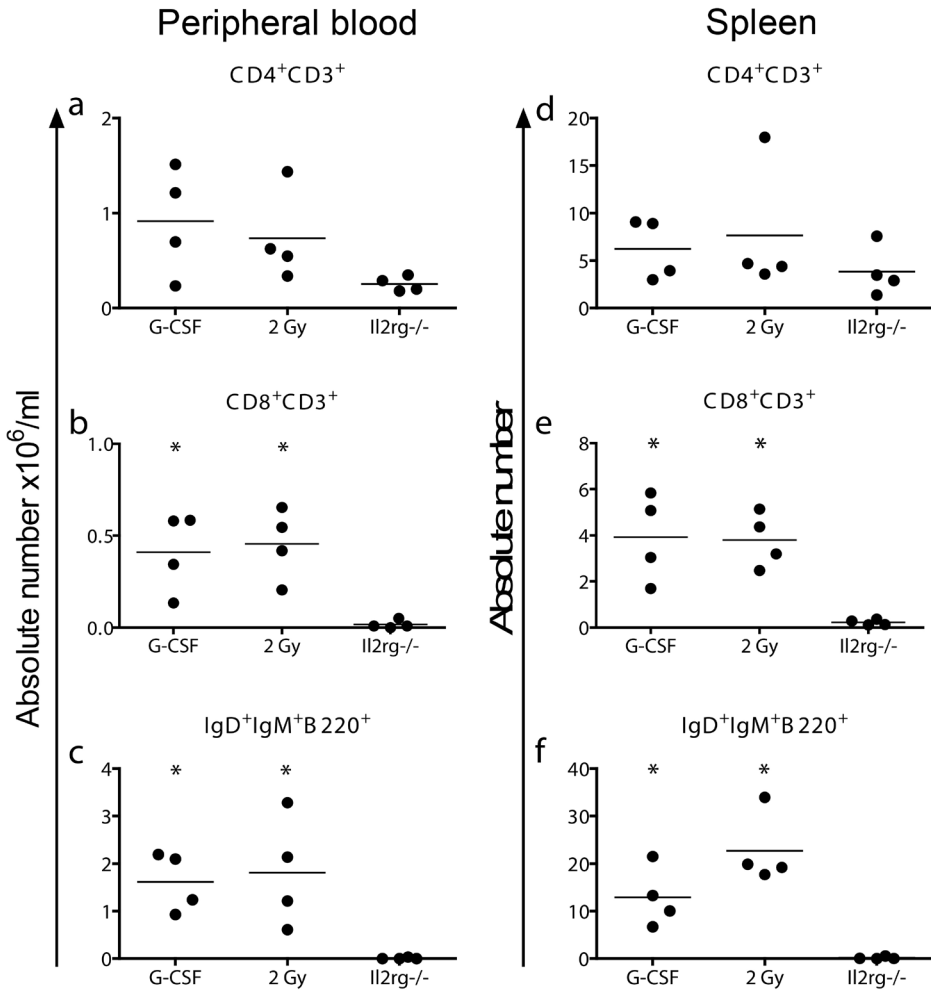


Figure 3: Comparison of G-CSF and 2Gy conditioning regimens in mice treated with SF.coIL2RG lentiviral vectors. Figure shows reconstitution of lymphocytes in peripheral blood (a-c) at six months and spleen (d-f) at seven months after transplantation of 3×10^5 SF.coIL2RG-transduced *Il2rg*^{-/-} Lin⁻ BM cells following conditioning regimens of G-CSF or 2 Gy irradiation (2Gy). Untreated age-matched *Il2rg*^{-/-} mice are included for comparison (* = significantly different from *Il2rg*^{-/-} group, p-value <0.05)

Table 2: Spleen and BM Y- chimerism levels and viral integrations per cell in *Il2rg*^{-/-} mice transplanted with SF.coIL2RG vector-treated Lin⁻ cells after various conditioning regimens.

Group	Spleen cell chimerism (%)	BM cell chimerism (%)	Integrations per cell spleen	Integrations per cell BM
G-CSF (n=4)	30.9 ± 15.5	25.3 ± 11.4	2.1 ± 0.9	1.1 ± 0.5
2 Gy irradiation (n=4)	60.9 ± 9.4	64.7 ± 11.1	1.4 ± 0.2	0.9 ± 0.5

Vector copy per cell is corrected for donor chimerism. Data are shown as mean ± SEM

In order to determine if G-CSF mobilization was sufficient to allow engraftment of Lin⁻ cells treated with lentiviral vectors containing a weaker eukaryotic promoter, we repeated the G-CSF protocol using the *SF.coIL2RG* vector or a vector driven by a 1.1 kb section of the native *IL2RG* promoter (γ cPr). A HeLa MOI of 3 was used with the goal of producing 1 vector copy per cell. The number of Lin⁻ BM cells transplanted was reduced slightly to 1×10^5 . Reconstitution of T and B lymphocytes in these mice can be found in Figure 4. The SF and γ cPr groups had significantly higher levels of CD8⁺ and IgM⁺IgD⁺B220⁺ lymphocytes in the spleen compared to the untreated *Il2rg*^{-/-} mice. Overall, the γ cPr vector was able to increase the T and B populations in *Il2rg*^{-/-} mice, but not to wild type levels. Y-chimerism and vector integration per cell analysis revealed lower levels of donor cell engraftment in spleen and bone marrow, as well as lower vector copy number per cell (Table 3), for the γ cPr group compared to the SF group.

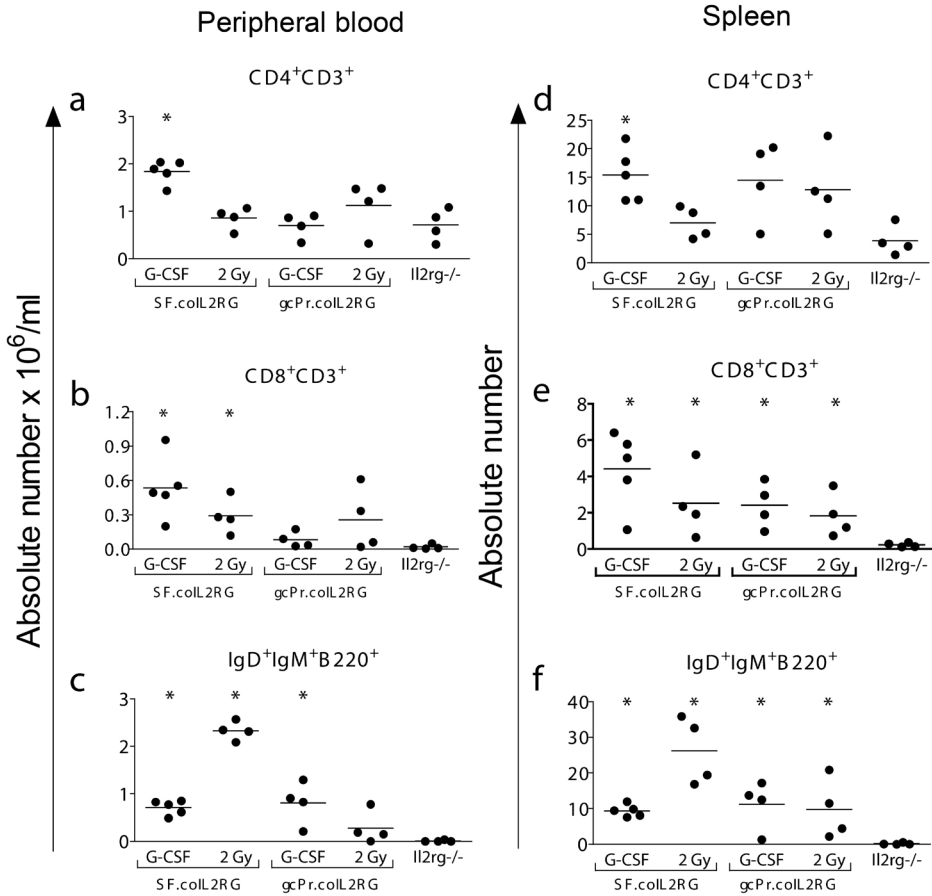


Figure 4: Comparison of various promoter elements in mice given coIL2RG vectors following a G-CSF conditioning regimen. Figure shows reconstitution of lymphocytes in peripheral blood (a-c) and spleen (d-f) after transplantation of 1×10^5 Lin- *Il2rg*^{-/-} cells transduced with coIL2RG lentiviral vectors containing SF or γ cPr promoter elements. Measurements calculated via cell counts and flow cytometry analysis six months post-transplant. Untreated age-matched *Il2rg*^{-/-} mice are included for comparison (* = significantly different from *Il2rg*^{-/-} group, p-value <0.05)

Table 3: Spleen and BM Y- chimerism levels and viral integrations per cell in *Il2rg*^{-/-} mice transplanted with LV vectors after various G-CSF conditioning.

Group	Spleen cell chimerism (%)	BM cell chimerism (%)	Integrations per cell spleen	Integrations per cell BM
<i>SF.coIL2RG</i> G-CSF (n=5)	23.1 ± 5.3	11.6 ± 2.6	1.1 ± 0.6	0.6 ± 0.2
<i>SF.coIL2RG</i> 2 Gy (n=4)	47.9 ± 12.0	46.5 ± 12.1	1.2 ± 0.2	1.1 ± 0.4
γ cPr: <i>coIL2RG</i> G-CSF (n=4)	5.4 ± 1.0	5.7 ± 2.2	0.4 ± 0.2	0.9 ± 0.3
γ cPr: <i>coIL2RG</i> 2 Gy (n=4)	13.4 ± 8.7	21.1 ± 5.6	0.3 ± 0.2	0.5 ± 0.3

Vector copy per cell is corrected for donor chimerism. Data are shown as mean ± SEM

To confirm that a specific T cell dependent antibody response was restored in the *Il2rg*^{-/-} mice transplanted with *SF.coIL2RG* or wild type cells, a tetanus toxoid immunization scheme was started at five months after transplantation as previously described[11]. Tetanus-specific IgG1 antibody levels increased in all groups given wild type cells or *SF.coIL2RG*-treated cells (Figure 5). Antibody titers were highest in the wild type group, but all G-CSF and 2Gy conditioned groups had a strong, specific anti-tetanus immune response.

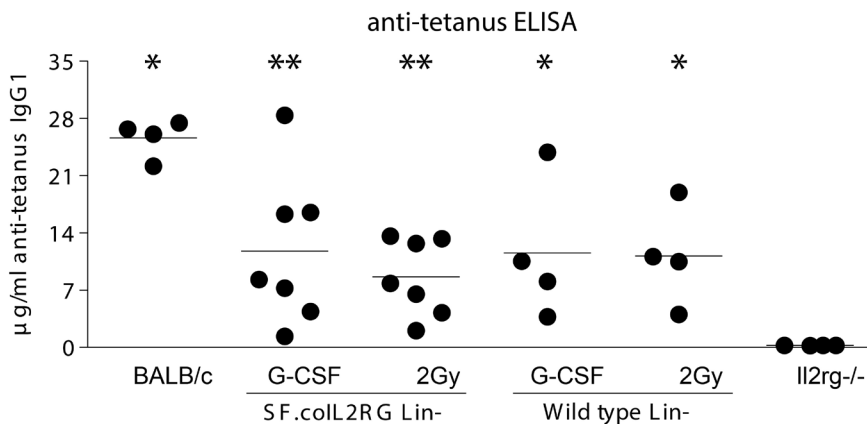


Figure 5: Anti-tetanus toxoid IgG1 immune response. Mice transplanted with wild type BALB/c Lin- cells or *SF.coIL2RG* vector-treated *Il2rg*^{-/-} Lin- cells after 2Gy irradiation or G-CSF conditioning were subjected to a tetanus toxoid immunization protocol beginning four months post-transplant. Concentrations of plasma anti-IgG1 tetanus toxoid antibodies were determined ten days after the last tetanus injection and compared to untreated BALB/c and *Il2rg*^{-/-} mice (* = p-value <0.05 compared to *Il2rg*^{-/-} group, ** = p-value <0.01)

DISCUSSION

We have previously shown that minimal pre-conditioning significantly improves reconstitution of gene-corrected cells in *Il2rg^{-/-}* mice [11]. Pre-transplant conditioning regimens, such as irradiation, myelosuppressive conditioning and chemotherapeutic agents, to suppress or eliminate endogenous bone marrow cells have been shown to improve the ability of donor HSCs cells to contribute to hematopoietic reconstitution[16, 17]. However, myeloablative and myelosuppressive conditioning is associated with significant morbidity and mortality [18]. SCID-X1 patients are often in such poor health from recurrent infections by the time they are approved for gene therapy intervention that even mild myelosuppressive regimens could be fatal[19]. Other agents that are less toxic may provide a safer way to treat patients with primary immune deficiencies, in particular SCID-X1 patients that in the current gene therapy protocols are not conditioned.

Granulocyte colony forming factor (G-CSF) is a cytokine which disrupts the interaction between SDF-1, produced by stromal cells, and CXCR4 in granulocyte precursors, resulting in a myeloid hyperplasia in the bone marrow and expansion of both mature and immature myeloid cell types as well as the release of quiescent HSCs into circulation[20, 21]. Repeated treatments of G-CSF over a period of 4-5 days results in the release of a considerable number of CD34⁺ cells into the peripheral blood, which can be harvested and subsequently administered into patients requiring bone marrow transplants [22]. For this reason, G-CSF is routinely used in the clinic to collect donor cells for bone marrow transplantation without resorting to more invasive procedures. There has been some investigation into the possibility of using mobilization to create open bone marrow niches in HSC transplant recipients [15], but to our knowledge no one has yet applied the use of mobilization factors to a gene therapy protocol.

Since engraftment of hematopoietic stem cells in the bone marrow is a competitive process, with donor HSC engraftment increasing based on the number of cells transplanted[23], we postulated that G-CSF mobilization of recipient HSC would improve donor cell engraftment and immune system reconstitution in immunodeficient *Il2rg^{-/-}* mice.

First, we tested the ability of G-CSF as an agent to mobilize the HSCs. By using a four day injection scheme a 7-fold increase was observed in the myeloid population in the periphery, but more importantly, also a 40- fold increase of c-Kit⁺ progenitor cells and a 6-fold increase c-Kit⁺Sca1⁺ HSCs was observed. This mobilization was roughly 2-fold lower than other reports in wild type mice using murine G-CSF [24, 25], possibly due to cross-species differences in the activity of this cytokine.

After validating the mobilization scheme, wild type Lin⁻ cells were transplanted in G-CSF treated mice. This resulted in high levels of spleen and peripheral blood T lymphocytes, indistinguishable from 2 Gy irradiated and saline treated mice. These results are similar to the SCID-X1 gene therapy protocols, in which high peripheral T cell populations can be observed in non-conditioned patients [10].

Possibly due to the lack of myelosuppressive conditioning, patients in the London and Paris SCID-X1 clinical trials generally had poor B cell reconstitution and low levels of gene-marking in the B cells, even though reconstitution levels and gene marking in T cells was high [6]. The gene therapy treated donor HSCs and B cell progenitors may have had difficulty engrafting in the bone marrow, a disadvantage not seen in T cell progenitors possibly due to: the higher selective advantage of these cell types, the lack of competing T cell progenitor cells in the host [26] or the ability of multipotent progenitors cells to home directly to the thymus [27]. Pre-transplant myelosuppressive conditioning, such as was done in the ADA-SCID trial in Milan [9], could improve HSC engraftment by creating open bone marrow niches. A non-cytoreductive method of creating open bone marrow niches could be safely attempted in SCID-X1 patients and may lead to improved engraftment and faster reconstitution.

In our mice, the number of B cells was similar for the groups given wild type cells after G-CSF or 2 Gy conditioning and higher than mice given wild type cells and saline. Donor cell chimerism levels in the bone marrow for the 2 Gy and G-CSF groups were higher than the saline group. Similar results were found for both peripheral T and B cell numbers and donor chimerism when *Il2rg*^{-/-} Lin- cells were transduced with a *SF.coIL2RG* lentiviral vector with low vector copy number per cell. Furthermore, when the number of cells transplanted was reduced 3-fold to 1×10^5 Lin- cells and the eukaryotic native IL2RG promoter was used at minimal vector copies per cell, the majority of mice in these groups had measureable levels T and B cells, and significant T cell dependent antibody responses. On average, the γ Pr groups had lower vector marking and lower donor cell chimerism in the BM and spleen compared to mice treated with the SF vector, which may suggest that the lower expression by the IL2RG promoter negatively affects the selective pressure to repopulate the bone marrow niches. The modest levels of lymphocyte reconstitution reached, comparable to 2 Gy irradiation, also confirms previously obtained results with this γ Pr.coIL2RG vector [11].

Although there was considerable peripheral reconstitution observed with G-CSF treatment, the levels of donor cell chimerism were reduced compared to those treated with 2 Gy irradiation. Chimerism levels in γ Pr.coIL2RG mice treated with G-CSF were only 2-fold higher than levels previously seen with this vector when no conditioning was used[11]. This may indicate that G-CSF-mediated HSC mobilization is not as effective as irradiation conditioning at creating open bone marrow niches. It is also known that not all patients respond to G-CSF treatment, and need additional agents such as plerixafor to improve mobilization[28]. Combinations of other mobilization agents may improve engraftment as well. However, the modest efficacy of G-CSF in our experiments is offset by the greatly reduced toxicity of G-CSF treatment compared to the myelosuppressive conditioning used in the clinical settings. Overall, mobilization of recipient bone marrow HSCs with G-CSF was found to have efficacy similar to mild irradiation conditioning when transplanting Lin- cells into an immunodeficient mouse model.

Unfortunately, one mouse in the *SF.coIL2RG* group contracted a CD4⁺CD8⁺ leukemia 6 months post-transplant. Analysis of BM cells revealed 2.4 vector integrations per cell, and although we do not yet know if this leukemic event was a direct consequence of insertional mutagenesis, this may highlight the risks inherent in the SF promoter that have been observed by others [29] and underlines that SF is not an attractive promoter for clinical gene therapy applications.

The success of the *SF.coIL2RG* vector after G-CSF treatment demonstrates the feasibility of mobilization factors as a way to improve engraftment of vector-treated donor HSCs in gene therapy of certain SCIDs. While the SF promoter is not an attractive option for clinical use, other, safer vectors, such as the self-activating gammaretroviral vector with the elongation factor 1 α short promoter [30] currently used in a multicentre clinical trial, or eukaryotic promoters, such as its native promoter [31], the ubiquitous chromatin opening element [32] or vav promoter [33] in combination with a lentiviral vector may be used. These promoters could retain sufficient transgene expression without the increased risk of insertional mutagenesis or methylation seen in SF [30, 33-35]. Administration of these vectors in conjunction with G-CSF mobilization may result in better reconstitution and a more effective treatment in future gene therapy clinical trials for SCID-X1.

MATERIALS AND METHODS

Mice

Il2rg^{-/-} mice and syngenic BALB/c wild type mice [11] were bred in the Experimental Animal Center of Erasmus MC. All mice were used at 6 to 10 weeks of age and were maintained in specified pathogen free conditions. Experiments were approved by the institutional Animal Ethical Committee of Erasmus MC in accordance with legislation in the Netherlands.

G-CSF mobilization protocols

Female *Il2rg*^{-/-} mice were given subcutaneous injections of 6 µg of filgrastim (recombinant methionylated human granulocyte-colony stimulating factor) (Sandoz, Frimley, UK) in 50 µl of saline solution or saline only daily for four consecutive days. Peripheral blood was collected at five hours after the last injection, at what is considered to be the peak mobilization time. The percentage of CD11b⁺, Sca-1⁺ and c-Kit⁺ cells were measured via flow cytometry. Absolute numbers of these cells were calculated based on total white blood cell counts and compared to the values in those animals one day prior to administration of G-CSF. Other cohorts of female *Il2rg*^{-/-} mice were given one of the aforementioned mobilization protocols and transplanted with 5×10^5 wild type or 1×10^5 LV vector treated lineage negative BM cells five hours after the last G-CSF injection.

Production of lentiviral vectors

Third generation self-inactivating (SIN) LV vectors incorporating codon-optimised *IL-2RG* cDNA driven by either the SFFV viral promoter or a 1.1kb section of the native *IL2RG* promoter were previously described [11]. LVs were produced by standard calcium phosphate transfection of HEK 293T cells [36] with the packaging plasmids pMDL-g/pRRE, pMD2-VSVg and pRSV-Rev. Vector particle concentration and titration were carried out as previously described [11].

Transduction and transplantation of lineage negative BM cells

BM cells from male *Il2rg*^{-/-} and congenic wild type BALB/c mice were purified by lineage depletion (Lin-) (BD Biosciences, Santa Clara, CA). Lin- BM cells were transduced overnight at 10^6 cells/ml in serum free modified Dulbecco's medium with supplements [37] in the presence of murine stem cell factor (mSCF) 100 ng/mL, human FMS-like tyrosine kinase 3-ligand (hFlt3-L) 50ng/ml and murine thrombopoietin (mTPO) 10 ng/ml at HTU MOI 10 (SF vector) or MOI 3 (γ cPr and UCOE vectors). Subsequently, 1 or 3×10^5 cells were injected into the tail vein of 2 Gy irradiated female or G-CSF treated *Il2rg*^{-/-} recipients. Mice were monitored monthly for six months via peripheral blood collection and analysis of hematopoietic cell counts and flow cytometry. Mice were sacrificed 6-7 months post transplant and bone marrow and spleen cells were isolated for further analysis.

Immunophenotyping by flow cytometry

Flow cytometric analyses were performed on cells obtained from blood, bone marrow and spleen. Peripheral blood was collected monthly in EDTA tubes by retro-orbital puncture under isoflurane anesthesia. Complete blood cell counts were measured using a Vet ABC hematology analyzer (Scil animal care company GmbH, Germany). Blood was lysed and leukocytes were washed three times with Hank's balanced salt solution (HBSS, Invitrogen) containing 0.5% (wt/vol) bovine serum albumin and 0.05% (wt/vol) sodium azide (HBN). Cells were incubated for 30 minutes at 4°C in HBN containing 2% heat-inactivated normal mouse serum and antibodies against CD3, CD4, CD8, B220, IgM, IgD, CD11b, Sca-1 and c-Kit directly conjugated to R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP) or allophycocyanin (APC; all antibodies BD Biosciences, Santa Clara, CA). Subsequently, cells were washed and measured on a FACSCanto (BD Biosciences). BM and spleen cells were evaluated similarly.

Real-time quantitative polymerase chain reactions

qPCR to quantify the integrated proviral copy number was performed as described before [11]. qPCR reactions were performed on spleen and BM DNA from *Il2rg^{-/-}* mice transplanted with male donor cells transduced with LV vectors to amplify a region of the Y chromosome as described by Pujal et al [38], using 100ng of genomic DNA and primers 5'-TCATCGGAGGGCTAAAGTGTCAC-3' and 5'-TGGCATGTGGGTTCCCTGTCC-3'. A standard curve was generated using spleen and BM DNA from male BALB/c mice. All PCR reactions were performed in triplicate and analyzed with SDS2.2.2 software.

Immunizations and ELISAs

T cell dependent specific antibody responses were determined by intraperitoneal immunization with 16 IU of tetanus toxoid. This process was repeated three times at two week intervals, with plasma collected two weeks after the last immunization and just prior to the initial immunization. An ELISA was performed on plates coated with tetanus toxoid, and HRP goat-anti-mouse IgG1 (InVitrogen) was used to measure signal. Anti-tetanus antibody TetE3 was used to make a standard line (AbCam). The specific ELISA protocol, using Covalink 96-well plates (Nunc A/S, Roskilde, Denmark) was identical to our previously described methods [11].

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Orit Kollet and Tsvee Lapidot of the Weizmann Institute of Science (Rehovot, Israel) for their assistance in titrating the dosage of G-CSF suitable for these experiments.

REFERENCES

1. Noguchi, M., et al., *Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans*. *Cell*, 1993. **73**(1): p. 147-57.
2. Fischer, A., *Thirty years of bone marrow transplantation for severe combined immunodeficiency*. *N Engl J Med*, 1999. **340**(7): p. 559-61.
3. Antoine, C., et al., *Long-term survival and transplantation of haemopoietic stem cells for immunodeficiencies: report of the European experience 1968-99*. *Lancet*, 2003. **361**(9357): p. 553-60.
4. Beatty, P.G., et al., *Marrow transplantation from related donors other than HLA-identical siblings*. *N Engl J Med*, 1985. **313**(13): p. 765-71.
5. Gaspar, H.B., et al., *Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector*. *Lancet*, 2004. **364**(9452): p. 2181-7.
6. Hacein-Bey-Abina, S., et al., *Efficacy of gene therapy for X-linked severe combined immunodeficiency*. *N Engl J Med*, 2010. **363**(4): p. 355-64.
7. Rocha, V., et al., *Graft-versus-host disease in children who have received a cord-blood or bone marrow transplant from an HLA-identical sibling*. *Eurocord and International Bone Marrow Transplant Registry Working Committee on Alternative Donor and Stem Cell Sources*. *N Engl J Med*, 2000. **342**(25): p. 1846-54.
8. Cavazzana-Calvo, M., et al., *Gene therapy for primary immunodeficiencies: part 1*. *Curr Opin Immunol*, 2012. **24**(5): p. 580-4.
9. Aiuti, A., et al., *Gene therapy for immunodeficiency due to adenosine deaminase deficiency*. *N Engl J Med*, 2009. **360**(5): p. 447-58.
10. Gaspar, H.B., et al., *Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency*. *Sci Transl Med*, 2011. **3**(97): p. 97ra79.
11. Huston, M.W., et al., *Correction of murine SCID-X1 by lentiviral gene therapy using a codon-optimized IL2RG gene and minimal pretransplant conditioning*. *Mol Ther*, 2011. **19**(10): p. 1867-77.
12. Petit, I., et al., *G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4*. *Nat Immunol*, 2002. **3**(7): p. 687-94.
13. Shier, L.R., et al., *Differential effects of granulocyte colony-stimulating factor on marrow- and blood-derived hematopoietic and immune cell populations in healthy human donors*. *Biol Blood Marrow Transplant*, 2004. **10**(9): p. 624-34.
14. Bensinger, W.I., et al., *Treatment of normal donors with recombinant growth factors for transplantation of allogeneic blood stem cells*. *Bone Marrow Transplant*, 1996. **17 Suppl 2**: p. S19-21.
15. Chen, J., et al., *Mobilization as a preparative regimen for hematopoietic stem cell transplantation*. *Blood*, 2006. **107**(9): p. 3764-71.
16. Giri, N., et al., *The effects of SCF/G-CSF prestimulation on radiation sensitivity and engraftment in non-myeloablated murine hosts*. *Exp Hematol*, 2001. **29**(6): p. 779-85.
17. Tomita, Y., D.H. Sachs, and M. Sykes, *Myelosuppressive conditioning is required to achieve engraftment of pluripotent stem cells contained in moderate doses of syngeneic bone marrow*. *Blood*, 1994. **83**(4): p. 939-48.
18. Armitage, J.O., *Bone marrow transplantation*. *N Engl J Med*, 1994. **330**(12): p. 827-38.
19. Cavazzana-Calvo, M., et al., *Gene therapy for severe combined immunodeficiency*. *Annu Rev Med*, 2005. **56**: p. 585-602.
20. Aiuti, A., et al., *The chemokine SDF-1 is a chemoattractant for human CD34+ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34+ progenitors to peripheral blood*. *J Exp Med*, 1997. **185**(1): p. 111-20.
21. Whetton, A.D. and G.J. Graham, *Homing and mobilization in the stem cell niche*. *Trends Cell Biol*, 1999. **9**(6): p. 233-8.

22. Elfenbein, G.J. and R. Sackstein, *Primed marrow for autologous and allogeneic transplantation: a review comparing primed marrow to mobilized blood and steady-state marrow*. *Exp Hematol*, 2004. **32**(4): p. 327-39.
23. Stewart, F.M., et al., *Lymphohematopoietic engraftment in minimally myeloablated hosts*. *Blood*, 1998. **91**(10): p. 3681-7.
24. Link, D.C., *Mechanisms of granulocyte colony-stimulating factor-induced hematopoietic progenitor-cell mobilization*. *Semin Hematol*, 2000. **37**(1 Suppl 2): p. 25-32.
25. Cynshi, O., et al., *Reduced response to granulocyte colony-stimulating factor in W/W^v and Sl/Sld mice*. *Leukemia*, 1991. **5**(1): p. 75-7.
26. Buckley, R.H., *Molecular defects in human severe combined immunodeficiency and approaches to immune reconstitution*. *Annu Rev Immunol*, 2004. **22**: p. 625-55.
27. Weerkamp, F., et al., *Human thymus contains multipotent progenitors with T/B lymphoid, myeloid, and erythroid lineage potential*. *Blood*, 2006. **107**(8): p. 3131-7.
28. DiPersio, J.F., et al., *Plerixafor and G-CSF versus placebo and G-CSF to mobilize hematopoietic stem cells for autologous stem cell transplantation in patients with multiple myeloma*. *Blood*, 2009. **113**(23): p. 5720-6.
29. Kustikova, O., et al., *Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking*. *Science*, 2005. **308**(5725): p. 1171-4.
30. Thornhill, S.I., et al., *Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency*. *Mol Ther*, 2008. **16**(3): p. 590-8.
31. Markiewicz, S., et al., *Tissue-specific activity of the gammac chain gene promoter depends upon an Ets binding site and is regulated by GA-binding protein*. *J Biol Chem*, 1996. **271**(25): p. 14849-55.
32. Zhang, F., et al., *A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors*. *Mol Ther*, 2010. **18**(9): p. 1640-9.
33. Almarza, E., et al., *Correction of SCID-X1 using an enhancerless Vav promoter*. *Hum Gene Ther*, 2011. **22**(3): p. 263-70.
34. Knight, S., et al., *Safer, silencing-resistant lentiviral vectors: optimization of the ubiquitous chromatin-opening element through elimination of aberrant splicing*. *J Virol*, 2012. **86**(17): p. 9088-95.
35. Greene, M.R., et al., *Transduction of Human CD34(+) Repopulating Cells with a Self-Inactivating Lentiviral Vector for SCID-X1 Produced at Clinical Scale by a Stable Cell Line*. *Hum Gene Ther Methods*, 2012. **23**(5): p. 297-308.
36. Follenzi, A. and L. Naldini, *HIV-based vectors. Preparation and use*. *Methods Mol Med*, 2002. **69**: p. 259-74.
37. Wognum, A.W., et al., *Stimulation of mouse bone marrow cells with kit ligand, FLT3 ligand, and thrombopoietin leads to efficient retrovirus-mediated gene transfer to stem cells, whereas interleukin 3 and interleukin 11 reduce transduction of short- and long-term repopulating cells*. *Hum Gene Ther*, 2000. **11**(15): p. 2129-41.
38. Pujal, J.M. and D. Gallardo, *PCR-based methodology for molecular microchimerism detection and quantification*. *Exp Biol Med (Maywood)*, 2008. **233**(9): p. 1161-70.

Met Tyr Xle Ala Cys
 Met Tyr Xle Ala Cys
 Cys Ala Lys Xle Ala Cys Asn Pyl Thr
 Ala Lys Ser Cys Lys Pyl Thr
 Xle Arg Glu Asn Ser SecSer
 PylAsn Ser SerThr LeuLeulle IleAsaAsn
 Lys Leu Glu Asn TrpIle LeuLeu His
 Cys Glu Ser GGAC GCC GGCATGCC TrpIle His Sec
 Ala Xle Thr AGCAGGTACGCGAGA His Asn Ser
 GluArg Thr AGGCTAGCTAG PylThr
 CAGGCTCGA Arg His
 CAGTGGGTAT PylAsn Arg AlaArg
 GTGAGCTTGCTC Thr Ile ArgIle
 Ser Glu Asn Pyl Thr His
 Sec Thr Pyl Thr His
 His SerSec

GGAC GCC GGCATGCC
 AGCAGGTACGCGAGA
 AGGCTAGCTAG
 CAGGCTCGA
 CAGTGGGTAT
 GTGAGCTTGCTC
 GATAGCGGGTGA
 TGCTAGCCCGCC
 GGGAAATGCAGG
 TACGCTTGCTAGA
 CCGCAGGCATGC
 CTGTATGCTTGCTC
 01001101011000010111
 00100 1110011011010000
 01100 001011011000110
 1100 0010000001010111
 011 010010110110001101
 10 0011010010110000101
 101101001 00000010010
 00011101 01011100110111010
 0011011 11011011100010110000
 100000 01010
 00001 1010
 000 100100



Low leukemia incidence in *Il2rg*^{-/-} mice treated with lentiviral vector gene therapy

Marshall W. Huston¹, Elnaz Farahbakhshian¹, Helen de Boer¹, Ali Nowrouzi², Martijn H. Brugman^{1,3}, Manfred Schmidt², Christof von Kalle², Niek P. van Til¹ and Gerard Wagemaker¹

¹Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands

²Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center (DKFZ), Heidelberg, Germany

³present address: Leiden University Medical Center, Department of Immunohematology and blood transfusion, Leiden, the Netherlands.

Manuscript in preparation

ABSTRACT

The efficacy of *ex vivo* hematopoietic stem cell gene therapy in treating severe combined immunodeficiency diseases (SCIDs) has been demonstrated in a number of clinical trials conducted in the early 2000s using gammaretroviral vectors. However the success of these early trials was tempered by the discovery of vector-derived insertional oncogenesis in some of the patients. HIV-1 derived lentiviral vectors have the potential to reduce the risk of genotoxicity due to their distinct integration profile compared to gammaretroviral vectors, but have yet to be tested in the clinic. We generated several self-inactivating (SIN) lentiviral vectors containing the IL2RG gene driven by promoters of varying strength and tested these vectors in a large number of *Il2rg*^{-/-} mice. During the process of monitoring these mice, we discovered a total of 14 leukemias in 212 animals transplanted with lentiviral vector-treated lineage negative bone marrow cells derived from *Il2rg*^{-/-} or wild type mice. The majority of leukemias observed were CD4⁺CD8⁺, but CD8⁺, CD4⁺, CD11b⁺, CD19⁺ and Sca1⁺ leukemias were also found. qPCR and integration site analysis suggests that vector-mediated insertional mutagenesis played a role in some of the leukemias found in vector-treated mice, while others contained no obvious oncogenic integration. Similar leukemia frequencies were seen in mice given IL2RG vectors, GFP vectors or untransduced cells and no increased risk of leukemogenesis was seen when using the viral SF promoter compared to a eukaryotic promoter. Three leukemias were also found in *Il2rg*^{-/-} mice transplanted with untransduced wild type cells. Considering the tumor-prone background of the *Il2rg*^{-/-} mice used in these experiments, these observations suggest that self-inactivating lentiviral vector HSC gene therapy has a low risk of genotoxicity.

INTRODUCTION

Gammaretroviral vector (RV) gene therapy has emerged as a treatment option for primary immunodeficiency, and in particular severe combined immunodeficiency (SCID) patients lacking a suitable bone marrow donor. The efficacy of these vectors to transduce hematopoietic stem cells (HSCs) *ex vivo* with a MFG Moloney-derived retrovirus vector and successfully reverse the SCID phenotype was first demonstrated in seminal clinical studies for SCID-X1 and ADA-SCID at the turn of the century [1-3]. The majority of patients in these trials recovered functional T and NK cell populations and healthy immunoglobulin levels. However, of the 20 SCID-X1 patients treated with gammaretroviral vectors, five later developed leukemia-like symptoms[4-6]. Careful investigation of these leukemias revealed that the vector had played a role in their development via insertional mutagenesis[7-9]. These leukemic events were treated successfully by chemotherapy in four of the patients without reducing the effectiveness of the gene therapy, while one patient succumbed as a result of treatment[5, 6, 10]. Evidence of clonal expansion was also seen in RV clinical trials for X-linked chronic granulomatous disease (CGD)[11] and X-linked primary immunodeficiency Wiskott-Aldrich syndrome (WAS)[12]. Conversely, none of the patients in the ADA-SCID trial have developed leukemia-like symptoms.

Further study identified a number of risk factors inherent to the clinical RV vectors used in the SCID-X1, ADA-SCID, CGD and WAS trials, including a strong viral promoter, lack of insulator elements and transcription start site (TSS) integration bias of the gammaretroviral vector itself [13, 14]. A redesigned gammaretroviral vector with an improved safety profile is currently in use in a multi-center clinical trial for SCID-X1. HIV-1 derived lentiviral vectors (LV) have been suggested as an alternative to gammaretroviral vectors for HSC gene therapy applications. Lentiviral vectors have several advantages over gammaretroviral vectors, including reduced silencing [15] and the ability to transduce quiescent cells. Unlike gammaretroviruses which have a preference towards integrating upstream of transcription start sites, lentiviruses prefer to integrate into the intronic DNA regions of active genes [14, 16, 17], which would in theory reduce the risk of genotoxicity. Further improvements by using the self-inactivating configuration and a eukaryotic promoter may even further reduce the risk of insertional oncogenesis[18-21]. Currently, there are LV clinical trials ongoing for X-linked adrenoleukodystrophy[22], β -thalassemia[23], Wiskott-Aldrich syndrome[24, 25] and metachromatic leukodystrophy[26]. The β -thalassemia trial reported expansion of a dominant myeloid-biased clone with a vector insertion in the third intron of the *HMG2A* gene. At the time of this writing, the dominant clone has been stable at approximately 10% of the clonal pool, shows no signs of additional expansion and has not disrupted haematopoietic homeostasis in the patient. Since the disease and genetic background can play an important role in the initiation of vector-induced leukemia[6], we have previously investigated the efficacy of lentiviral *IL2RG* gene therapy in an *in vivo* setting[19] by transplanting transduced and non-transduced lineage negative (Lin-) bone marrow cells into *Il2rg*^{-/-} mice for long-term follow-up and retransplantation

into secondary recipients for a maximum up to 21 months. In this paper we assess the risk of oncogenesis in this disease model. By comparing the relative rates of leukaemogenesis, vector copy numbers and integration sites between experimental groups, a rough estimate of the risk of using lentiviral *IL2RG* gene therapy in the *Il2rg*^{-/-} mouse model can be assessed.

RESULTS

Experimental overview

This manuscript provides a summary of the results of 11 separate gene therapy experiments, involving 239 *Il2rg*^{-/-} primary transplants and 118 secondary transplants given Lin-bone marrow cells after irradiation conditioning (Table 1). Male donor cells were transplanted into female recipient mice in order to track donor cell chimerism.

Table 1: Overview of *Il2rg*^{-/-} transplantation and LV gene therapy experiments

Exp.	No. of 1° mice	No. of 2° mice	No. of Lin- cells transplanted	No. of mice transplanted after LV transduction	No. of mice transplanted without transduction	No. of leukemias observed	vectors used (no. of leukemias observed)
1	10	12	4x10 ⁵	10	0	1	SF.IL2RG(1) SF.GFP
2	15	24	5x10 ⁵	15	0	0	SF.coIL2RG γcPr.coIL2RG, γcPr.GFP
3	50	9	4x10 ⁵	50	0	4	SF.IL2RG(1) SF.poly(2) PGK.coIL2RG(1) EFS.coIL2RG
4	19	14	5x10 ⁵	13	6	0	γcPr.coIL2RG
5	19	16	5x10 ⁵	15	4	3	γcPr.coIL2RG(2)
6	17	0	3x10 ⁴	0	17	2	n/a
7	24	20	1x10 ⁵	24	0	2	γcPr.coIL2RG(2) SF.poly
8	21	10	3x10 ⁵	21	0	1	γcPr.coIL2RG PGK.coIL2RG γcPr.GFP(1)
9	41	13	3x10 ³	41	0	4	γcPr.coIL2RG PGK.GFP(4)
10	12	0	3x10 ⁵	12	0	0	PGK.GFP
11	11	0	3x10 ⁵	11	0	0	PGK.GFP

Abbreviations: Exp., experimental ID reference number; 1°, primary transplant; 2°, secondary transplant; Lin-, lineage negative cells (bone marrow derived)

The cells transplanted included untreated wild type Lin⁻ cells (27 mice), *Il2rg*^{-/-} Lin⁻ cells transduced with lentiviral vectors containing no transgene or a GFP gene (95 mice), and *Il2rg*^{-/-} Lin⁻ cells transduced with lentiviral vectors containing the wild type human *IL2RG* cDNA or a codon optimized version (*coIL2RG*) (117 mice, Table 2). The number of Lin⁻ cells transplanted per mouse ranged from 3x10⁵ to 1x10⁵, with the exception of one titration experiment where 10,000 to 300 wild type Lin⁻ cells were transplanted. Transduction efficiency varied based on the HeLa MOI of the vector used (MOI 1 to 10), resulting in 60% to 90% of the total cell pool containing at least 1 viral integration. Mice were monitored monthly via peripheral blood cell counts and FACS analysis for 6 to 8 months, and subsequently sacrificed for FACS analysis of bone marrow (BM), spleen and thymus. Mice which fell ill during the course of the experiment were sacrificed and potential leukemic phenotypes were determined via flow cytometry analysis and confirmed via secondary transplants. Bone marrow and spleen DNA from leukemic secondary transplants were used to study potential vector influence on oncogenesis via qPCR and LAM-PCR investigation of integration sites. A total of 17 leukemias were confirmed in the course of these experiments, including 7 in mice treated with *IL2RG* vectors (out of 117 mice, a 6.0% leukaemia rate), 7 in mice treated with GFP vectors or vectors containing no transgene (7.4%) and 3 in mice given untransduced wild type Lin⁻ cells (11.1%).

Table 2: Overview of lentiviral vectors used in *Il2rg*^{-/-} transplantation experiments

vector	no. of mice	Lin ⁻ donor cells	observation time (months)		no. of leukemias seen
			1° transplants	2° transplants	
EFS.coIL2RG	6	<i>Il2rg</i> ^{-/-}	8	x	0
PGK.coIL2RG	12	<i>Il2rg</i> ^{-/-}	8	9	1
PGK.GFP	47	BALB/c	7	12	4
SF.coIL2RG	4	<i>Il2rg</i> ^{-/-}	6	9	0
SF.IL2RG	16	<i>Il2rg</i> ^{-/-}	6	9	2
SF.IL2RG	10	BALB/c	6	9	0
SF.poly	14	<i>Il2rg</i> ^{-/-}	7	x	1
SF.poly	15	BALB/c	7	x	0
SF.GFP	4	<i>Il2rg</i> ^{-/-}	6	9	0
γcPR.coIL2RG	69	<i>Il2rg</i> ^{-/-}	8	12	4
γcPR.GFP	4	<i>Il2rg</i> ^{-/-}	6	9	0
γcPR.GFP	11	BALB/c	7	12	2
no vector	27	BALB/c	7	9	3

Abbreviations: 1°, primary transplant; 2°, secondary transplant; poly, vector containing a polylinker sequence but no transgene

Leukemias attributed to viral vector integrations: SF promoter

63 mice were transplanted with cells transduced by vectors containing the SF promoter (Table 3a). Two leukemias were seen in mice given the SF.IL2RG vector in separate experiments. The first mouse fell ill 21 weeks after transplantation with an extremely high white blood cell (WBC) count of $243 \times 10^6/\text{ml}$. Flow cytometric analysis of BM in the primary mouse and three secondary transplants confirmed that the leukemia was Sca-1⁺, with 93% of BM cells having this phenotype (Table 4). qPCR analysis of BM cell DNA revealed a high Y (donor)-chimerism (95% in secondary recipients) and an average of 2.2 integrations per BM cell. LAM-PCR and sequencing of the flanking genomic sequence revealed a dominant integration site on chromosome 13 at base 83644267 (Supplemental Table 1). This integration is situated 1.1kb from the transcription start site (TSS) of gene *Mef2c* (myocyte enhancer factor 2C), in the first intron. *Mef2c* is a known oncogene and is the likely initiator of leukemogenesis in this mouse. A second mouse treated with SF.IL2RG fell ill with a WBC count of $80 \times 10^6/\text{ml}$ with a peripheral blood CD4⁺CD8⁺ leukemia, with 87% of bone marrow cells having this phenotype. qPCR revealed that the leukemia was donor-derived (BM Y-chimerism = 80%) and had an average of 2.8 vector integrations per BM cell. LAM-PCR and sequencing showed a prominent IS on chromosome 5 at base 24812116, located in an intron of oncogene *Mll3* (myeloid/lymphoid or mixed-lineage leukemia 3). Additional integration sites were located on chromosomes 2 and 10 (Supplemental Table 1). One mouse given *Il2rg*^{-/-} cells after transduction with SF.poly, a vector containing the SF promoter and a polylinker sequence but no transgene, developed a donor-derived CD11b⁺ leukemia. Three prominent integration sites were found in the leukemic clones (2.9 vector integrations per BM cell) on chromosomes 1, 11 and 16. No known oncogenes were located within 500kb of these IS, but the integration on chromosome 1 was located in the first intron of gene *Pyhin1* (pyrin and HIN domain family, member 1), which has been suggested as a tumor suppressor gene[27]. Another mouse transplanted with wild type Lin⁻ cells after SF.poly transduction became sick with a high peripheral white blood cell count ($137 \times 10^6/\text{ml}$) and was found to have a CD4⁺CD8⁺ leukemia. Although Y-chimerism levels in this mouse were high, qPCR and LAM-PCR of BM cells revealed no vector integrations, suggesting that this oncogenic event was not vector-initiated.

Table 3: Experimental *Il2rg*^{-/-} gene therapy mice, sorted by vector

a) Sorted by vector promoter.

vector	no. mice treated	no. leukemias seen	leukemia frequency	overall frequency
EFS.coIL2RG	6	0	0.0%	0.0%
PGK.coIL2RG	12	1	8.3%	
PGK.GFP	47	4	8.5%	6.3%
SF.coIL2RG	4	0	0.0%	
SF.IL2RG	26	2	7.7%	
SF.poly	29	2	6.9%	
SF.GFP	4	0	0.0%	6.0%
γ cPR.coIL2RG	69	4	5.8%	
γ cPR.GFP	15	1	6.7%	

b) Sorted by vector transgene.

vector	no. mice treated	no. leukemias seen	leukemia frequency	overall frequency	frequency, all transplanted mice
EFS.coIL2RG	6	0	0.0%	6.0%	7.1%
PGK.coIL2RG	12	1	8.3%		
SF.coIL2RG	4	0	0.0%		
SF.IL2RG	26	2	7.7%		
γ cPR.coIL2RG	69	4	5.8%		
SF.poly	29	2	6.9%		
SF.GFP	4	0	0.0%		
PGK.GFP	47	4	8.5%	7.4%	
γ cPR.GFP	15	1	6.7%	11.1%	
no vector	27	3	11.1%		

Leukemias attributed to viral vector integrations: eukaryotic promoters

84 mice were transplanted with Lin⁻ cells transduced by vectors driven by the γ cPr promoter, a 1.1kb section of the native *IL2RG* promoter region. Four leukemias were seen in mice given the γ cPr.coIL2RG vector: all of these leukemias were donor-derived and had between 1.4 and 1.8 vector copies per bone marrow cell (Supplemental Table 1). One mouse had a dominant IS on chromosome 6, 51kb upstream of the TSS of proto-oncogene *Ccnd2* (cyclin D2). Another mouse had a prominent IS on chromosome 4, 74kb upstream of the TSS of oncogene *Pax5* (paired box gene 5) and a second IS on chromosome 10. A third mouse had a prominent IS located in a gene-dense region on chromosome 2 in the first intron of gene *St6galnac4*. An additional IS was found on chromosome 12, 267kb downstream of gene *Il12b* (interleukin 12b). No known oncogenes were found within 500kb of either integration site. The last γ cPr.coIL2RG which developed leukemia had

a single dominant integration site 69kb upstream of transcriptional repressor gene *Kctd1* (potassium channel tetramerisation domain containing 1)[28] on chromosome 18.

One leukemia was found mice given wild type Lin- cells after γ Pr.GFP transduction. The leukemic clones were donor-derived and CD4⁺CD8⁺ GFP⁻, with a dominant IS on chromosome 16, in the first intron of gene *Rbfox1* (RNA binding protein fox-1 homolog). No known oncogenes were found within 500kb of the IS flanking *Kctd1* or *Rbfox1*.

59 mice were transplanted with *Il2rg*^{-/-} cells transduced with vectors driven by the PGK promoter. One donor-derived leukemia was found in a mouse treated with the PGK.coIL-2RG vector. The leukemic clone contained an average of 1.2 vector copies per cell. There was one dominant IS on in a gene-dense region of chromosome 11, located in intron 1 of P4hb (prolyl 4-hydroxylase, beta polypeptide). The nearest known oncogene adjacent to this IS is *Aspscr1* (alveolar soft part sarcoma chromosome region, candidate 1), 113kb upstream.

Four leukemias were found in mice treated with PGK.GFP, all in the secondary transplants of a single experiment. This experiment involved transplanting a low number of transduced wild type cells (10,000 to 300 Lin- cells per mouse) into 6 Gy conditioned *Il2rg*^{-/-} recipients following seven days of expansion *in vitro* with protein Angiopoietin-like 3. No primary mice in this experiment developed any hematopoietic abnormalities during 7 months of monitoring. However, multiple secondary transplants developed CD4⁺CD8⁺ leukemias 8 to 24 weeks after receiving BM cells from the primary mice. Three of the leukemias were GFP⁺ and shared an integration sites on chromosome 5 and chromosome 10, flanking genes *Phf2* (putative homeodomain transcription factor 2) and *Shprh* (SNF2 histone linker PHD RING helicase), respectively. No known oncogenes were found within 500kb of these IS, however, *Shprh* has been implicated as a tumor suppressor gene[29]. The IS on chromosome 10 lies in an intron between the 17th and 18th exons of *Shprh*. The final leukemia seen in this experiment was also CD4⁺CD8⁺ GFP⁺ and contained integration sites on chromosomes 4, 5 and 11. The IS on chromosome 11 is located within 100kb of several genes involved in transcriptional regulation or repression, including Sox15 and Zbtb4[30, 31].

Leukemias seen in mice transplanted with Lin- cells without transduction

Three leukemias were observed in *Il2rg*^{-/-} mice transplanted with untransduced wild type BALB/c Lin-cells. Two leukemias were discovered in mice transplanted with the same pool of cells at 27 weeks post-transplant. These mice did not have elevated WBC counts but one did have an enlarged spleen (spleen weights of 860mg and 180 mg). Their leukemias had different phenotypes: the mouse with the enlarged spleen had a CD4⁺CD8⁺ (75% of total BM cells) leukemia, while the other mouse's leukemic phenotype was CD8⁺CD4⁻ (84% of total BM cells) with a small subpopulation which was CD4⁺CD8⁺ (4% of BM cells). These divergent phenotypes were confirmed in secondary transplants. qPCR for Y-chimerism revealed that both leukemias were donor-derived. In another experiment,

total BM cells from a healthy *Il2rg*^{-/-} recipient of wild type cells were transplanted into a secondary *Il2rg*^{-/-} recipient. This mouse fell ill 40 weeks after transplant and had to be sacrificed. White blood cell count was within normal range ($17 \times 10^6/\text{ml}$), but the spleen

Table 4: Overview of leukemia phenotypes seen in *Il2rg*^{-/-} mice during Lin- cell transplantation and gene therapy experiments

ID#	trans-plant type	vector	donor cells	TBI dose	onset of illness (weeks)	PB WBC count	leukemia phenotype	% phenotype in BM	Y-chimerism (BM)
1	2°	PGK.coIL2RG	Il2rg--	6 Gy	66	67.8	CD4CD8	89	86%
2	2°	PGK.GFP	BALB/c	6 Gy	44	3.2	CD4CD8	79	87%
3	2°	PGK.GFP	BALB/c	6 Gy	68	4.5	CD4CD8	44	71%
4	2°	PGK.GFP	BALB/c	6 Gy	68	2.5	CD4CD8	37	71%
5	2°	PGK.GFP	BALB/C	6 Gy	52	20.5	CD4CD8	64	70%
6	1°	SF.II2RG	Il2rg--	6 Gy	21	243	Sca-1	93	95%
7	1°	SF.II2RG	Il2rg--	6 Gy	25	80	CD4CD8	87	62%
8	1°	SF.poly	BALB/C	6 Gy	35	153	CD11b	96	72%
9	1°	SF.poly	Il2rg--	6 Gy	22	137	CD4CD8	82	75%
10	1°	γ cPR.coIL2RG	Il2rg--	2 Gy	30	29.8	CD4CD8	57	95%
11	1°	γ cPR.coIL2RG	Il2rg--	2 Gy	33	1.7	CD4CD8	92	57%
12	1°	γ cPR.coIL2RG	Il2rg--	2 Gy	28	65.9	CD4	86	63%
13	1°	γ cPR.coIL2RG	Il2rg--	2 Gy	28	71	CD4CD8	94	82%
14	1°	γ cPR.GFP	BALB/c	0 Gy	35	28.8	CD4CD8	83	88%
15	2°	n/a	BALB/c	6 Gy	67	16.8	CD19	96	95%
16	1°	n/a	BALB/c	2 Gy	27	2.9	CD8	84	97%
17	1°	n/a	BALB/c	2 Gy	27	16.5	CD4CD8	75	12%

Abbreviations: 1°, primary transplant; 2°, secondary transplant; TBI, total body irradiation; PB, peripheral blood; WBC, white blood cell

was enlarged (590 mg) and a B cell leukemia was found during FACS analysis (96% of BM cells were B220⁺CD19⁺) were found in this secondary recipient, 65 weeks after the initial transplantation into the primary *Il2rg*^{-/-} mouse. 50% of the leukemic clones were IgM⁺, but none were IgD⁺. BM Y-chimerism levels were high, suggesting that the leukemic clone was donor-derived.

DISCUSSION

Risks of SF vector and IL2RG transgene on insertional oncogenesis

Following the initial reports of leukemia in the 2000 Paris and London gene therapy trials for SCID-X1 [5, 6], the MFG Moloney-gammaretrovirus derived vector used in these trials was implicated as having a role in oncogenesis. The case against viral enhancer/promoters was made due to the strength of activation of nearby genes and its possible influence on the integration site profile [13, 14]. *In vitro* immortalisation assays seemed to confirm the higher genotoxic risks of viral promoters, such as the SF promoter, compared to weaker eukaryotic promoters [13].

In our experiments with SIN lentiviral vector gene therapy for murine SCID-X1, the choice of promoter seemed to have no effect on the rate of leukemogenesis. Mice treated with the SF promoter had a leukemia rate of 6.3%, similar to the rates seen in eukaryotic promoters PGK (8.5%) and γ cPr (6.0%). There was no significant difference in the rate of leukemogenesis between the various promoter elements.

The IL2RG gene itself was implicated as potentially having a role in oncogenesis [32], although this link has since been disputed [33-35] and a mouse strain expressing high levels of *Il2rg* showed no predisposition to tumor development [36]. Further investigation in mice has suggested that *Lmo2*, the principle oncogene in the clinical SCID-X1 trials, and *Il2rg* are often deregulated together in leukemias, and that there is strong selection for *Il2rg* deregulation in *Lmo2*-initiated instances of T cell leukemia [37]. However, deregulation of these two genes is insufficient to trigger oncogenesis: leukemia onset requires other contributing mutations including overexpression of other oncogenes and downregulation of tumor suppressor genes. In the *Il2rg*^{-/-} murine gene therapy experiments conducted here, the rates of oncogenesis for vectors containing the *IL2RG* transgene versus a “neutral” *GFP* transgene or no transgene at all were very close (6.0% to 7.4%, respectively) and similar to the background rate of leukemiogenesis in animals transplanted with wild type BALB/c cells without viral transduction (11.1%).

Thus, no additional risk of generating leukemias was seen in vectors containing either the SF promoter or the *IL2RG* transgene gene in our lentiviral gene therapy experiments in *Il2rg*^{-/-} mice. This may be due to the presence of additional safety elements such as internal promoters and the deletion of enhancers elements in our 3rd generation SIN lentiviral vectors, features that were not present in the gammaretroviral vectors used in the initial clinical trials. Previous studies conducted with gammaretroviral and lentiviral vectors demonstrated that transcriptionally active LTRs, such as those used in the initial SCID-X1 clinical trials, are major determinants of genotoxicity [38]. The same studies suggested that lentiviral vectors require a 10-fold higher viral load to trigger leukemogenesis than gammaretroviral vectors, likely due to the preference of gammaretroviruses to integrate into the promoter regions of active genes and into genes involved in cell growth and cancer [9].

Effects of *Il2rg*^{-/-} background and irradiation conditioning on leukemogenesis

While it has been previously reported that *Il2rg*^{-/-} mice are highly susceptible to MLV vector leukemogenesis, the rates of leukemogenesis in our vector-treated mice were not higher than those seen in mice transplanted after mock transduction (Table 2) or background levels of tumor formation in untreated *Il2rg*^{-/-} mice (data not shown). Aside from the viral vector, several other factors could trigger oncogenesis in gene therapy protocols. The majority of recipient mice in our experiments received either 2 Gy or 6 Gy TBI prior to cell transplantation, sufficient to damage nuclear DNA and trigger leukemogenesis. However, leukemias generated due to irradiation would be of recipient origin and have low levels of Y(donor)-chimerism. Naturally, they would also have no evidence of viral vector integration. Out of the 17 leukemias described here, only one had sufficiently low Y chimerism to be considered of recipient origin. The remaining 16 leukemias are clearly donor-derived.

Another factor which could trigger oncogenesis is the proliferative stress the donor HSC are subjected to after transplantation as they repopulate the hematopoietic system: in this case, the leukemias would be expected to be donor derived and have high Y-chimerism levels. Two donor-derived leukemias were seen in mice transplanted with untransduced wild type Lin⁻ cells. As no viral vector was administered to these cells, proliferative stress is the most likely culprit for leukemogenesis. Similarly, the donor leukemia in a mouse treated with the LV.SF.poly vector, which had no vector DNA present in the (donor-derived) leukemic clones, was also probably due to proliferative stress triggering an oncogenic event in an untransduced donor HSC.

The potential for HSC transplantation to trigger leukemogenesis in *Il2rg*^{-/-} mice makes it difficult to state with confidence that insertional mutagenesis is responsible for the leukemias seen in lentiviral treated mice, even when the leukemic clones contain vector integrations near proto-oncogenes. In some instances (such as the Sca-1⁺ leukemia containing a dominant integration site in the first intron of oncogene *Mef2c*) it is very likely that the vector played a key role in oncogenesis. However, there were several leukemias in vector-treated mice had no dominant integration site within 500kb of any known oncogenes, or tumor suppressor genes or transcription repressor genes. Some or all of these leukemic clones may have been caused by mutations generated during proliferative stress without vector involvement. This does not exclude the possibility that these dominant IS may lie near an as-yet unidentified oncogene. It is also possible that in some samples the oncogenic IS was missed by the LAM-PCR method used here, which is only able to detect 60-70% of all IS clones and covers at most offers 80% genome coverage[39, 40].

The large majority of the leukemias seen in these mice derived from T cell lineages (14 out of 17, including one each of CD4⁺ and CD8⁺ and 12 that were CD4⁺CD8⁺). This was highly significant and suggests that the T cell lineages were especially prone to oncogenesis in these transplantation experiments. It is noteworthy that all 5 leukemias seen in the human SCID-X1 gene therapy trials were T-ALL, as was the leukemia seen in the

trial for Wiskott-Aldrich syndrome (WAS)[41]. Both WAS and SCID-X1 are T immunodeficiencies. Similarly, the two adverse events seen in the chronic granulomatous disease (CGD) clinical trial were myelodysplasias[42]. It is possible that corrected hematopoietic progenitor cells belonging to the deficient cell type are subjected to additional proliferative stress and are therefore more susceptible to transformative events, particularly when this stress is combined with disruptive viral integrations near key proto-oncogenes.

Overall, the analysis of these experiments suggests that the relative risk of developing leukemia in gene therapy treated *Il2rg*^{-/-} mice using 3rd generation SIN lentiviral vectors is unaffected by the use of the SF promoter at low vector copy numbers, and further argues against the idea that the *IL2RG* gene is oncogenic. The use of 3rd gen SIN LV vectors for treatment of murine SCID-X1 is relatively benign and does not significantly induce insertional mutagenesis when used at low copy number and when transplanting a modest number of lineage negative cells. These vectors should be considered for future SCID clinical trials as they may offer a significant benefit to safety without reducing the efficacy of the gene therapy treatment.

METHODS

Mice

BALB/c and *Il2rg*^{-/-} mice were bred in the Experimental Animal Center of Erasmus MC. *Il2rg*^{-/-} single KO mice with a targeted γ c deletion were derived from a 10th generation backcross of BALB/c Rag2^{-/-}/ γ c^{-/-} mice, kindly provided by Dr. H Spits[43], with syngenic BALB/c wild type mice. All mice were used at 6 to 10 weeks of age at the time of transplant and were maintained in specified pathogen free conditions. Experiments were approved by the institutional Animal Ethical Committee of Erasmus MC in accordance with legislation in the Netherlands.

Production of lentiviral particles and titration

Third generation self-inactivating (SIN) LV incorporating the *IL2RG* cDNA were constructed using the HIV-1 vector backbone[44, 45] with SF promoter and a modified woodchuck posttranslational regulatory element (bPRE4*)[46]. Codon-optimized human *IL2RG* sequence predicted by GeneOptimizer[®] software (GeneArt AG, Regensburg, Germany). A 1.1 kb fragment of the human *IL2RG* promoter (γ cPr) and other promoter elements were cloned into LV-SF-*IL2RG* as described previously [19].

LVs were produced by standard calcium phosphate transfection of HEK 293T cells[47] with the packaging plasmids pMDL-g/pRRE, pMD2-VSVg and pRSV-Rev. LV were concentrated by ultracentrifugation at 20,000 r.p.m. for 2 hours at 4°C and stored at -80°C. Viral vector titration was performed on HeLa cells by serial dilutions to determine transducing units (HeLa transducing units, HTU) per mL. The transduced cells were harvested on day 11, DNA and total RNA were purified and the copy number per cell was determined by real-time quantitative polymerase chain reaction (qPCR).

Lentiviral vector gene therapy protocols

Lentiviral gene therapy vectors containing the native human *IL2RG* cDNA, or a codon optimized version (*coIL2RG*), were created under the control of one of the following promoters: spleen focus forming virus promoter (SF), phosphoglycerate kinase promoter (PGK), elongation factor-1 alpha promoter (EFS) or a 1.1kb section of the native *IL2RG* promoter region (γ cPr). Lineage negative (lin-) *Il2rg*^{-/-} bone marrow cells from male donors were transduced overnight (18 hours transduction time) with these vectors at HeLa MOI 1, 3 or 10 and then injected into female *Il2rg*^{-/-} mice after irradiation conditioning, as previously described[19]. Mice were monitored for 6 to 8 months for reconstitution of lymphocyte populations and signs of hematopoietic abnormalities. Subsequently, mice were sacrificed and total bone marrow cells from selected mice were retransplanted into secondary recipients to be monitored for up to an additional 12 months. Other *Il2rg*^{-/-} mice were transplanted with wild type or *Il2rg*^{-/-} Lin- cells transduced with GFP vectors driven by the SF, PGK or γ cPr promoters, or a vector containing the polylinker sequence but no

transgene (LV-SF-poly), in order to monitor the rate of insertional mutagenesis in non-*IL-2RG* vectors. Untransduced wild type Lin⁻ cells were transplanted into *Il2rg*^{-/-} mice as efficacy controls on a per-experiment basis. 1x10⁶ total bone marrow cells from selected mice (primarily mice treated with therapeutic vectors, i.e. vectors containing the *IL2RG* transgene) were transplanted into secondary recipients following 6 Gy TBI conditioning.

A total of 239 *Il2rg*^{-/-} primary transplants were generated over 11 separate gene therapy experiments, including 212 mice transplanted with lineage negative BM cells after overnight transduction with lentiviral vectors and 27 mice transplanted with untransduced BALB/c cells.

Immunophenotyping by flow cytometry

Flow cytometric analyses were performed on cells obtained from blood, bone marrow, thymus and spleen. Peripheral blood was collected monthly in EDTA tubes by retro-orbital puncture under isoflurane anesthesia. Complete blood cell counts were measured using a Vet ABC hematology analyzer (Scil animal care company GmbH, Germany). Blood was lysed and leukocytes were washed three times with Hank's balanced salt solution (HBSS, Invitrogen) containing 0.5% (wt/vol) bovine serum albumin and 0.05% (wt/vol) sodium azide (HBN). Cells were incubated for 30 minutes at 4°C in HBN containing 2% heat-inactivated normal mouse serum and antibodies against CD3, CD4, CD8, B220, CD19, IgM, IgD, CD11b Sca-1, and C-kit directly conjugated to R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP) or allophycocyanin (APC; all antibodies BD Biosciences). Subsequently, cells were washed and measured on a FACSCalibur or FACSCanto (Becton Dickinson). BM and spleen cells were evaluated similarly. Additionally, GFP expression was measured in mice treated with the SF-GFP, PGK-GFP or γ Pr-GFP vectors.

Identification and confirmation of leukemic mice

Mice which fell ill during the course of the experiment or who were found to have hematopoietic aberrations during the course of monthly blood analysis were sacrificed and tissue samples of spleen, lungs, heart, liver, kidney, and intestines were collected and fixed with 4% paraformaldehyde overnight. Potential leukemic phenotypes were determined via FACS analysis of peripheral blood, thymic, BM and spleen cells. DNA was extracted from 10x10⁶ spleen and BM cells for PCR analysis, with the remaining cells being frozen viably at -190°C for later transplantation into secondary recipients. If a leukemia was suspected, 1x10⁶ spleen or BM cells were thawed and transplanted into 3 secondary recipients to confirm the leukemic phenotype. These secondary transplants were monitored daily and sacrificed at the first sign of illness, with similar analysis performed as for the primary mice. A total of 17 leukemias were confirmed in primary or secondary recipients during the course of experiments.

Real-time quantitative polymerase chain reactions

qPCR to quantify the integrated proviral copy number[48] was performed on the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) with 100 ng genomic DNA template and SYBR Green PCR Master Mix (Applied Biosystems) with HIV primer, 5'-CTGGAAGGGCTAATTCCTC-3', and 5'-GGTTTCCCTTTC-GCTTTCAG-3' resulting in a 274 bp fragment of the proviral DNA. The PCR conditions were 95 °C for 10 min and 45 cycles as follows: 95°C for 15 s and 62°C for 1 min. A standard calibration curve was determined from flow sorted mouse 3T3 cells transduced with LV-SF-GFP at a multiplicity of infection (MOI) of 0.06 containing one integration per genome. All PCR reactions were performed in triplicate and analyzed with SDS2.2.2 software. qPCR reactions were performed on spleen and BM DNA from *Il2rg^{-/-}* mice transplanted with male donor cells to amplify a region of the Y chromosome as described by Pujal et al[49], using 100ng of genomic DNA and primers 5'-TCATCGGAGGGCTA-AAGTGTCAC-3' and 5'-TGGCATGTGGGTTCTCCTGTCC-3'. A standard curve was generated using spleen and BM DNA from male BALB/c mice.

LM-PCR and LAM-PCR

High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR) and ligation-mediated PCR (LM-PCR) was performed on BM or spleen DNA from *Il2rg^{-/-}* mice transduced with LV vectors. For LAM-PCR restriction enzymes Tsp509I and MseI were used with the lentiviral (HIV) primer set. PCR products were run on high resolution polyacrylamide (Spreadex) gel. To the nested primers of the LM-PCR[51, 52] protocol, the following primers were added to create the sequencing tags: 3'-GCCTCCCTCGGCCATCAG-5'+ MID (multiplex identifier) to the primer annealing to the viral LTR, and 3'-GCCTTGCCAGCCCGCTCAG-5'+ MID to the primer annealing to the linker. High throughput sequencing of LAM-PCR products was performed at GATC Biotech (Konstanz, Germany). Following sequencing, the barcode and LTR sequences were trimmed and the remaining genomic sequences uploaded to the MAVRIC analysis tool[53] for alignment to the mouse genome and annotation of nearby genes. Genes flanking each integration site (within 100kb) were checked against the Mouse Retrovirus Tagged Cancer Gene Database (RTCGD) and the Sanger Cancer Gene Census to identify oncogenes. Integration site prevalence was determined by calculating dividing the number of replicate sequences in a given sample by the total sequences for that sample: integrations making up more than 50% of the total sequence pool, or 5-fold higher fraction than the next most prevalent integration, were considered dominant. Dominant IS were re-checked for oncogenes within 500kb.

REFERENCES

1. Cavazzana-Calvo, M., et al., *Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease*. Science, 2000. **288**(5466): p. 669-72.
2. Gaspar, H.B., et al., *Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector*. Lancet, 2004. **364**(9452): p. 2181-7.
3. Aiuti, A., et al., *Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning*. Science, 2002. **296**(5577): p. 2410-3.
4. Hacein-Bey-Abina, S., et al., *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1*. Science, 2003. **302**(5644): p. 415-9.
5. Hacein-Bey-Abina, S., et al., *Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1*. J Clin Invest, 2008. **118**(9): p. 3132-42.
6. Howe, S.J., et al., *Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients*. J Clin Invest, 2008. **118**(9): p. 3143-50.
7. Bushman, F.D., *Retroviral integration and human gene therapy*. J Clin Invest, 2007. **117**(8): p. 2083-6.
8. Deichmann, A., et al., *Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy*. J Clin Invest, 2007. **117**(8): p. 2225-32.
9. Schwarzwaelder, K., et al., *Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo*. J Clin Invest, 2007. **117**(8): p. 2241-9.
10. Fischer, A. and M. Cavazzana-Calvo, *Gene therapy of inherited diseases*. Lancet, 2008. **371**(9629): p. 2044-7.
11. Ott, M.G., et al., *Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EV71, PRDM16 or SETBP1*. Nat Med, 2006. **12**(4): p. 401-9.
12. Boztug, K., et al., *Stem-cell gene therapy for the Wiskott-Aldrich syndrome*. N Engl J Med, 2010. **363**(20): p. 1918-27.
13. Modlich, U., et al., *Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity*. Blood, 2006. **108**(8): p. 2545-53.
14. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration*. Nat Biotechnol, 2006. **24**(6): p. 687-96.
15. Pfeifer, A., et al., *Transgenesis by lentiviral vectors: lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos*. Proc Natl Acad Sci U S A, 2002. **99**(4): p. 2140-5.
16. Schambach, A., et al., *Equal potency of gammaretroviral and lentiviral SIN vectors for expression of O6-methylguanine-DNA methyltransferase in hematopoietic cells*. Mol Ther, 2006. **13**(2): p. 391-400.
17. Schroder, A.R., et al., *HIV-1 integration in the human genome favors active genes and local hotspots*. Cell, 2002. **110**(4): p. 521-9.
18. Thornhill, S.I., et al., *Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency*. Mol Ther, 2008. **16**(3): p. 590-8.
19. Huston, M.W., et al., *Correction of murine SCID-X1 by lentiviral gene therapy using a codon-optimized IL2RG gene and minimal pretransplant conditioning*. Mol Ther, 2011. **19**(10): p. 1867-77.
20. Zhang, F., et al., *A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors*. Mol Ther, 2010. **18**(9): p. 1640-9.
21. Zychlinski, D., et al., *Physiological promoters reduce the genotoxic risk of integrating gene vectors*. Mol Ther, 2008. **16**(4): p. 718-25.
22. Cartier, N., et al., *Lentiviral hematopoietic cell gene therapy for X-linked adrenoleukodystrophy*. Methods Enzymol, 2012. **507**: p. 187-98.
23. Cavazzana-Calvo, M., et al., *Transfusion independence and HMG2 activation after gene therapy of human beta-thalassaemia*. Nature, 2010. **467**(7313): p. 318-22.
24. Galy, A. and A.J. Thrasher, *Gene therapy for the Wiskott-Aldrich syndrome*. Curr Opin Allergy Clin Immunol, 2011. **11**(6): p. 545-50.

25. Aiuti, A., et al., *Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome*. Science, 2013.
26. Biffi, A., et al., *Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy*. Science, 2013.
27. Ding, Y., et al., *Interferon-inducible protein IFIXalpha1 functions as a negative regulator of HDM2*. Mol Cell Biol, 2006. **26**(5): p. 1979-96.
28. Ding, X.F., et al., *Characterization and expression of a human KCTD1 gene containing the BTB domain, which mediates transcriptional repression and homomeric interactions*. DNA Cell Biol, 2008. **27**(5): p. 257-65.
29. Sood, R., et al., *Cloning and characterization of a novel gene, SHPRH, encoding a conserved putative protein with SNF2/helicase and PHD-finger domains from the 6q24 region*. Genomics, 2003. **82**(2): p. 153-61.
30. Filion, G.J., et al., *A family of human zinc finger proteins that bind methylated DNA and repress transcription*. Mol Cell Biol, 2006. **26**(1): p. 169-81.
31. Wilson, M. and P. Koopman, *Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators*. Curr Opin Genet Dev, 2002. **12**(4): p. 441-6.
32. Woods, N.B., et al., *Gene therapy: therapeutic gene causing lymphoma*. Nature, 2006. **440**(7088): p. 1123.
33. Pike-Overzet, K., et al., *Gene therapy: is IL2RG oncogenic in T-cell development?* Nature, 2006. **443**(7109): p. E5; discussion E6-7.
34. Pike-Overzet, K., et al., *Ectopic retroviral expression of LMO2, but not IL2Rgamma, blocks human T-cell development from CD34+ cells: implications for leukemogenesis in gene therapy*. Leukemia, 2007. **21**(4): p. 754-63.
35. Thrasher, A.J., et al., *Gene therapy: X-SCID transgene leukaemogenicity*. Nature, 2006. **443**(7109): p. E5-6; discussion E6-7.
36. Scobie, L., et al., *A novel model of SCID-X1 reconstitution reveals predisposition to retrovirus-induced lymphoma but no evidence of gammaC gene oncogenicity*. Mol Ther, 2009. **17**(6): p. 1031-8.
37. Dave, U.P., et al., *Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy*. PLoS Genet, 2009. **5**(5): p. e1000491.
38. Montini, E., et al., *The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy*. J Clin Invest, 2009. **119**(4): p. 964-75.
39. Paruzynski, A., et al., *Genome-wide high-throughput integrative analyses by nrLAM-PCR and next-generation sequencing*. Nat Protoc, 2010. **5**(8): p. 1379-95.
40. Harkey, M.A., et al., *Multiair high-throughput integration site detection: limitations of LAM-PCR technology and optimization for clonal analysis*. Stem Cells Dev, 2007. **16**(3): p. 381-92.
41. Avedillo Diez, I., et al., *Development of novel efficient SIN vectors with improved safety features for Wiskott-Aldrich syndrome stem cell based gene therapy*. Mol Pharm, 2011. **8**(5): p. 1525-37.
42. Stein, S., et al., *Genomic instability and myelodysplasia with monosomy 7 consequent to EVII activation after gene therapy for chronic granulomatous disease*. Nat Med, 2010. **16**(2): p. 198-204.
43. Gimeno, R., et al., *Monitoring the effect of gene silencing by RNA interference in human CD34+ cells injected into newborn RAG2-/- gammac-/- mice: functional inactivation of p53 in developing T cells*. Blood, 2004. **104**(13): p. 3886-93.
44. Ailles, L.E. and L. Naldini, *HIV-1-derived lentiviral vectors*. Curr Top Microbiol Immunol, 2002. **261**: p. 31-52.
45. Follenzi, A. and L. Naldini, *Generation of HIV-1 derived lentiviral vectors*. Methods Enzymol, 2002. **346**: p. 454-65.
46. van Til, N.P., et al., *Lentiviral gene therapy of murine hematopoietic stem cells ameliorates the Pompe disease phenotype*. Blood, 2010. **115**(26): p. 5329-37.
47. Follenzi, A. and L. Naldini, *HIV-based vectors. Preparation and use*. Methods Mol Med, 2002. **69**: p. 259-74.

48. van Til, N.P., et al., *Kupffer cells and not liver sinusoidal endothelial cells prevent lentiviral transduction of hepatocytes*. Mol Ther, 2005. **11**(1): p. 26-34.
49. Pujal, J.M. and D. Gallardo, *PCR-based methodology for molecular microchimerism detection and quantification*. Exp Biol Med (Maywood), 2008. **233**(9): p. 1161-70.
50. Schmidt, M., et al., *High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR)*. Nat Methods, 2007. **4**(12): p. 1051-7.
51. Kustikova, O.S., U. Modlich, and B. Fehse, *Retroviral insertion site analysis in dominant haematopoietic clones*. Methods Mol Biol, 2009. **506**: p. 373-90.
52. Pfeifer, G.P., et al., *Genomic sequencing and methylation analysis by ligation mediated PCR*. Science, 1989. **246**(4931): p. 810-3.
53. Huston, M.W., et al., *Comprehensive Investigation of Parameter Choice in Viral Integration Site Analysis and Its Effects on the Gene Annotations Produced*. Hum Gene Ther, 2012.

SUPPLEMENTAL TABLES

Supplemental Table 1: details of integration sites found in LV vector leukemias

ID #	vector	integrations per BM cell	% clonal dominance	chromosome	IS locus	flanking gene	distance from IS to gene 5' end
1	PGK.coIL2RG	1.1	96	11	120421360	Arl16	92.6 kb US
						Gm17586	73.7 kb US
						Npb	48.4 kb US
						Anapc11	38.4 kb US
						Gm11789	28.4 kb US
						Fam195b	10.5 kb US
						Ppp1r27	9.0 kb US
						P4hb	13.1 kb DS
						Arhgdia	21.4 kb DS
						Gcgr	29.3 kb DS
						Alyref	38.2 kb DS
						Pcyt2	57.8 kb DS
						Sirt7	65.1 kb DS
						Slc25a10	68.2 kb DS
						Mafg	73.4 kb DS
						Mrpl12	75.4 kb DS
Pycr1	83.6 kb DS						
Myadml2	88.0 kb DS						
Hgs	92.4 kb DS						
2	PGK.GFP	2.1	44	5	20378460	Tmem60	9.6 kb US
						Phtf2	9.4 kb DS
			55	10	10904550	Fbxo30	96.9 kb US
						Shprh#	35.0 kb DS

Abbreviations: US, upstream of gene transcription start site; DS, downstream of gene transcription start site;

* proto-oncogene

tumor suppressor gene or transcriptional repressor gene

Supplemental Table 1: details of integration sites found in LV vector leukemias (cont.)

ID #	vector	integrations per BM cell	% clonal dominance	chromosome	IS locus	flanking gene	distance from IS to gene 5' end
3	PGK.GFP	3.8	53	11	69550965	Mpdu1	74.9 kb US
						Cd68	71.4 kb US
						Eif4a1	65.1 kb US
						Senp3	55.5 kb US
						Tnfsf13	51.8 kb US
						BC096441	41.4 kb US
						Zbtb4	28.4 kb US
						Polr2a	21.1 kb DS
						Amac1	24.4 kb DS
						2010012P19Rik	55.3 kb DS
						Chrb1	58.4 kb DS
						Fgf11	64.3 kb DS
						Tmem102	68.1 kb DS
						Mir1934	74.4 kb DS
						4933402P03Rik	80.9 kb DS
						Sox15	82.1 kb DS
Spem1	84.6 kb DS						
4	PGK.GFP	2.6	5	4	139002617	Mrto4	94.1 kb US
						Iffo2	83.9 kb US
						Ubr4	94.0 kb DS
						Adamts3	94.5 kb DS
						Tmem60	9.5 kb US
						Phtf2	9.4 kb DS
5	PGK.GFP	3.1	29	5	90217673	Fbxo30	96.6 kb US
						Shprh#	35.3 kb DS
						Psd2	44.9 kb DS
						Tmem60	9.5 kb US
						Phtf2	9.4 kb DS
						Rsb11	79.1 kb DS
						Fbxo30	96.6 kb US
						Shprh#	35.3 kb DS
						Irgm1	74 kb US
						Gm5431	43.5 kb US
5	PGK.GFP	3.1	24	11	48759187	Olfir56	33.5 kb US
						Psme2b-ps	0.5 kb DS
						9930111J21Rik1	33.7 kb DS
						Gm12185	46.5 kb DS
						Tgtp1	48.5 kb DS

* proto-oncogene

tumor suppressor gene or transcriptional repressor gene

Supplemental Table 1: details of integration sites found in LV vector leukemias (cont.)

ID #	vector	integrations per BM cell	% clonal dominance	chromosome	IS locus	flanking gene	distance from IS to gene 5' end
6	SF.IL2RG	2.2	7	3	75428013	Pcd10	67.4 kb US
						Serpini1	66.5 kb DS
						Mef2c*	1.1 kb DS
						Utrn	291.6 kb DS
7	SF.IL2RG	2.8	52	5	24812154	Galnt15	124.8 kb DS
						Galnt11	83.4 kb DS
						Mll3*	192.3 kb DS
						Mettl15	158.9 kb DS
8	SF.poly	2.9	31	1	175565248	BC094916	99.2 kb US
						Pydc3	38.8 kb US
						Pydc4	35.9 kb US
						Pyhin1#	4.3 kb DS
						Abca13	221.4 kb DS
						Pmm2	4.7 kb US
						Tmem186	4.6 kb DS
Carhsp1	39.1 kb DS						
9	SF.poly	0	n/a	n/a	n/a	n/a	n/a
10	γcPR.coIL-2RG	1.4	10	6	127152223	9630033F20Rik	92.7 kb US
						Cend2*	51.2 kb US
						9330179D12Rik	52.8 kb DS
						Papola	10.3 kb DS
						Ak7	84.6 kb DS
11	γcPR.coIL2RG	1.6	68	2	32444602	Klhl34	51.9 kb DS
						Mir1954	63.2 kb US
						Eng	57.5 kb US
						Ak1	32.7 kb US
						St6galnac6	10.6 kb US
						St6galnac4	2 kb DS
						Pip5kl1	14.3 kb DS
						Dpm2	18.2 kb DS
						Fam102a	53.8 kb DS
						Fpgs	114.7 kb DS
22	11	43946382	Il12b	267.2 kb DS			
12	γcPR.coIL2RG	1.8	98	18	15379080	Kctd1	69 kb US
13	γcPR.coIL2RG	1.7	58	4	44798766	Pax5*	74.5 kb US
						Gm12463	32.8 kb US
						Zcchc7	30 kb DS
14	γcPR.GFP	0.8	32	10	9576640	Stxbp5	44.2 kb DS
						6884523	Rbfox1

* proto-oncogene

tumor suppressor gene or transcriptional repressor gene



Met TyrXle
Ala Cys
Met Ala Cys
Cys Xle Ala
Ala Lys Ser Cys
Xle Arg Pyl Lys
PylAsn Glu Asn Sar
Ser Leu SerThr LeuLeulle LeuLeu His
Cys Glu LeuGlu Asn TrpIle TrpIle Sec
Ala Ser GGAC GCC GGCATGCC
Xle Thr AGCAGGTACGCGAGA His Ala Asn Ser
GluArg AGGCTAGCTAG PylThr
CAGGCTCGA PylAsn Arg His
CAGTGGGTAT Thr Arg AlaArg ArgIle
GTGAGCTTGCT Ser Ile Asn Glu
GATAGCGGGTA Sec Thr Pyl Thr His
TGCTAGCCGGC His SerSec
GGGAATGCAGG
TACGCTTGCTAGA
CCGCAGGCATGC
CTGTATGCTTGCTC
01001101011000010111
00100 1110011011010000
01100 001011011000110
1100 0010000001010111
011 010010110110001101
10 0011010010110000101
101101001 00000010010
00011101 01011100110111010
0011011 11011011100010110000
100000 01010
00001 1010
000 100100

Comprehensive investigation of the parameters of viral integration site analysis and their effects on the gene annotations produced

Marshall Huston^{1*}, Martijn Brugman^{1,3*}, Sebastiaan Horsman², Andrew Stubbs², Peter van der Spek² and Gerard Wagemaker^{1**}

¹Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands

²Department of Bioinformatics, Erasmus University Medical Center, Rotterdam, the Netherlands.

*These authors contributed equally to this study.

ABSTRACT

Introducing therapeutic genes into hematopoietic stem cells using retroviral vector mediated gene transfer is an effective treatment for monogenic diseases. The risks of therapeutic gene integration include aberrant expression of a neighbouring gene, resulting in oncogenesis at low frequencies (10^{-7} - 10^{-6} /transduced cell). Mechanisms governing insertional mutagenesis are the subject of intensive ongoing studies which produce large amounts of sequencing data representing genomic regions flanking viral integration sites (IS). Validating and analyzing these data require automated bioinformatics applications. The exact methods used vary between applications, based on the requirements and preferences of the designer. The parameters used to analyze sequence data are capable of shaping the resulting integration site annotations, but a comprehensive examination of these effects is lacking. Here we present a web-based tool for integration site analysis, called Methods for Analyzing ViRal Integration Collections (MAVRIC), and use its highly customizable interface to look at how IS annotations can vary based on the analysis parameters. We used the integration data of the previously published ADA-SCID gene therapy trials for evaluation of MAVRIC. The output illustrates how MAVRIC allows for direct multi-parameter comparison of integration patterns. Careful analysis of the SCID data and re-analyses using different parameters for trimming, alignment and repeat masking revealed the degree of variation that can be expected to arise due to changes in these parameters. We observed mainly small differences in annotation, with the largest effects caused by masking repeat sequences and by changing the size of the window around the IS.

INTRODUCTION

The availability of PCR based techniques (inverseX, LAM-PCR, LM-PCR, splinkerette PCR) to identify viral integration sites has led to efforts to classify the relationship between virus integration and leukemia when using replicating retroviruses (Joosten *et al.*, 2002; Mikkers *et al.*, 2002; Erkeland *et al.*, 2006) as well as in gene therapy to identify clonal dominance and insertional mutagenesis leading to oncogenic events (Kustikova *et al.*, 2005; Modlich *et al.*, 2005). Further investigations into the integration profile of retroviruses and lentiviruses (Mitchell *et al.*, 2004; Montini *et al.*, 2006; Biasco *et al.*, 2011) has demonstrated the relationship between the occurrence of insertional oncogenesis and the integration profile (Montini *et al.*, 2006; Zychlinski *et al.*, 2008). Analysis of the amplicons retrieved after sequencing of viral genome boundaries (Kustikova *et al.*, 2009) is usually automated due to the large number of sequences produced by modern high throughput sequencing. The specific analysis methods applied to raw sequencing data vary between applications, which raises the question of how adjusting the analysis parameters might affect the annotations produced. We developed a bioinformatics pipeline aimed at high throughput data analysis for virus genome boundaries, which could be easily modified to compare the results of analyzing the same data set while varying a wide range of parameters. Similar tools have been described (Appelt *et al.*, 2009; Hawkins *et al.*, 2011) for alignment of genome virus boundary sequences, but differ in customization options compared to our program. Understandably, analysis parameters in these tools are locked to ensure consistent results. For our bioinformatics tool we chose to allow several parameters to be adjusted prior to analysis to determine the impact on the final annotation results.

MAVRIC (Methods for Analyzing ViRal Integration Collections, <http://mavric.erasmusmc.nl>) automatically BLASTs input sequences according to user-specified parameters and returns information on the genes surrounding the viral integration site (IS). The analysis parameters used by MAVRIC can be set by the user, including repeat masking, minimum accepted sequence length, maximum E-value threshold, the maximum distance from the IS to search for genes, and the ability to use previous versions of BLAST and older genome builds in Ensembl. Analysis of nearby gene expression levels using user-generated expression data is an additional option, with MAVRIC automatically generating binning charts. Users can upload viral or adapter sequence information such as LTRs to be screened prior to analysis. This paper outlines the design choices made in the process of developing tools for virus integration analysis and examines the differences obtained when analyzing the same dataset with different database builds and analysis methods.

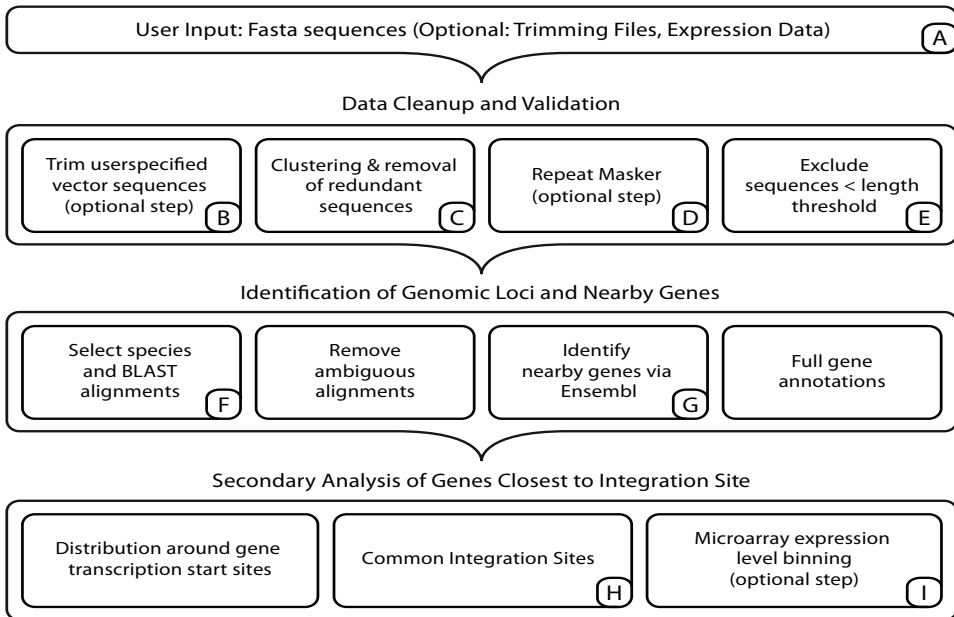


Figure 1. MAVRIC workflow.

- A (user input) - Fasta formatted sequences of human, murine or macaque origin. The user may also input vector sequences to be trimmed and/or expression data for additional analysis
- B (vector trimming step) - MAVRIC uses cross-match to identify any user-specified vector sequence fragments and removes them, leaving the genomic sequence
- C (clustering step) - Sequences which are >95% identical to another sequence are automatically removed from analysis
- D (RepeatMasker) - Different masking criteria are available for various species (human, murine, macaque or mammalian)
- E (sequence length threshold) - Sequences shorter than the cut-off are excluded from analysis. Recommended range is 11 bp to 25bp
- F (BLAST step) - Some older versions of BLAST are available for easy comparison to historical analysis
- G (identify nearest gene to integration site) - Identifies the “nearest” gene (defined as the gene with the 5’ end closest to the IS) and all genes within the user-specified distance from the IS on either strand
- H (identify common integration sites) - Based on the list of genes nearest to IS. Any gene linked to 3 or more sequences is considered a CIS gene
- I (expression level binning) - User may upload expression data to be used in generating IS nearest gene expression binning charts

MATERIALS AND METHODS

Integration analysis

The annotation workflow can be summarized as series of steps which consist of: data clean up, integration site detection, and identification of the nearest gene(s) to that site (Figure 1). First, the virus-genome boundary sequences are uploaded to the MAVRIC server. Sequence requirements are as follows: fasta format, 5'-3' orientation, and at least 11 bp in length. MAVRIC currently supports human, murine and macaque sequence data. The user may choose to include RepeatMasker criteria, upload vector sequence data to be trimmed, or change the minimum acceptable sequence length. Sequences that do not meet the minimum length requirement after any trimming or repeat masking step are excluded from further analysis. If the user is loading pre-trimmed sequences that contain no vector DNA, the genomic sequence fragments should be loaded such that the viral LTR would be adjacent to the 5' end of the sequence.

If the sequences contain fragments of vector, adapters or linkers that must be removed prior to annotation, the user should upload them as separate files. Up to three such trimming files can be included. The upload windows are helpfully named for vector sequence, LTR and polylinker, but any external residues that may contaminate the genomic sequence can be uploaded and screened. MAVRIC detects and removes designated sequences via cross-match, an implementation of the Smith-Waterman-Gotoh algorithm developed by Phil Green (<http://www.phrap.org/>, University of Washington). Vector and linker/adaptor sequences will be trimmed regardless of their orientation. The user can also choose to exclude data that do not contain certain external sequences, such as viral LTRs, that are used to validate bona fide integration sites.

After trimming, MAVRIC clusters the approved sequences by an implementation of cd-hit-est (Li and Godzik, 2006). Sequences which are found to be more than 95% identical to another sequence in the dataset are considered redundant integrations and excluded from further analysis. The number of redundant sequences in each cluster is tracked and this information is presented in the “skipped sequences” summary table.

Submitted sequences can also be pre-screened using the online tool RepeatMasker (www.repeatmasker.org, Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0 1996-2004*). RepeatMasker screens sequences for regions of known genomic repetitive sequence using an implementation of the cross-match algorithm. Simple repeats, tandem repeats, segmental duplications and interspersed repeats such as pseudogenes, retrotranscripts, SINEs, DNA transposons, retrovirus retrotransposons and non-retroviral transposons (LINEs) are masked by this algorithm. RepeatMasker can be adjusted to known repetitive sequences for several species (currently human, mouse, macaque and mammalian are available on MAVRIC).

MAVRIC aligns the screened sequences in the genome via BLAST (Basic Local Alignment Search Tool (Altschul *et al.*, 1990), a fast algorithm that allows alignment of query

sequences (DNA or protein) to a predefined database. It finds the largest subsection of the query and tries to expand it in both directions, assigning penalties for the creation of gaps or mismatches. Given a query sequence, BLAST identifies the best fitting alignment in addition to similar alignments with higher penalty scores. In MAVRIC, the user can select the maximum E-value to be returned, with the default maximum E-value set to 0.01. The E-value is a statistical measure that shows how likely it is that the given alignment occurred by chance. Lowering the maximum E-value will result in fewer BLAST hits accepted but increases the confidence that those alignments are accurate. The alignments of integration sites identified by BLAST and accepted by MAVRIC are subsequently annotated using data available from Ensembl. Integration sites with multiple alignments possessing the same lowest E-value are considered non-informative since the alignment is ambiguous and these sites are excluded from further analysis.

Versions of BLAST linked to older NCBI builds are available for each species in order to precisely compare new data to previously-analyzed sets. NCBI v34-v37, v36-v37 and v1.41 are available for human, mouse and macaque, respectively. The user chooses the maximum acceptable E-value (Expect value) for the BLAST results. BLAST hits that exceed this maximum E-value are excluded from further analysis. The Ensembl database used is based on the BLAST version and species selected. Note that some BLAST versions are linked to more than one Ensembl database (for example, mouse v37 is linked to Ensembl mouse 49 and 56), in which case the user may select the Ensembl database they wish to use. Finally, the user sets the size of the window flanking the integration site to search for genes, and chooses whether to return only Refseq genes (Pruitt *et al.*, 2007) or all genes and genomic features (miRNAs, pseudogenes, snRNAs, etc) found by Ensembl.

Annotation of retrieved integration locations

Genes adjacent to the integration site are annotated as follows: gene symbol, Ensembl ID, Entrez ID, Unigene ID, Refseq ID, biotype, chromosome, strand, location of the hit start and hit end, and both the raw distance (defined as the minimum distance from the IS to the 5' or 3' end of the gene) and 5' distance (defined as the distance from the gene transcription start site to the IS). Two outputs are generated for Ensembl hits: one file that contains every gene with at least 1 bp within the user-defined maximum distance from the integration site (default set to 100 kb), and another file which only contains the gene whose 5' end (adjacent to the transcription start site) is closest to the integration site and is within the specified maximum distance. The 'all genes' file, assuming the user-defined distance is sufficiently large, gives an overview of every gene that could potentially be affected by the integrated vector. The 'nearest gene' file contains less information about the genomic loci surrounding the integrations, but allows each integration site to have an equal "weight" for certain downstream analyses. MAVRIC uses the 'nearest gene' file to generate a histogram showing the distribution of distances from the integration site to the nearest gene transcription start site. It also lists common integration sites (defined here as

3 or more hits for a given gene in the ‘nearest gene’ dataset) and a summary of the analysis parameters and excluded sequences. Graphical output and a summary of the IS annotations is presented in the results page, with the full output available for download in a zip file (Supplemental Figure 1)

Nearest gene expression level binning

MAVRIC includes an optional feature which associates genes annotated by MAVRIC as nearest to an integration site with user-generated expression level data, sorted in equally sized bins (Schróder, *et al.*, 2002). The user submits a table of gene expression values generated from Affymetix arrays or similar analyses. MAVRIC sorts the genes in this table into 10 equally-sized bins as published previously) (Deichmann *et al.*, 2007; Schwarzwaelder *et al.*, 2007) based on their relative expression levels, and then checks each gene in the ‘nearest gene’ file against this binning profile. MAVRIC creates a histogram displaying the number of IS genes allotted to each bin, giving an overview of the relation between gene expression level and integration frequency (Supplemental Figure 1D). Common integration site genes are also plotted separately on the same graph. Users can upload own expression data based on other experimental parameters along with their sequences. Note that MAVRIC cannot process raw array data: expression values must be determined by the user prior to upload. MAVRIC is able to process expression values from any source, provided that the user uploads the data in the proper format (tab-delimited text, see the help file at <http://mavric.erasmusmc.nl/help.php?page=runs> for further details). Users can upload up to two expression values per gene, which will result in two separate binning graphs. This allows easy comparison of two gene expression datasets, for example before and after cell stimulation.

Test set for comparison of parameters: a published sequence set for ADA-SCID gene therapy

We compared the MAVRIC annotation process with the published annotation of a dataset retrieved from a clinical gene therapy trial for ADA-SCID (Aiuti *et al.*, 2007). The published data contained 706 sequences along with the gene found nearest to each integration site. These sequences were analyzed via MAVRIC, using the NCBI 37 assembly and Ensembl version 56. A total of six variations of RepeatMasker settings and E-value conditions were performed (Table 1). The ‘nearest gene’ and ‘all genes’ annotations for each MAVRIC variant, as well as the originally published analysis, were used as input for Ingenuity functional analysis (Ingenuity Systems). Ingenuity identifies biological functions and/or diseases that are most significant to the set of genes provided by analyzing RNA expression data in the context of known biological responses and regulatory networks as well as other higher-order response pathways. The following Ingenuity settings were used for pathways analysis: Direct and Indirect Relationships; Data Sources = all; Confidence = experimentally observed; Species = all, Tissues and Cell Lines = all. This analysis was repeated using KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis, a manually curated list of genes and their interactions.

Table 1a. Analysis parameters for MAVRIC using ADA-SCID data as input

name of analysis criteria	max E-value	uses Human Repeat-Masker?	minimum length threshold	distance around IS searched	# of sequences processed	# of IS with BLAST hits	# of IS annotated	# of unique genes
MAV1 (permissive)	10	no	11	100k	701	697	636	574
MAV2	0.01	no	11	100k	701	671	613	558
MAV3	0.00001	no	11	100k	701	621	567	522
MAV4	10	YES	11	100k	633	625	572	517
MAV5	0.01	YES	11	100k	633	592	547	495
MAV6	0.00001	YES	11	100k	633	540	495	457
conservative	0.01	YES	16	100k	630	592	542	495
super-permissive	10	no	11	500k	701	697	687	620

Table outlines analysis parameters and the number of annotations returned for each of eight MAVRIC settings. The number of sequences annotated can be greatly increased by skipping RepeatMasker and extending the size of the window around the IS to 500 kb, however this may result in lower quality annotations.

Table 1b. Gene ontology annotation results based on MAVRIC analysis parameters

"nearest gene" dataset		INGENUITY functional analysis	
Dataset	number of unique genes annotated	number of mapped genes	% mapped
MAV1	574	547	95.3%
MAV2	558	531	95.2%
MAV3	522	496	95.0%
MAV4	517	493	95.4%
MAV5	495	471	95.2%
MAV6	457	434	95.0%
ADA-SCID analysis	706	639	90.5%
"all genes within 100 kb" dataset		INGENUITY functional analysis	
Dataset	number of unique genes annotated	number of mapped genes	% mapped
MAV1	2059	1901	92.3%
MAV2	2014	1827	90.7%
MAV3	1929	1780	92.3%
MAV4	1854	1716	92.6%
MAV5	1796	1656	92.2%
MAV6	1709	1576	92.2%
ADA-SCID analysis	1106	983	88.9%

Results of loading MAVRIC nearest gene annotations into Ingenuity functional analysis. The analysis parameters have little effect on the percentage of genes mapped by Ingenuity, but they do have an effect on the top pathways returned (see Supplemental Table 2).

For direct comparison to the published analysis, we used the MAVRIC output with the following settings, hereby referred to as the ‘conservative criteria’: human RepeatMasker, minimum sequence length = 16 bp, maximum E-value = 0.01, maximum distance from gene to IS = 100 kb. These criteria returned 542 annotated sequences, or 77% of the dataset. We then re-analyzed the dataset in MAVRIC with no RepeatMasker step, reducing the minimum sequence length to 11 bp and increasing the maximum E-value to 10. These parameters, dubbed the ‘permissive criteria,’ annotated a total of 636 sequences, for a 90% return rate. The dataset was re-analyzed several more times with various parameters, described in detail in the next section.

To determine the relationship between retroviral integration and the expression of nearby genes in hematopoietic stem cells, the target cells for gene therapy trials in inherited disorders of the blood cell / immune systems (Aiuti *et al.*, 2002; Hacein-Bey-Abina *et al.*, 2002; Gaspar *et al.*, 2004), we referenced the ‘nearest gene’ list generated using the ‘permissive criteria’ to gene expression level data in hematopoietic stem cells. Gene expression values obtained from CD34⁺ human cord blood cells as measured on a U133 plus 2.0 gene array were sorted into 10 bins based on relative expression. Genes identified by MAVRIC as the unique nearest gene to each IS were labelled according to their bin, and the resulting total for each bin was placed into a histogram.

ADA-SCID integrations compared to XSCID integration data.

To assess to what extent the ADA-SCID data analyzed in this study overlap with data from the XSCID clinical trial, previously published data from the London and Paris XSCID trials (Deichmann *et al.*, 2007; Schwarzwaelder *et al.*, 2007) were compared using MAVRIC analysis with the permissive criteria.

RESULTS

MAVRIC performance testing

The current application annotates over 2500 unique integrations per hour, the rate limiting step being the BLAST alignment (60% of the total processing time). MAVRIC was not designed to process the extremely high number of sequencing reads generated by high throughput sequencing applications, but it does contain several features which make annotation of medium-sized datasets possible. High-throughput sequencing methods produce a high volume of redundant sequence reads. The use of cross-match to recognize and remove redundant sequences (classified here as >95% identical to another sequence in the dataset) allows MAVRIC to process large high-throughput sequencing datasets in a relatively short amount of time. In our testing environment, MAVRIC was able to trim, cluster and annotate up to 30,000 high throughput LAM-PCR sequences per hour.

The user can choose to exclude sequences that do not contain certain external fragments, such as viral LTRs. High throughput sequencing methods produce a number of aspecific genomic fragments, meaning that the presence of a viral LTR flanking genomic sequence is necessary to confirm a bona fide insertion site.

Using repeat masking to improve annotation throughput

Genomes contain sequences of interspersed repeats and regions of low complexity: repetitive sequences make up almost 50% of the human genome (Levy *et al.*, 2007). These repeat regions cause difficulties when intersecting with a virus genome boundary sequence. When multiple alignments retrieved by BLAST share the lowest E-value score, the sequence will not point to a unique place in the genome. Repeat sequences also cause the throughput of the BLAST alignment to decrease. We approached this problem by pre-screening the virus genome boundary sequences using the online tool RepeatMasker. RepeatMasker allows batch uploads of sequences and returns sequences with the desired repeats masked. The resulting trimmed sequences can then be aligned using BLAST without the throughput penalties or alignment ambiguities that repeat sequences introduce. The number of different repeat consensus sequences varies between genomes and the user should indicate from which species the sequences are derived when selecting RepeatMasker. Figure 2A-B shows an example of the impact of repeat masking on sequence lengths. Repeat masking clearly shortens the sequences by removal of the repeated regions and reduces ambiguous alignments. Repeat masking also reduces the time required to BLAST alignments, leading to a net decrease in annotation time. RepeatMasker does cause some sequences to lose the majority of the data, and in a few cases the entirety of the sequence is deleted (Figure 2B). Users choosing to implement RepeatMasker should expect to have anywhere from 2% to 12% of their sequence data excluded prior to the BLAST step due to being trimmed below the acceptable length threshold, depending on the initial length and quality of the sequences.

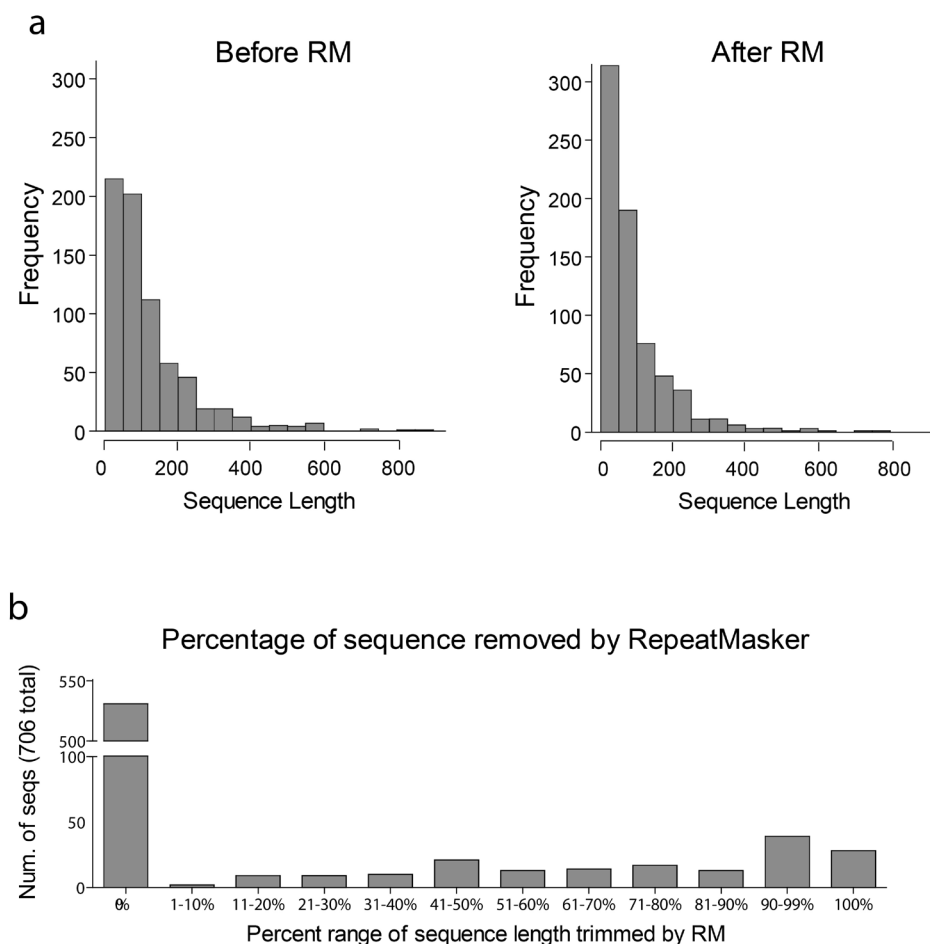


Figure 2. Changes in sequence length distribution caused by RepeatMasker. (a) The distribution of sequence lengths in the original data (left) and after repeat masking (right). (b) Percentage of each sequence removed by RepeatMasker. Dataset was the 706 sequences published by Aiuti et al., 2007.

Differences in annotation between genome databases

The annotation of virus integration sites is dependent on the analysis parameters provided. While the different databases (UCSC, Ensembl, NCBI) use the same underlying genomic information, (<http://genome.ucsc.edu/FAQ/FAQreleases>, although UCSC mouse genome assembly contains only the C57Bl/6J reference strain data), the annotations are slightly different. For instance, Ensembl (http://www.ensembl.org/info/docs/genbuild/genome_annotation.html) utilizes an automated annotation pipeline, using mRNA and protein data, combined with manually reviewed curated data from the Vega project (<http://vega.sanger>).

ac.uk/index.html) and reviewed protein coding transcripts from CCDS (<http://www.ensembl.org/info/docs/genebuild/ccds.html>) (Flicek *et al.*, 2011). Any integration dataset therefore depends on the annotation database (NCBI, Ensembl, UCSC) used. Although the majority of the genes identified will be identical regardless of the database, there might be some differences between the annotations, possibly leading to different genes being assigned as the nearest gene to an integration. For example, the murine Mecom locus, in which several splice variants for Evi1 and Mds1 are situated, shows differences in annotation between NCBI, the UCSC genome browser and Ensembl (Supplemental Figure 2 and Supplemental Table 1).

The use of NCBI Reference sequences or broader annotation

The option to annotate only Refseq genes has the advantage of returning only well described genes, but has the disadvantage of filtering out potentially interesting genomic features such as miRNAs, snRNAs and pseudogenes. In an analysis of a sample dataset (the published ADA-SCID data described below), filtering for Refseq genes resulted in only 3.5% reduction in the number of sequences annotated but a 19.5% mismatch in the genomic features identified via Ensembl as nearest to the IS (data not shown).

Annotation of retrieved integration locations

For each integration site, there are surrounding genes upstream and downstream on both strands. In addition to the closest gene, an integration may have several more neighbouring genes that could be influenced by elements of the inserted provirus. Additionally, integrated promoter/enhancer regions can influence an area surrounding the integration of at least 92 kb (Bartholomew and Ihle, 1991) and possibly as large as 500 kb (Kustikova *et al.*, 2005). Studies of deregulation of the virus integration locus should therefore be focused on such large areas and might include up to 1 Mb regions. With this in mind, our tool generates two outputs for Ensembl hits: one file that contains every gene located within a user-defined window around the integration site (default is 100 kb), and another file which only contains the gene whose 5' end is nearest to each site. Assigning one unique gene that maps 'nearest' to each integration site allows for easier automated downstream analysis and allows every integration site to have equal impact, but user investigation of all the genes near each IS is highly recommended.

Impact of genome builds on annotation output

In addition to RefSeq filtering and the size of the window around the IS, changes between versions of genome builds can have an effect on the annotations returned. Table 2a shows how the number of sequences hit and the number of genes identified increase with successive builds of the human genome. Table 2b illustrates how an identical dataset can produce different outputs depending on which version of the NCBI human genome build was used by the BLAST algorithm. In particular, the gene symbols are updated frequently. For

instance, comparing the outputs for the published ADA-SCID sequences using v35 and v37, we find a 99% match in sequences with alignments, but only 91% of the Ensembl IDs and 89% of the EntrezIDs for those sequences match, and only 72% of the common gene names. We therefore highly recommend including the Ensembl ID, which exhibited the most stability between genome builds, for any large scale analysis and in published data.

Table 2a. Comparison of results returned when using different builds of the human genome.

	Human NCBI 34 (October 2003)	Human NCBI 35 (October 2004)	Human NCBI36 (March 2008)	Human NCBI37 (July 2009)
num. of integration sites mapped (unique BLAST hit)	600	628	630	636
num. of integration sites within 100kb of a Refseq gene	543 (90.5%)	569 (90.6%)	571 (90.6%)	574 (90.2%)

The number of integrations mapped and genes identified by NCBI build, using published ADA-SCID data as input. The consensus sequence is updated often, which can cause integration sites to be mapped to different places or to change the nearest gene. Sequences were annotated via MAVRIC, using the following parameters: no repeatmasking, maximum e-value = 10, minimum sequence length = 11 bp.

Table 2b. Comparison of the overlap in integration sites (IS) and gene identifiers linked to each sequence (seq) over multiple human genome builds (versions 34 through 37).

	v.34 vs 35	v.34 vs 36	v.34 vs 37	v.35 vs 36	v.35 vs 37	v.36 vs 37
num. of matching sequences with annotated IS	596	590	595	618	624	627
num. of seqs with matching nearest gene Ensembl IDs	508	479	455	551	519	548
num. of seqs with matching nearest gene Entrez IDs	n/a	n/a	n/a	519	505	517
num. of seqs with matching nearest gene names	355	348	330	469	411	490

The number of annotated integration sites in different genome builds are nearly identical, but there are some discrepancies in the gene Ensembl IDs and Entrez IDs. This means that the IS occupies a slightly different place in the genome between the various builds. There are even greater discrepancies between gene names, indicating that the commonly used names for genes are subject to a great deal of turnover. Genome information and build dates were retrieved from ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/. Entrez ID information was not available (n/a) for v.34.

Comparison of MAVRIC analysis settings via gene ontology

A total of six MAVRIC analysis parameters (MAV1-MAV6, Table 1a) were used to generate ‘nearest gene’ and ‘all genes’ data as the input for Ingenuity functional analysis and KEGG pathway analysis. In general, the same pathways were found regardless of the analysis settings but the ranking of the top 5 pathways had some minor adjustments (Supplemental Table 2). There was slightly more variation seen in the top pathways for the ‘nearest gene’ datasets compared to the ‘all genes’ datasets. On average, Ingenuity was able to annotate approximately 95% of the genes in the ‘nearest gene’ datasets and 92% of the ‘all genes’ datasets (Table 1b).

Comparison of MAVRIC and a published analysis for ADA-SCID gene therapy

We compared MAVRIC’s performance with the manual annotation of a dataset retrieved from a clinical gene therapy trial for ADA-SCID, in which patients were transplanted with CD34⁺ cells transduced with the adenosine deaminase (ADA) gene using a retroviral vector (Aiuti *et al.*, 2007). We analyzed 706 sequences using MAVRIC, which returned a varying number of annotated results depending on the RepeatMasker and E-value conditions (Table 1a). Increasing the E-value cut-off allows lower quality alignments to be annotated, however it also increases the processing time as more BLAST hits are generated. Sequences that have two or more alignments with equal minimum E-values cannot be confidently assigned to a specific place in the genome and so are excluded by MAVRIC from further analysis. This often occurs when the IS is located in repeat-dense regions of the genome. Repeat masking shortens the input sequences by removing highly repeated

regions from the input sequences, and occasionally will shorten an input sequence below the minimum length threshold for BLAST. RepeatMasker can also sometimes have an unintended effect on the BLAST alignments. For instance, the ADA-SCID sequences S3_022 and S3_110 exceed the length threshold but no longer align to a unique location in the human genome after human repeat masking.

For comparison to the published ADA-SCID data, we used the MAVRIC output using the following settings, hereby referred to as the ‘conservative criteria’: human RepeatMasker, minimum sequence length = 16 bp, maximum E-value = 0.01, maximum distance from gene to IS = 100 kb. These criteria were selected in order to return the highest confidence alignments. Using these settings, MAVRIC returned 542 annotated IS, or 77% of the previously published sequences. Five sequences were removed during the clustering stage due to a more than 95% homology with another sequence in the dataset. An additional 71 sequences were excluded due to being under the minimum length requirement after repeat masking. Of the remaining excluded sequences, 38 did not have an unambiguous lowest E-value < 0.01 and 50 had no Refseq gene within 100 kb of the IS.

In order to annotate more of the ADA-SCID dataset, we adjusted some of MAVRIC’s parameters. The following changes were made: no repeat masking step, reduce the minimum sequence length to 11 bp and increase the maximum acceptable E-value from 0.01 to 10. These settings, dubbed the ‘permissive criteria’, resulted in 697 seqs returning unique lowest BLAST hits. 636 IS were within 100 kb of a Refseq gene, for a 90% annotation rate.

To annotate the maximum number of sequences from this dataset, we extended the window around the IS to look for genes to up to 500 kb distant in either direction. This resulted in 687 sequences returning gene annotations, for a final return rate of 97%. Of the 19 ADA-SCID sequences MAVRIC was unable to annotate, five were excluded during clustering for being >95% similar to another sequence in the dataset. Four more sequences were excluded due to having a non-unique lowest E-value (i.e., these sequences aligned with the same confidence to two or more places in the genome). The remaining 10 sequences had no Refseq genes within 500 kb of the IS (in the original ADA-SCID analysis, 11 sequences had this property).

The distribution of IS is highly similar between the previously published analysis (Aiuti *et al.*, 2007) and MAVRIC permissive criteria (Table 3a). For the combined sequences, MAVRIC analysis returned the same 44%/56% split between intragenic and intergenic integrations. The percentage of integration sites less than 30 kb upstream, 10 kb upstream and within 5 kb of the transcription start site (TSS) is higher in the MAVRIC analysis, likely due to a slightly higher gene density in the newer human genome builds. The distribution of IS around the nearest gene TSS are also similar (data not shown). Using MAVRIC’s conservative criteria had a very minor effect on the distribution of integrations, skewing slightly away from distribution seen in the original ADA-SCID analysis.

Table 3a. Retroviral integration site distribution in hematopoietic stem cells from ADA-SCID patients, comparison between analysis methods

analysis method	IS annotated	intragenic	intergenic	<30 kb upstream	<10 kb upstream	+/-5 kb from TSS
ADA-SCID in vitro	212	50.9%	49.1%	19.4%	12.3%	23.6%
MAVRIC in vitro	185	47.6%	52.4%	30.3%	19.5%	27.6%
ADA-SCID ex vivo	496	41.3%	58.7%	25.6%	19.6%	28.8%
MAVRIC ex vivo	451	42.8%	57.2%	29.5%	20.4%	35.9%
ADA-SCID total	706	44.3%	55.7%	23.7%	17.4%	27.3%
MAVRIC total	636	44.2%	55.8%	29.7%	20.1%	33.5%
MAVRIC total, conservative criteria	542	44.8%	55.2%	30.1%	21.0%	33.9%

Comparing the published ADA-SCID analysis with MAVRIC's analysis of the same dataset using the permissive criteria (see Table 1A). Analysis of the total dataset was also performed with the MAVRIC conservative criteria. Although the distributions vary slightly between analysis methods, the overall integration patterns are similar.

Table 3b. Comparison of unique genes and tumorigenesis-related genes found by each analysis method.

analysis method	sequences annotated	unique genes returned	oncogenes annotated
ADA-SCID analysis nearest gene only	706	637	30
ADA-SCID analysis all genes	706	972	36
MAVRIC permissive criteria nearest gene only	636	567	26
MAVRIC permissive criteria all genes within 100 kb	636	2049	49
MAVRIC conservative criteria nearest gene only	542	495	24
MAVRIC conservative criteria all genes within 100 kb	542	1790	46

Expanding the MAVRIC search parameters to list all RefSeq genes within 100 kb of the IS result in a roughly 4-fold increase in the number of genes returned. The number of annotated oncogenes also increases. The ADA-SCID analysis had no limit on distance from the IS to the nearest gene, so it was able to annotate all sequences.

The impact of choosing one gene per integration versus all genes within a defined window

As expected, searching for all genes within 100 kb of the integration site resulted in many more genes annotated compared to listing one 'nearest' gene per integration (Table 3b). The number of Refseq genes returned was approximately 4-fold higher relative to linking only a single gene to each IS. Checking these genes against a list of 464 genes listed in the Sanger Cancer Genome project (<http://www.sanger.ac.uk/genetics/CGP/Census/>) revealed that the 'all genes' datasets contained many more potential oncogenes than the 'nearest gene' datasets. Additionally, using the 'all genes' dataset in Ingenuity and KEGG pathway analysis results in a large shift in the top functional pathways returned

(Supplemental Table 2). Restricting the annotations to those genes closest to the insertion site therefore masks potentially interesting genomic features and risks.

Expression level binning reveals that CIS genes are clustered in highly expressed bins

MAVRIC used the “nearest gene” list from the permissive analysis parameters as input for gene expression level binning. The list of genes was referenced to hematopoietic stem cell gene expression data obtained from CD34⁺ umbilical cord blood cells, as measured on an U133 plus 2.0 gene array. The list was then sorted into ten equally sized bins based on expression level. The results are charted based on the number of genes per bin, demonstrating that in the ADA-SCID trial integration occurred more often near highly expressed genes (Figure 3). Genes identified as common integration site (CIS) genes are also charted separately, and are present exclusively in the top three bins. Similar results were found when using the conservative criteria (data not shown). This suggests that the vector used in the ADA-SCID trial has a preference for integrating near highly expressed genes in hematopoietic stem cells.

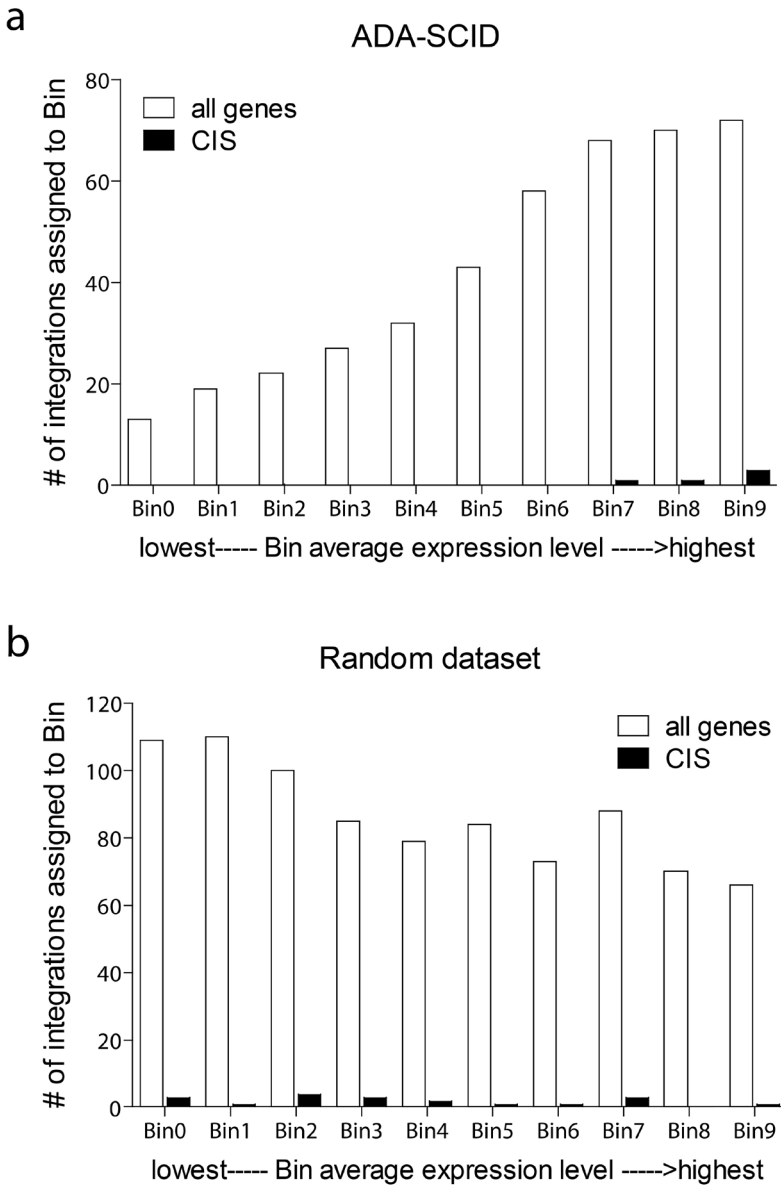


Figure 3. Expression level binning of genes identified in ADA-SCID trial by MAVRIC analysis or randomly selected genes. (a) Binning of 574 genes located proximal to IS from the ADA-SCID clinical trial: highly-expressed genes are more likely to have nearby IS. All of the CIS genes found in this dataset were assigned to the three highest bins. Genes annotated using MAV1 (permissive) criteria. (b) Binning of 864 genes nearby randomly-generated IS: a slight bias towards IS in low-expressed genes, and CIS are found in most bins. Expression levels generated from human HSC using an Affymetrix U133 array.

ADA-SCID integrations compared to XSCID integration data.

To assess to what extent the ADA-SCID data analyzed in this study overlap with data from the XSCID clinical trial, previously published data from the London and Paris XSCID trials (Deichmann *et al.*, 2007; Schwarzwaelder *et al.*, 2007) were compared with the MAVRIC tool using the permissive criteria. The ADA SCID and XSCID trials used similar gammaretroviral vectors, which is reflected in a remarkable overlap in the list of genes nearest to an IS (14-16% compared to 4-6% vs random datasets, Table 4).

Table 4. Genes co-occurring between X-SCID clinical data and ADA-SCID clinical data during MAVRIC analysis.

Trial	# of IS annotated	# of unique genes annotated as nearest to an IS	overlap with ADA-SCID nearest genes
ADA-SCID	636	574	--
X-SCID Paris	445	390	64 (16.4%)
X-SCID London	524	487	69 (14.2%)
X-SCID Combined	969	815	117 (14.4%)
Random 1	998	882	33 (5.8%)
Random 2	999	840	31 (5.4%)
Random 3	999	872	23 (4.0%)
Random 4	998	870	29 (5.1%)

The integrations obtained from clinical samples retrieved in the London and Paris X-SCID trials, compared to the ADA-SCID integration data used in this study. Analysis consists of unique genes designated as nearest to an IS using MAVRIC's permissive criteria (see Table 1). The ADA-SCID data is also compared against four sets of randomly-generated integration sites. The overlap between the clinical datasets is rough 3-fold higher than the overlap with the random data.

DISCUSSION

Automated analysis tools allow efficient annotation of virus integration sites according to a predefined set of criteria, reduce the risk of user input errors and allow for convenient reanalysis in case annotations or database builds change. We created the MAVRIC web tool to automatically validate integration site sequences, align those using BLAST and annotate them using data retrieved from the Ensembl databases. MAVRIC can remove external or redundant sequences, returns a list of genes near each integration site and annotates the genomic loci (Figure 1).

To efficiently process larger datasets, we have employed a clustering strategy, which accelerates BLAST alignments for high throughput sequencing data. Sequences which are found to be >95% similar to a previously-uploaded sequence in the dataset are removed. Sequences removed this way are tracked in a table which details why they were excluded and how many times that redundant sequence was found.

Obviously, a method to annotate integrations following a defined set of criteria is less likely to make mistakes such as typing errors. Yet there is a problem even when reporting the automatically annotated integrations. Usually, spreadsheets are a convenient way of displaying moderately sized datasets. The automatic correction methods built in to spreadsheets such as Microsoft Excel, however, can cause difficulties when the built-in correction methods are inadvertently used on gene symbols. This problem has been reported previously (Zeeberg *et al.*, 2004) and has also occurred in our analysis

Another factor to consider when analyzing integration site data is the distance from the IS to nearby genes. Some IS may have no genes within 1 Mb or more, while others have multiple genes less than 50 kb away. We recommend searching for genes within 100 kb on either side of the IS, as viral integrations have been shown to have an effect on nearby gene regulation within this distance, but it is possible to widen this window at the cost of increased processing time. Widening the range to 500 kb allowed 7% more sequences to be annotated and increased the total number of genes returned by almost 4-fold but increased the annotation time by 125%.

Investigation of gene expression level binning, included as an optional step in the MAVRIC workflow, can give the user additional insights into the integration preferences of their vector. Referencing the nearest genes list of the ADA-SCID dataset against expression data generated from hematopoietic stem cells revealed that genes in the three highest expression bins were overrepresented, especially those genes that were also common integration sites. This suggests that the vector used in the ADA-SCID trial has a preference for integrating near highly expressed genes. Users interested in this type of analysis can use their cell type of interest and compare up to two expression values per gene. Any type of gene-centric expression values could be suitable for binning, with the caveat that it's the user's responsibility to assess the relevance of the resulting analyses.

Linking each integration to a single gene (generally the gene nearest to the integra-

tion site) is a common method of analysis. This allows every IS to have the same impact on secondary analyses such as gene expression arrays, Ingenuity Pathway Analysis, or other methods of looking for integration patterns in viral vectors. However, it is important to highlight what is meant by “nearest.” In general, we suggest that the nearest gene be whichever gene has its 5' end closest to the IS, as viral integrations are known to alter expression of nearby genes via their long-terminal repeats (Wu *et al.*, 2003; Lewinski *et al.*, 2006) or their internal promoters (Modlich *et al.*, 2009). There will be occasional integrations where this method of determining the gene most influenced by the IS will result in important information being missed. For example, consider a hypothetical viral vector integration located on human chromosome 13 at position 32,860,000 (Supplemental Figure 3). This integration is situated in an intron of the FRY gene, but it is within 25 kb of the 5' end of gene ZAR1L and within 30kb of the 5' end of known oncogene BCRA2. Which gene should be considered ‘nearest’? Which is the most significant for determining vector safety? Is the orientation of the IS relative to nearby genes also a factor? Investigation of the genomic region around each integration site is highly recommended. Limiting the analysis to one gene per integration or a narrow window around the IS will exclude a high percentage of the genes close enough to potentially be affected by the integration, including genes related to tumorigenesis (Table 3B).

The difference in datasets obtained with or without Refseq filtering, using different NCBI builds, adjusting the alignment criteria, etc., demonstrates how analysis results can be altered by minor changes to annotation settings. It is therefore necessary when describing IS annotations to very explicitly state all parameters which were used in the analysis. Without these parameters, reproduction of alignments and annotations will prove difficult even when all sequence information is available.

Comparing the output of the automated MAVRIC tool to the published analysis of the ADA-SCID gene therapy study demonstrated large similarities in the overall integration patterns while highlighting subtle differences in the annotation processes. The distributions of integration site distance from TSS were very similar, as were the common integration sites. Some of the sequences identified in the ADA-SCID study were excluded by MAVRIC due to failure to achieve a quality threshold (primarily minimum sequence length before or after repeat masking). The remainder were excluded because of multiple identical lowest E-value BLAST alignments.

Integration site analysis of ADA-SCID and X-SCID clinical trials revealed a large overlap in genes that have viral integrations nearby (14-16%). Ingenuity analysis showed the same functional pathways as frequent integration targets in both clinical trials and in mouse studies (Supplemental Figure 4, Montini *et al* 2009, Deichmann *et al*, 2011). This suggests that closely related vector types may have similar integration profiles across multiple treatments and species. Further cross-study analysis of vector integration sites could lead to new insights into viral vector integration patterns.

In basic structure, MAVRIC is similar to other automated sequence analysis tools such

as QuickMap (Appelt *et al.*, 2009). Compared to QuickMap, MAVRIC annotation speed is slower but it annotates more integration sites. In a test run with the ADA-SCID sequences, QuickMap returned 515 unique protein coding genes vs 574 for MAVRIC, with annotations for 601 out of 706 sequences vs 636 for MAVRIC). MAVRIC produces less annotation data than some other annotation tools but has a variety front-end options for the user with regards to the analysis parameters. MAVRIC is also able to incorporate microarray expression data to further investigate viral integration tendencies. We feel it is important to stress that MAVRIC was not designed for processing the number of reads generated by high-throughput sequencing. For these extremely large datasets, MAVRIC should be viewed as a supplementary tool. MAVRIC's ease of use makes it an attractive option for investigators with modest amounts of sequencing data and those who lack a background in bioinformatics.

In summary, automated analysis tools can greatly increase the speed and efficiency of viral integration site annotation. Our tool, MAVRIC, returns detailed information on the genomic location of the integration sites and gives an overview of viral integration patterns. It is well-suited to generating annotations for re-analysis of historical datasets or comparison of data from different studies. Regardless of the annotation tool used, the user should carefully choose the quality thresholds and note which database versions and filtering options are used, as small changes in the analysis parameters can have a significant effect on the resulting annotations. Analyzing all datasets with identical analysis parameters helps to ensure that the resulting annotations are optimally suited for direct comparison to enable rapid cumulative data processing. As high throughput sequencing becomes routine and more gene therapy applications move towards clinical implementation, automated analysis tools are becoming essential for rapid patient monitoring.

MAVRIC is available for use upon free registration at the following url: <http://mavric.erasmusmc.nl>.

ACKNOWLEDGMENTS

The authors would like to thank Prof. A. Aiuti for the use of the ADA-SCID sequencing data used for testing MAVRIC. Funding for this study was provided the Netherlands Health Research Organization ZonMw (www.zonmw.nl), Translational Gene Therapy Program, projects 43100016 and 43400010, by the German Federal Ministry of Education and Research, (BMBF project iGene 01GU0813) and by the European Commission's 5th and 6th Framework Programs (<http://ec.europa.eu/research>), Contracts QLK3-CT-2001-00427-INHERINET, LSHB-CT-2004-005242-CONCERT, LSHB-CT-2006-018933 and LSHB-CT-2006-19038. The authors acknowledge Drs. F.J.T. Staal and K. Pike-Overzet, Dept. Immunology, Erasmus University Medical Center, for assistance in the microarray analyses.

REFERENCES

Aiuti, A., Slavin, S., Aker, M., Ficara, F., Deola, S., Mortellaro, A., Morecki, S., Andolfi, G., Tabucchi, A., Carlucci, F., Marinello, E., Cattaneo, F., Vai, S., Servida, P., Miniero, R., Roncarolo, M.G., and Bordignon, C. (2002). Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science* 296, 2410-2413.

Aiuti, A., Cassani, B., Andolfi, G., Mirolo, M., Biasco, L., Recchia, A., Urbinati, F., Valacca, C., Scaramuzza, S., Aker, M., Slavin, S., Cazzola, M., Sartori, D., Ambrosi, A., Di Serio, C., Roncarolo, M.G., Mavilio, F., and Bordignon, C. (2007). Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. *The Journal of Clinical Investigation* 117, 2233-2240.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of molecular biology* 215, 403-410.

Appelt, J.U., Giordano, F.A., Ecker, M., Roeder, I., Grund, N., Hotz-Wagenblatt, A., Opelz, G., Zeller, W.J., Allgayer, H., Fruehauf, S., and Laufs, S. (2009). QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. *Gene Therapy* 16, 885-893.

Bartholomew, C., and Ihle, J.N. (1991). Retroviral insertions 90 kilobases proximal to the Evi-1 myeloid transforming gene activate transcription from the normal promoter. *Molecular and Cellular Biology* 11, 1820-1828.

Biasco, L., Ambrosi, A., Pellin, D., Bartholomae, C., Brigida, I., Roncarolo, M.G., Di Serio, C., von Kalle, C., Schmidt, M., and Aiuti, A. (2011). Integration profile of retroviral vector in gene therapy treated patients is cell-specific according to gene expression and chromatin conformation of target cell. *EMBO Molecular Medicine* 3, 89-101.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* 110, 521-529.

Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., Fischer, N., Schwarzwaelder, K., Siegler, M.L., De Ridder, D., Pike-Overzet, K., Howe, S.J., Thrasher, A.J., Wagemaker, G., Abel, U., Staal, F.J., Delabesse, E., Villeval, J.L., Aronow, B., Hue, C., Prinz, C., Wissler, M., Klanke, C., Weissenbach, J., Alexander, I., Fischer, A., Von Kalle, C., and Cavazzana-Calvo, M. (2007). Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *The Journal of Clinical Investigation* 117, 2225-2232.

Deichmann, A., Brugman, M.H., Bartholomae, C., Schwarzwaelder, K., Verstegen, M.M.A., Howe, S.J., Arens, A., Ott, M.G., Hoelzer, D., Seger, R., Grez, M., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., Fischer, A., Paruzynski, A., Gabriel, R., Glimm, H., Abel, U., Cattoglio, C., Mavilio, F., Cassani, B., Aiuti, A., Dunbar, C.E., Baum, C., Gaspar, B.H., Thrasher, A.J., von Kalle, C., Schmidt, M., and Wagemaker, G. (2011). *Molecular Therapy* 19, 2031-2039.

Erkeland, S.J., Verhaak, R.G., Valk, P.J., Delwel, R., Lowenberg, B., and Touw, I.P. (2006). Significance of murine retroviral mutagenesis for identification of disease genes in human acute myeloid leukemia. *Cancer Research* 66, 622-626.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Chen, Y., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Overduin, B., Pritchard, B., Riat, H.S., Rios, D., Ritchie, G.R., Ruffier, M., Schuster, M., Sobral, D., Spudich, G., Tang, Y.A., Trevanion, S., Vandrovcova, J., Vilella, A.J., White, S., Wilder, S.P., Zadissa, A., Zamora, J., Aken, B.L., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernandez-Suarez, X.M., Herrero, J., Hubbard, T.J., Parker, A., Proctor, G., Vogel, J. and Searle, S.M. (2011). Ensembl 2011. *Nucleic Acids Research* 39, D800-806.

- Gaspar, H.B., Parsley, K.L., Howe, S., King, D., Gilmour, K.C., Sinclair, J., Brouns, G., Schmidt, M., Von Kalle, C., Barington, T., Jakobsen, M.A., Christensen, H.O., Al Ghonioum, A., White, H.N., Smith, J.L., Levinsky, R.J., Ali, R.R., Kinnon, C., and Thrasher, A.J. (2004). Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector. *Lancet* 364, 2181-2187.
- Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.P., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., Fischer, A., Davies, E.G., Kuis, W., Leiva, L., and Cavazzana-Calvo, M. (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. *The New England Journal of Medicine* 346, 1185-1193.
- Hawkins, T.B., Dantzer, J., Peters, B., Dinuer, M., Mockaitis, K., Mooney, S., and Cornetta, K. (2011). Identifying viral integration sites using SeqMap 2.0. *Bioinformatics* 27, 720-722.
- Joosten, M., Vankan-Berkhoudt, Y., Tas, M., Lunghi, M., Jenniskens, Y., Parganas, E., Valk, P.J., Lowenberg, B., Van Den Akker, E., and Delwel, R. (2002). Large-scale identification of novel potential disease loci in mouse leukemia applying an improved strategy for cloning common virus integration sites. *Oncogene* 21, 7247-7255.
- 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.
- Kustikova, O.S., Modlich, U., and Fehse, B. (2009). Retroviral insertion site analysis in dominant haematopoietic clones. *Methods in Molecular Biology* 506, 373-390.
- Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., Macdonald, J.R., Pang, A.W., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., Mcintosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., and Venter, J.C. (2007). The diploid genome sequence of an individual human. *PLoS Biology* 5, e254.
- Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathogens* 2, e60.
- Li, W. and Gozick, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.
- Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., and Berns, A. (2002). High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nature Genetics* 32, 153-159.
- 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 Biology* 2, E234.
- Modlich, U., Kustikova, O.S., Schmidt, M., Rudolph, C., Meyer, J., Li, Z., Kamino, K., Von Neuhoff, N., Schlegelberger, B., Kuehlcke, K., Bunting, K.D., Schmidt, S., Deichmann, A., Von Kalle, C., Fehse, B., and Baum, C. (2005). Leukemias following retroviral transfer of multidrug resistance 1 (MDR1) are driven by combinatorial insertional mutagenesis. *Blood* 105, 4235-4246.
- Modlich, U., Navarro, S., Zychlinski, D., Maetzig, T., Knoess, S., Brugman, M.H., Schambach, A., Charrier, S., Galy, A., Thrasher, A.J., Bueren, J., and Baum, C. (2009) Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Molecular Therapy* 17, 1919-1928.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Ponzoni, M., Bartholomae, C., Sergi Sergi, L., Benedicenti, F., Ambrosi, A., Di Serio, C., Doglioni, C., Von Kalle, C., and Naldini, L. (2006). Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nature Biotechnology* 24, 687-696.

Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C., Ranzani, M., Benedicenti, F., Sergi Sergi, L., Ambrosi, A., Ponzoni, M., Doglioni, C., Di Serio, C., Von Kalle, C., and Naldini, L. (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. *The Journal of Clinical Investigation* 119, 964-975.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* 35, D61-65.

Schwarzwaelder, K., Howe, S.J., Schmidt, M., Brugman, M.H., Deichmann, A., Glimm, H., Schmidt, S., Prinz, C., Wissler, M., King, D.J., Zhang, F., Parsley, K.L., Gilmour, K.C., Sinclair, J., Bayford, J., Peraj, R., Pike-O-verzet, K., Staal, F.J., De Ridder, D., Kinnon, C., Abel, U., Wagemaker, G., Gaspar, H.B., Thrasher, A.J., and Von Kalle, C. (2007). Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *The Journal of Clinical Investigation* 117, 2241-2249.

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 (New York, N.Y)* 300, 1749-1751.

Zeeberg, B.R., Riss, J., Kane, D.W., Bussey, K.J., Uchio, E., Linehan, W.M., Barrett, J.C., and Weinstein, J.N. (2004). Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics* 5, 80.

Zychlinski, D., Schambach, A., Modlich, U., Maetzig, T., Meyer, J., Grassman, E., Mishra, A., and Baum, C. (2008). Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Molecular Therapy* 16, 718-725.

SUPPLEMENTAL DATA

Supplemental Table 1. Mecom annotations in NCBI, Ensembl and UCSC.

Database	Annotated start site	Annotated end site
NCBI GRCh37.p5	29850218	30408409
Ensembl 56.37i	29850221	30446930
UCSC GRCh37/hg19	30135491	30408409

The murine Mecom locus has different start and end sites depending on which version of the human genome is used. While the differences are minor, they could potentially effect automated annotations such as those used by MAVRIC and other bioinformatics software.

Supplemental Table 2. Top 5 pathways/functions returned by Ingenuity and KEGG for each of six MAVRIC analysis datasets (see Table 1).

Ingenuity analysis – nearest gene

rank	MAV1	MAV2	MAV3	MAV4	MAV5	MAV6
1	Gastrointestinal Disease	Gastrointestinal Disease	Gastrointestinal Disease	Genetic Disorder	Gastrointestinal Disease	Gastrointestinal Disease
2	Genetic Disorder	Genetic Disorder	Endocrine System Disorders	Inflammatory Disease	Inflammatory Disease	Endocrine System Disorders
3	Inflammatory Disease	Inflammatory Disease	Metabolic Disease	Skeletal & Muscle Disorders	Genetic Disorder	Metabolic Disease
4	Neurological Disease	Endocrine System Disorders	Gene Expression	Immunological Disease	Skeletal & Muscle Disorders	Genetic Disorder
5	Gene Expression	Metabolic Disease	Genetic Disorder	Gastrointestinal Disease	Immunological Disease	Gene Expression

Ingenuity analysis – all genes

rank	MAV1	MAV2	MAV3	MAV4	MAV5	MAV6
1	Gene Expression	Gene Expression	Gene Expression	Gene Expression	Gene Expression	Gene Expression
2	Cellular Growth & Proliferation	Cellular Growth & Proliferation	Cellular Growth & Proliferation	Cellular Growth & Proliferation	Cellular Growth & Proliferation	Cellular Growth & Proliferation
3	Cell Death	Hematological Development & Function	Hematological Development & Function	Genetic Disorder	Immunological Disease	Inflammatory Response
4	Hematological Development & Function	Tissue Morphology	Tissue Morphology	Immunological Disease	Skeletal & Muscle Disorders	Genetic Disorder
5	Tissue Morphology	Cell Death	Inflammatory Response	Skeletal & Muscle Disorders	Inflammatory Disease	Immunological Disease

KEGG analysis – nearest gene

rank	MAV1	MAV2	MAV3	MAV4	MAV5	MAV6
1	Metabolic pathways	Metabolic pathways	Metabolic pathways	Metabolic pathways	Metabolic pathways	Metabolic pathways
2	Pathways in cancer	MAPK Signal. Pathways	MAPK Signal. Pathways	Pathways in cancer	Pathways in cancer	Pathways in cancer
3	MAPK Signal. Pathways	Pathways in cancer	Pathways in cancer	MAPK Signal. Pathways	MAPK Signal. Pathways	MAPK Signal. Pathways
4	Focal adhesion	Endocytosis	Endocytosis	Focal adhesion	Calcium Signal. Pathways	Calcium Signal. pathway
5	Endocytosis	Calcium Signal. Pathways	Calcium Signal. Pathways	Neurotrophin Signal. Pathways	Cell cycle	Cell cycle

KEGG analysis – all genes

rank	MAV1	MAV2	MAV3	MAV4	MAV5	MAV6
1	Metabolic pathways	Metabolic pathways	Metabolic pathways	Metabolic pathways	Metabolic pathways	Metabolic pathways
2	Systemic lupus erythematosus	Systemic lupus erythematosus	Systemic lupus erythematosus	MAPK Signal. Pathways	Systemic lupus erythematosus	Systemic lupus erythematosus
3	Pathways in cancer	MAPK Signal. Pathways	MAPK Signal. Pathways	Pathways in cancer	MAPK Signal. Pathways	MAPK Signal. pathway
4	MAPK Signal. Pathways	Pathways in cancer	Pathways in cancer	Systemic lupus erythematosus	Pathways in cancer	Pathways in cancer
5	Olfactory transduction	Cytokine-Cytokine Recept. Inter.	Cytokine-Cytokine Rec. Inter.	Chemokine Signal. Path.	Cyto.-Cyto. Rec. Inter.	Cyto.-Cyto. Rec. Inter.

a

MAVRIC

METHODS FOR ANALYZING VIRAL INTEGRATION CLUSTERS



Upload Your Runs Help

Your runs

Run ID	User	Type	Description	Status	Start	Elapsed	Current	Entries	Link	message
2463	infusion	BNNNO	testMAnimal_Fu_LUK_XSCDseep	FINISHED	2012-03-12 18:59:04	0s	0	0	Results	
2462	infusion	ENSEMBL	testMAnimal_Fu_LUK_XSCDseep	FINISHED	2012-03-12 16:42:29	2h 16m 25s	1056	1056	4 skipped	
2461	infusion	BLAST	testMAnimal_Fu_LUK_XSCDseep	FINISHED	2012-03-12 15:41:11	1h 1m 16s	1056	1056		
2460	infusion	UPLOAD	testMAnimal_Fu_LUK_XSCDseep	FINISHED	2012-03-12 15:41:11	0s	4	4	5 skipped	
2459	admin	BNNNO	test mouse 68 - 1	FINISHED	2012-03-12 14:34:37	2s	3	3	Results	
2458	admin	ENSEMBL	test mouse 68 - 1	FINISHED	2012-03-12 14:03:26	1m 12s	33	33	1 skipped	
2457	admin	BLAST	test mouse 68 - 1	FINISHED	2012-03-12 14:02:55	29s	33	33		
2456	admin	UPLOAD	test mouse 68 - 1	FINISHED	2012-03-12 14:02:39	16s	4	4	78 skipped	

b

Skipped sequences

type	reason	count
upload	sequence > 11 bp after trimming file 1	217
upload	sequence excluded because it contains no part of file 1	372
upload	sequence > 11 bp after trimming file 2	37
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6LHMJ	1653
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6BOMJ	23
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6SYVJ	6
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6BNHJ	15
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6CQWJ	6
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6ALHJ	10
upload	sequence was > 95% identical to sequence Phn21c2_CTATC_G31COWD2C6BGMJ	21

c

Analysis Results

Download all results
Skipped sequences summary

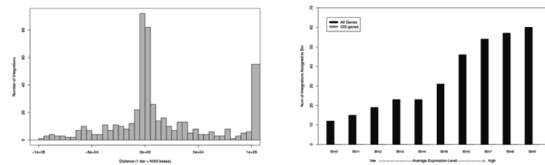
Summary

Run code: L2WtgdQ
Run description: up22_UCB samples
Sequence file: T136-127UCB.se
Trim file 1: L2WtgdQ
Trim file 2: L2WtgdQ_consens.txt
Trim file 3: consens.txt
Expression file: none
Repeat masking using: none
BLAST against human_G043737 with e-value: 0.01
0 sequences were skipped during validation.
1731 sequences were skipped during trimming.
9339 sequences were removed during tabling.
0 sequences were skipped during repeat masking.
Minimum sequence lengths: 11 bp
Using gene window: 10000
Refset genes only: yes

Viral Insertion Profile

Integration summary	CIS NAME	# OF INTEGRATIONS	BIOTYPE
Number of insertions with protein coding genes as nearest feature:	WNA7	7	protein_coding
Number of insertions with regulatory elements as nearest feature:	SAPC01	7	protein_coding
Number of unique genes found within distance threshold:	SNORC148	7	snoRNA
Number of genes listed as nearest 3+ times (CID genes):	C2	7	protein_coding
Number of ribotic integrations:	MCH5	7	protein_coding
Average distance to nearest gene:	LEW2	7	protein_coding
Average distance to nearest transcription start site:	CLC1	7	protein_coding
	WRS	7	protein_coding

d



Results

Sequence	Gene ID	Biotype	Gene Symbol	Chr	Gene start	Gene end	Strand	Description
S1_003	ENSG00000102651	protein_coding	GRASP1	8	114220569	144206044	+	glycerolphosphatidylinositol anchorage high density lipoprotein binding
S1_004	ENSG00000204136	pseudogene	GOTAF1P	9	124207269	124232206	-	glucosyl, alpha-galactosyltransferase 1 pseudogene [Source:HGNC. Sy.
S1_006	ENSG00000102481	protein_coding	OLPRL2L	12	75784850	75824468	+	OLI pathogenesis-related 1 like 2 [Source:HGNC Symbol;Acc:295923]
S1_009	ENSG00000210893	nontranscribed_self	LTR4-401	6	6346659	6622871	+	LTR4 antisense RNA 1 (non-protein coding) [Source:HGNC Symbol;Acc:2653
S1_010	ENSG00000104391	protein_coding	NEK10	3	27151576	27410951	-	NIMA (never in mitosis gene A)-related kinase 10 [Source:HGNC Symbol;Acc:231721]
S1_011	ENSG00000120278	protein_coding	MVCT1	6	153019039	153045702	+	mvc target 1 [Source:HGNC Symbol;Acc:187923]
S1_013	ENSG00000115206	protein_coding	ARC5	17	29156009	29222265	+	ATPase family AAA domain containing 5 [Source:HGNC Symbol;Acc:251526]
S1_017	ENSG00000200567	miRNA	MIR-42	17	56480893	56480879	-	microRNA 142 [Source:HGNC Symbol;Acc:31529]
S1_018	ENSG00000108091	protein_coding	CDC6B	10	61548921	61666414	-	coiled-coil domain containing 8 [Source:HGNC Symbol;Acc:18792]
S1_019	ENSG00000109926	protein_coding	HLF13	15	31619059	31727668	+	Kisspep-like factor 13 [Source:HGNC Symbol;Acc:13073]

Download all results

Figure S1: MAVRIC graphical output.

(a) Shows the Your Runs page, with real-time updates of current runs and an archive of all runs by the user in the last 30 days. Clicking on one of the “___ skipped” link (marked in blue font, right-hand side of the page) brings the user to the Skipped Sequences page (b), where the sequences which were excluded for each annotation criteria are enumerated. Clicking on the Results link in the Your Runs page leads to the Analysis Results page (c-d). An overview of the datasets integration profile and common integration sites (CIS) are shown at the top of the page, with graphical output and the annotations for the first few sequences listed below. (d) shows the histogram and HSC expression binning profiles for the ADA-SCID dataset.

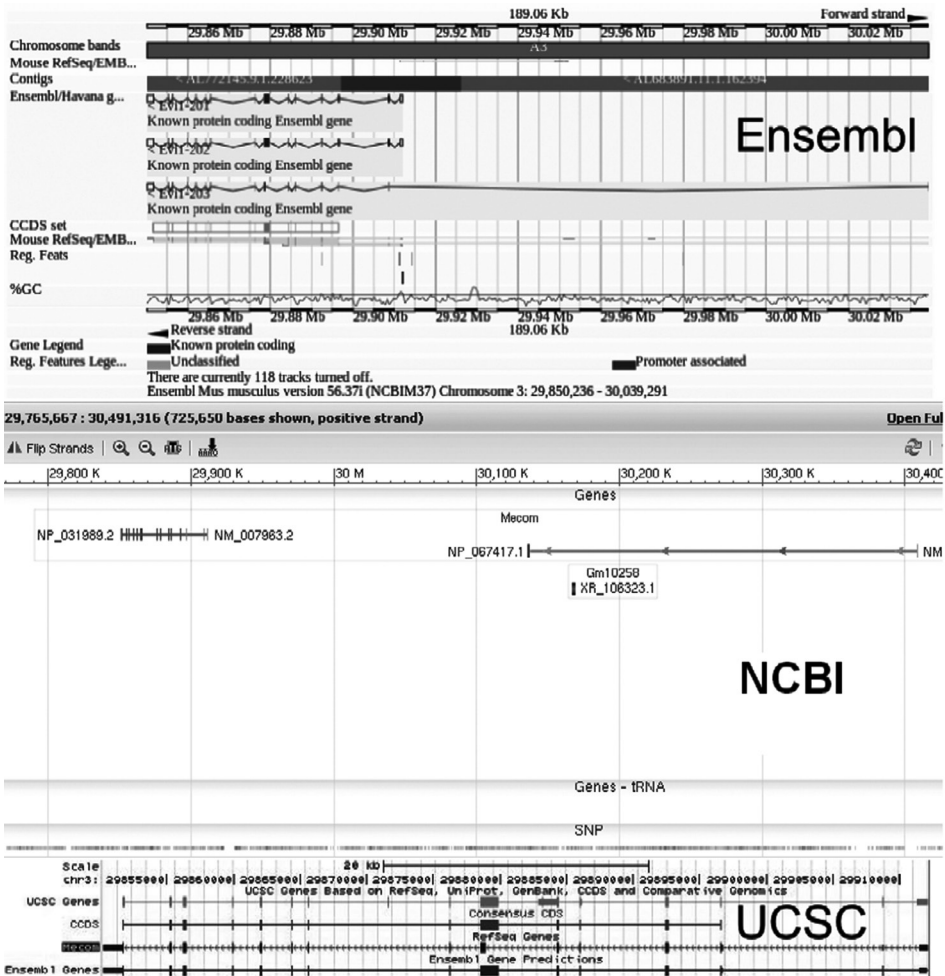


Figure S2: The MECOM locus as annotated by Ensembl, NCBI and UCSC.

Figure shows the MECOM MDS1 and EVI1 complex locus located on mouse chromosome 3 as annotated by three different genome databases: Ensembl, NCBI and UCSC. Shown are Ensembl version 56.37i, NCBI assembly GRCh37.p5 and UCSC assembly GRCh37/hg19.

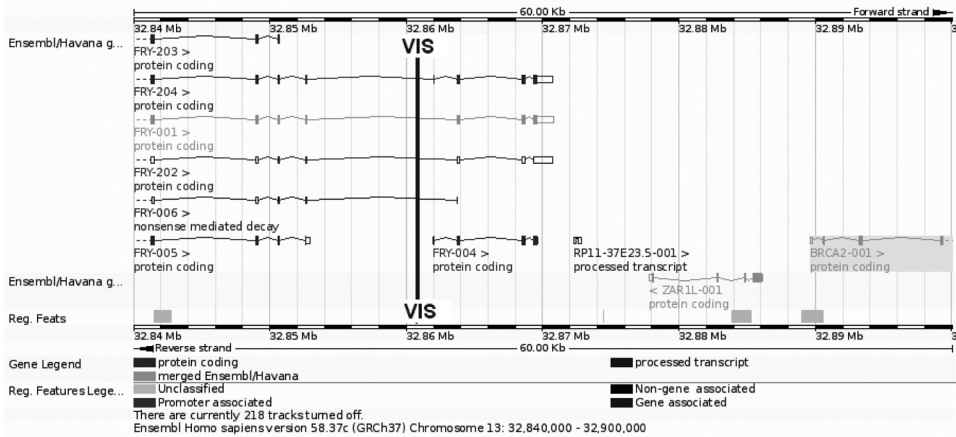
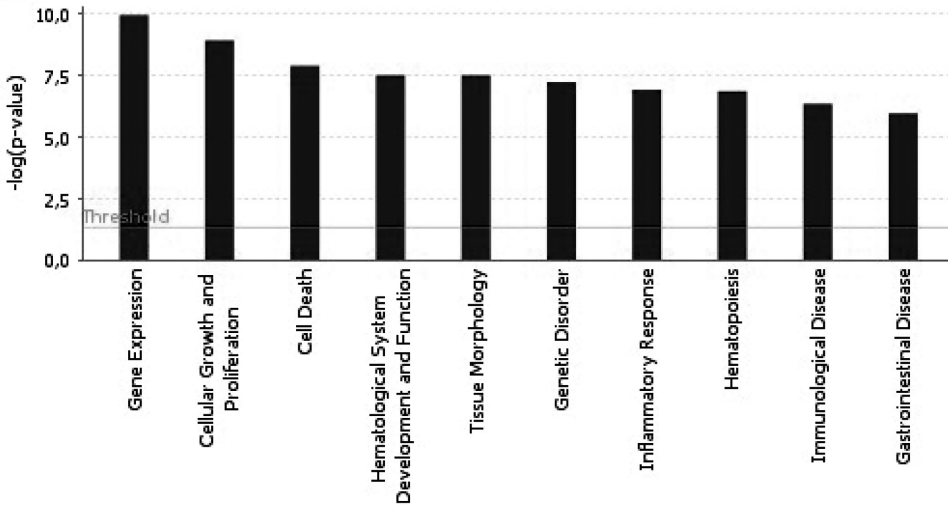


Figure S3: A hypothetical integration site situated in an intron of human gene FRY near the 5' end of the BRCA2 gene.
 The hypothetical vector integration site (VIS) is indicated by the vertical blue line and is in the 5'-3' (forward) orientation. Based on the position of this IS, the 'nearest' gene could be considered to be FRY (IS is intronic), ZAR-IL 001 (gene with nearest 5' end), or the BRCA2 gene (gene with the nearest 5' end on the same strand).



© 2000-2011 Ingenuity Systems, Inc. All rights reserved.

Figure S4: Ingenuity functional pathways analysis of all genes within 100 kb of the ADA-SCID clinical trial IS.
 The list of 2059 genes annotated as within 100kb of a published ADA-SCID IS using the permissive criteria (ie MAV1, see FIGURE 1A) was uploaded to Ingenuity Pathways Analysis software and analyzed using the following settings: Direct and Indirect Relationships; Data Sources = all; Confidence = experimentally observed; Species = all, Tissues and Cell Lines = all.

Met TyrXle Ala Cys
Met Xle Ala Cys
Cys Xle Ala Cys
Ala Lys Ser Cys
Xle Ser Pyl Lys
PylAsn Arg Glu Asn Sar
Ser Leu SerThr LeuGlu Asn TrpIle LeuLeu IleAsaAsn His Sec
Lys Cys Ala Xle Ser GGAC GCC GGCATGCC TrpIle LeuLeu Pyl His Ser
Ala Xle Thr AGCAGGTACGCGAGA His Ala Asn PylThr
GluArg AGGCTAGCTAG PylAsn Arg His AlaArg ArgIle Glu
CAGTGGGTAT Thr Ile Ser Thr Pyl Thr His
GTGAGCTTGCTC Ser Sec Thr Thr His
GATAGCGGGTA His Thr SerSec
TGCTAGCCGGC
GGGAATGCAGG
TACGCTTGCTAGA
CCGCAGGCATGC
CTGTATGCTTGCTC
01001101011000010111
00100 1110011011010000
01100 001011011000110
1100 0010000001010111
011 010010110110001101
10 0011010010110000101
101101001 00000010010
00011101 010111001101111010
0011011 11011011100010110000
100000 01010
00001 1010
000 100100



Lentiviral integration profiles are stable across multiple species, disease phenotypes and tissue types

Marshall W. Huston¹, Ali Nowrouzi², Merel Stok¹, Fatima Aerts¹, Yuedan Li¹, Elnaz Farahbakhshian¹, Martijn H. Brugman¹, Yvette van Helsdingen¹, Manfred Schmidt², Christof von Kalle², Monique M.A. Versteegen¹, Niek P. van Til¹ and Gerard Wagemaker¹.

¹Department of Hematology, Erasmus University Medical Center, Rotterdam, the Netherlands

²Department of Translational Oncology, National Center for Tumor Diseases and German Cancer Research Center (DKFZ), Heidelberg, Germany

Manuscript in preparation

ABSTRACT

Clinical trials have demonstrated the potential of *ex vivo* hematopoietic stem cell gene therapy to treat a variety of monogenic inherited disorders, particularly severe combined immunodeficiency disease (SCID), while also underlining the inherent risks in using retroviral vectors which integrate in a semi-random manner and the potential role the disease background may play in genotoxicity. In order to better understand the integration patterns of third generation self-inactivating lentiviral vectors, a broad range of vectors were created and used to transduce hematopoietic stem cells for transplantation into murine models for severe combined immunodeficiency disorders SCID-X1, RAG1-SCID and RAG2-SCID, as well as the lysosomal storage disorder Pompe disease. GFP control vectors were used to transduce cultured hematopoietic stem and progenitor cells of murine, rhesus and human origin. Vector-genome boundary sequences obtained from cultured cells and various murine tissues were pooled to create a large database of over 16,000 lentiviral vector integration sites. Using this database, integration sites profiles were stratified by disease phenotype, tissue, transgene, promoter and *in vitro* vs *in vivo* samples in order to compare lentiviral integration patterns under a wide variety of conditions and determine if any of the above criteria had an effect on the integration pattern. The resulting analysis demonstrates that lentiviral integration profiles differ based on the tissue and transgene, but not from the promoter, and shows some evidence of integration site-specific selection *in vivo*.

INTRODUCTION

The clinical gene therapy trials implemented since 2000 have successfully reversed the disease phenotype in a number of patients with life-threatening inherited diseases, but also underlined the risks associated with the integration pattern of the viral vectors used. Trials for X-linked severe combined immunodeficiency (SCID-X1)[1, 2], adenosine deaminase deficiency (ADA-SCID)[3, 4], X-linked chronic granulomatous disease (CGD)[5], Wiskott-Aldrich syndrome (WAS)[6, 7], β -thalassemia[8], X-linked adrenoleukodystrophy[9] and metachromatic leukodystrophy[10] all reported that the majority of patients obtained a clear clinical benefit. However, in the SCID-X1 trials initiated in 2000 five patients (out of the 20 treated) developed acute clonal T cell lymphoproliferative disorders caused by the insertional activation of proto oncogenes *LMO2* or *CCND2*[11, 12]. Four of these patients were successfully treated with chemotherapy without compromising the effectiveness of their gene therapy treatment, while the fifth patient succumbed in spite of chemotherapy. Two CGD patients treated with gene therapy developed myelodysplasia due to insertional activation of proto-oncogene *MDS1-EVII*, with one patient dying of sepsis and the other undergoing allogeneic hematopoietic stem cell transplantation[13]. At least one WAS patient has been reported to have developed acute lymphocytic leukemia similar to that seen in the SCID-X1 trials[6].

As a result of these vector-derived adverse events, which had been observed in experimental settings only at a very low frequency prior to the SCID clinical trials, a great deal of research has been done on the mechanisms of viral vector integration into hematopoietic cells. Numerous studies have demonstrated that retroviral vector integration into hematopoietic cells is semi-random, with the viral subtype[14], long terminal repeats (LTRs)[15], promoter[16], transgene[17] and number of integrations per cell all potentially having an influence on either the site of integration or the genotoxic potential of the transduced cell. For example, it has been clearly demonstrated that gammaretroviral vectors have a clear bias towards for integrating near gene transcription start sites[18], whereas lentiviral vectors show a preference for integrating into active genes[19]. Transcriptionally active LTRs and strong viral enhancer/promoter regions have been shown to increase the risk of oncogenesis[15], while self-inactivating vectors and well insulated or weaker promoters often confer a reduced genotoxicity risk. Despite the prodigious amount of research that has been performed on retroviral vectors, many of their mechanisms for cell targeting and integration are still poorly understood.

In the course of developing new therapeutic lentiviral vectors for treatment of SCID-X1[20], recombination activating gene (RAG)-SCID[21] and the lysosomal storage disorder Pompe disease[22], we have tested a wide range of vectors in murine models by transducing and subsequently transplanting hematopoietic stem and progenitor cells (HSPCs). Viral vector integration sites (IS) from these animals were generated via linear amplification mediated PCR (LAM-PCR)[23] and high throughput sequencing.

Integration sites were collected from a variety of tissues, including lineage negative bone marrow cells (Lin-) pre-transplantation and bone marrow spleen, thymus, lymph nodes and peripheral blood post-transplant. Additional lentiviral integrations were generated *in vitro* via transduction of HSPCs derived from mouse bone marrow, rhesus bone marrow or human umbilical cord blood. In total, 16,370 unique IS were annotated, including 9699 *in vivo* and 6671 *in vitro* integrations.

As all of these IS were all generated using the same general transduction protocol in the same laboratory, they offer the opportunity to perform a broad comparative analysis of lentiviral integration profiles across disease phenotypes, tissues, and transgenes. Here we have performed an in-depth analysis, investigating oncogene prevalence, common integration site frequency, the tendency of IS to be located near a TSS or near/within a Refseq gene, Ingenuity Pathways Analysis and nearby gene expression levels to determine if any integration profile characteristics are over or underrepresented in certain sample sets. This analysis generates new insights into the integration profiles of lentiviral vectors under various conditions *in vivo* and *in vitro*.

RESULTS

Table 1: Integration sites (IS) recovered from cultured hematopoietic stem/progenitor cells

	mouse ^a	rhesus ^b	human ^c	<i>Il2rg</i> ^{-/-d}	<i>Rag1</i> ^{-/-d}	<i>Rag2</i> ^{-/-d}	Pompe ^d
no. of unique IS	1645	1279	835	1676	272	642	322
no. of unique genes	1353	1066	737	1345	256	588	309
no. of IS within 10kb of TSS ^e	303	207	120	294	52	126	63
no. of IS within 10kb of gene	1214	878	658	1214	216	485	213
no. of unique oncogenes	40	27	26	43	13	14	7
total no. of oncogenes	50	33	32	60	14	14	7
no. of CIS ^f	126	84	44	147	9	36	4
CIS oncogenes	2	0	3	9	1	0	0
% within 10kb of gene	73.8%	68.6%	78.8%	72.4%	79.4%	75.5%	66.1%
% of IS that are CIS	7.7%	6.6%	5.3%	8.8%	3.3%	5.6%	1.2%
% of IS that are oncogenes	3.0%	2.6%	3.8%	3.6%	5.1%	2.2%	2.2%

^alineage negative bone marrow cells from wild type BALB/c mice

^bCD34⁺ bone marrow cells

^cCD34⁺ cells derived from umbilical cord blood

^dlineage negative bone marrow cells from mutant mice

^etranscription start site

^fcommon integration sites

Integration site profiles of cultured mouse, rhesus and human HPSCs

Lineage depleted mouse bone marrow (BM) cells, CD34⁺ cells derived from rhesus bone marrow and CD34⁺ human umbilical cord blood cells were cultured overnight with the pRRL.PPT.SF.GFP.bPRE4*.SIN lentiviral vector[24, 25] in order to determine if lentiviral integration site preferences had a species-specific component. A total of 8 murine samples, 10 rhesus samples and 4 human samples were analyzed, yielding 1645, 1278 and 835 unique integration sites, respectively. A comparison of the integration profiles for all three lineages can be found in Table 1. As expected, the percentage of IS within 10kb of a gene TSS was far lower than reported for gammaretroviral vectors, ranging from 14.4% of mapped IS for the human samples up to 18.4% for the murine samples. The percentage of IS within a gene or less than 10kb distant ranged from 68.6% (rhesus samples) to 78.8% (human). The number of common integration sites (CIS) scaled with the total number of IS analyzed: a list of CIS 5th order or higher for each dataset can be found in Supplemental Table 1. The number of IS found near proto-oncogenes in the rhesus, mouse and human samples containing oncogene frequencies of 2.6%, 3.0% and 3.8%, respectively. The human dataset also contained the highest frequency of CIS oncogenes (3 out of 44 CIS were in or flanking oncogenes, compared to 2 out of 126 and 0 out of 84 for the murine and rhesus

datasets, respectively), although this difference was not significantly different from either of the other species (Fisher's exact test). For each dataset, Ingenuity Pathways analysis was applied to the genes within 10kb of an IS. All three species had Gene Expression and Organismal Survival in their top 10 pathways (Supplemental Figure 1). Interestingly, the top functional pathway for the murine Lin- samples was cancer, with 46.4% of mapped genes linked to this pathway (Figure 1a). This overrepresentation of cancer-linked genes in the mouse Lin- cells was highly significant even after applying Bonferroni's correction for multiple tests ($p\text{-value} = 2.41 \times 10^{-9}$, threshold $\alpha = 0.05$ corrected to 0.0002). Cancer was not in the top ten Ingenuity pathways for either human or rhesus samples, and was represented by a far lower fraction of the mapped genes (20.9% and 19.7%, respectively).

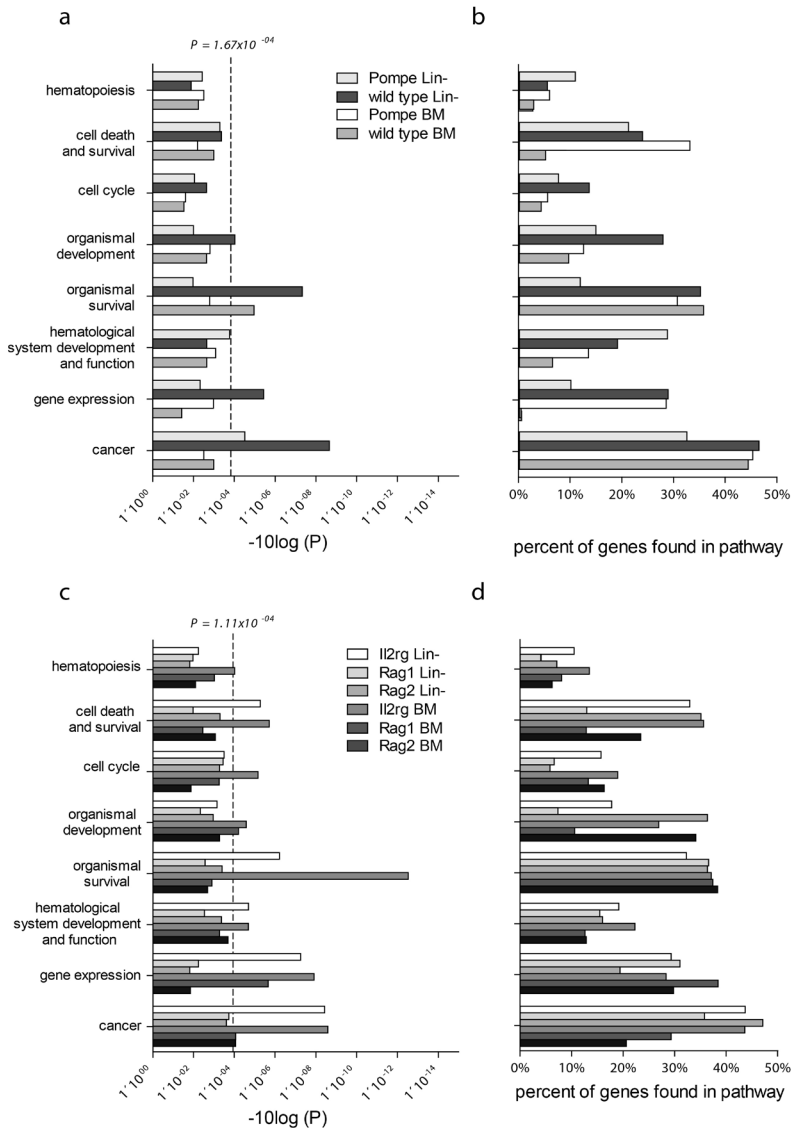


Figure 1: Ingenuity Functional Pathways analysis. (a) Overrepresented Ingenuity functional pathways for genes flanking integration sites in transduced bone marrow (BM) and lineage negative (lin-) cells from wild type and Pompe mice. Significance is shown on y-axis as $-\log_{10} P$ value. The base significance threshold of $P < 0.05$ was adjusted for multiple comparisons using the Bonferroni correction to reach a new threshold of $P < 0.00017$ (75 functional classes and 4 cell types = 300 function comparisons). (b) Percentages of genes belonging to the indicated Ingenuity pathway for wild type and Pompe integrations (c) Overrepresented functional pathways in transduced bone marrow (BM) and lineage negative (lin-) cells from severe combined immunodeficiency (SCID) mice. The base significance threshold of $P < 0.05$ was adjusted for multiple comparisons using the Bonferroni correction to reach a new significance threshold of $P < 0.00011$ (75 functional classes and 6 cell types = 450 function comparisons). (d) Percentages of genes belonging to the indicated Ingenuity pathway for each SCID dataset.

Lineage negative bone marrow cells derived from murine models for SCID-X1 (*Il2rg*^{-/-}), RAG-SCID (*Rag1*^{-/-} and *Rag2*^{-/-}), and Pompe's disease (*Gaa*^{-/-}), as well as wild type BALB/c mice, were cultured overnight with therapeutic lentiviral vectors containing the human version of the appropriate transgene or a GFP control vector. The cells were subsequently cultured for 10-14 days without viral vectors before DNA was purified for LAM-PCR and integration site analysis (Table 1). The frequency of oncogenes was lowest in the Pompe and *Rag2* samples (2.2%) and highest in the *Rag1* samples (5.1%), although this difference in oncogene frequency between groups was not significant after Bonferroni correction (p values: *Rag1* vs *Rag2* = 0.0395, *Rag1* vs Pompe = 0.725, corrected α = 0.0125). The percentage of IS within 10kb of a gene TSS or within 10kb of a gene are similar for all four mouse phenotypes. The *Il2rg* Lin- dataset has 1676 IS and thus can be directly compared to the wild type mouse Lin- dataset (1645 IS). The frequency of IS flanking a gene or TSS are similar for the two datasets, while the *Il2rg* dataset has a slightly higher incidence of CIS (8.8% compared to 7.7% for the wild type Lin- dataset) and oncogenes (3.6% vs 3.0%). Strikingly, the *Il2rg* dataset has a much higher percentage of CIS oncogenes: 6.1% (9 out of 147 CIS) vs. 1.6% for the wild type dataset (2 out of 126 CIS). All of the *Il2rg* oncogenes are 2nd order CIS, while the wild type dataset had one 2nd order and one 3rd order CIS oncogene (*Akap13* and *Mef2c*, respectively). The full list of CIS oncogenes for each dataset can be found in Supplemental Table 2.

Comparison of integration profiles between murine in vitro and in vivo samples

Lineage negative bone marrow cells derived from *Il2rg*^{-/-}, *Rag1*^{-/-}, *Rag2*^{-/-} or *Gaa*^{-/-} mice were cultured overnight with therapeutic lentiviral vectors or a GFP control vector and subsequently transplanted into the appropriate mouse phenotype. Other mice were transplanted with wild type Lin- cells after overnight transduction with a GFP control vector. Transgene expression was driven by the SF, PGK, UCOE, γ cPr or TCR β promoter elements. DNA from bone marrow, spleen, and/or thymus was collected 6-12 months after transplantation. Subsequent LAM-PCR and sequencing of vector-genome boundary sequences yielded over 12,600 unique integration sites for analysis. The distribution of samples and integration sites based on mouse strain can be seen in Table 2. Notably, the percentage of IS within a gene or less than 10kb distant was significantly lower for wild type mice given GFP vectors (58.5%) than in any of the disease phenotypes (66.5% to 73.9%, p value <0.001 for all disease phenotypes compared to the wild type group). The percentage of the IS flanking proto-oncogenes ranged from 3.3% to 4.0%, depending on the mouse phenotype, with the *Rag1* dataset having the highest oncogene frequency, although these differences were not significant. CIS of 5th order or greater for these datasets can be found in Supplemental Table 1, and the full list of CIS oncogenes can be found in Supplemental Table 2.

Table 2: Overview of all integration sites (IS) recovered from murine *in vivo* samples

	Il2rg ^{-/-}	Rag1 ^{-/-}	Rag2 ^{-/-}	Pompe	Wild type
no. of unique IS	2960	1384	2561	798	1996
no. of unique genes	1829	986	1804	713	1277
no. of IS within 10kb of TSS	430	286	515	161	213
no. of IS within 10kb of gene	1968	965	1893	556	1168
no. of unique oncogenes	61	36	59	24	46
total no. of oncogenes	105	56	84	27	76
no. of CIS	441	205	63	49	224
CIS oncogenes	16	9	12	4	10
% within 10kb of gene	66.5%	69.7%	73.9%	69.7%	58.5%
% of IS that are CIS	14.9%	14.8%	2.5%	6.1%	11.2%
% of IS that are oncogenes	3.5%	4.0%	3.3%	3.4%	3.8%

Comparing the list of CIS for the various *in vivo* and *in vitro* datasets reveals a number of overlapping CIS (Table 3). No more than one of these lentiviral integration “hot spots” was found on any single chromosome, and none of them were in the vicinity of any known oncogenes. One CIS, found in wild type mice transplanted with SF.GFP transduced cells, contained 8 IS clustered in or near proto-oncogene pleomorphic adenoma gene 1 (*Plagl1*), which encodes for a developmentally-regulated zinc finger protein (Supplemental Table 2). An additional CIS oncogene of 5th order, serologically defined colon cancer antigen 8 (*Sdccag8*), was found in the *in vivo* Il2rg dataset. Overall, out of 73 CIS of 5th order or higher found in the various *in vivo* and *in vitro* subsets, only 2 were flanking oncogenes.

Table 3: Common integration sites found in multiple datasets (3rd order or higher)

Gene	Chromosome	Dataset	CIS order	Locus
Sfi1	11	Rag2 <i>in vivo</i>	8th	3031603-3110289
		Il2rg <i>in vivo</i>	7th	3031770-3202635
		Il2rg <i>in vitro</i>	7th	3030445-3110058
		Pompe <i>in vivo</i>	3rd	3083899-3095580
1110028C15Rik	1	WT <i>in vivo</i>	3rd	66827410-66833325
		Rag2 <i>in vivo</i>	3rd	66808037-66850310
		Il2rg <i>in vivo</i>	3rd	66823216-66836930
Ccnh	13	Rag2 <i>in vivo</i>	4th	85353162-85418014
		Il2rg <i>in vivo</i>	4th	85357018-85411736
		WT <i>in vitro</i>	3rd	85367950-85401745
Cggbp1	16	Pompe <i>in vivo</i>	3rd	64795977-64840559
		Il2rg <i>in vivo</i>	3rd	64795733-64825315
		Il2rg <i>in vitro</i>	3rd	64822461-64859627
Epha7	4	Rag2 <i>in vivo</i>	3rd	28771873-28806620
		Rag1 <i>in vivo</i>	3rd	28832563-28833606
		Il2rg <i>in vivo</i>	3rd	28753182-28775600
Plxdc2	2	Il2rg <i>in vivo</i>	6th	16422537-16565516
		Rag2 <i>in vivo</i>	4th	16402704-16428376
		WT <i>in vitro</i>	3rd	16467374-16494708
Uba6	5	Il2rg <i>in vivo</i>	11th	86459481-86582241
		WT <i>in vivo</i>	3rd	86562050-86583025
		WT <i>in vitro</i>	3rd	86575222-86583707
Ube3a	7	Il2rg <i>in vivo</i>	4th	66528406-66562531
		Il2rg <i>in vitro</i>	3rd	66493095-66524429
		Rag2 <i>in vivo</i>	3rd	66498812-66538856
Vegfc	8	WT <i>in vivo</i>	6th	55188886-55262672
		Rag2 <i>in vivo</i>	6th	55199088-55366242
		WT <i>in vitro</i>	3rd	55309415-55345413

Pooling all of the *in vivo* IS (consisting of integrations recovered from spleen, BM, peripheral blood and lymph nodes) and all of the murine *in vitro* IS (consisting of IS recovered from Lin⁻ cells) and comparing their CIS genes revealed that only 60% of the 5th order or higher CIS present in the Lin⁻ integrations were present as CIS in the integration sites derived from murine tissues (Figure 2). Of the ten highest-ranking CIS for the *in vitro* and *in vivo* datasets, only two overlapped: a 114kb region on chromosome 11 near gene *Sfi1* (*Sfi1* homolog, spindle assembly associated (yeast)), and a 155kb region on chromosome 14 flanking gene *Oxsm* (3-Oxoacyl-ACP Synthase, Mitochondrial). While no enrichment of oncogenes was seen when comparing the two datasets, the *in vivo* data had more oncogenes present in high level CIS compared to the *in vitro* data: 11 oncogenes were found in the top 10% highest order CIS for the *in vitro* data (corresponding to CIS order 7 and higher), including *Camk2d*, *Cr2*, *Evi1*, *Gimap7*, *Ikzf1*, *Lnpep*, *Mef2c*, *Meis1*, *Plag1*, *Rps13* and *Sdccag8*. Only one oncogene, *Mef2c*, was found in the top 13% of CIS for the *in vivo* data (corresponding to CIS order 4 and higher).

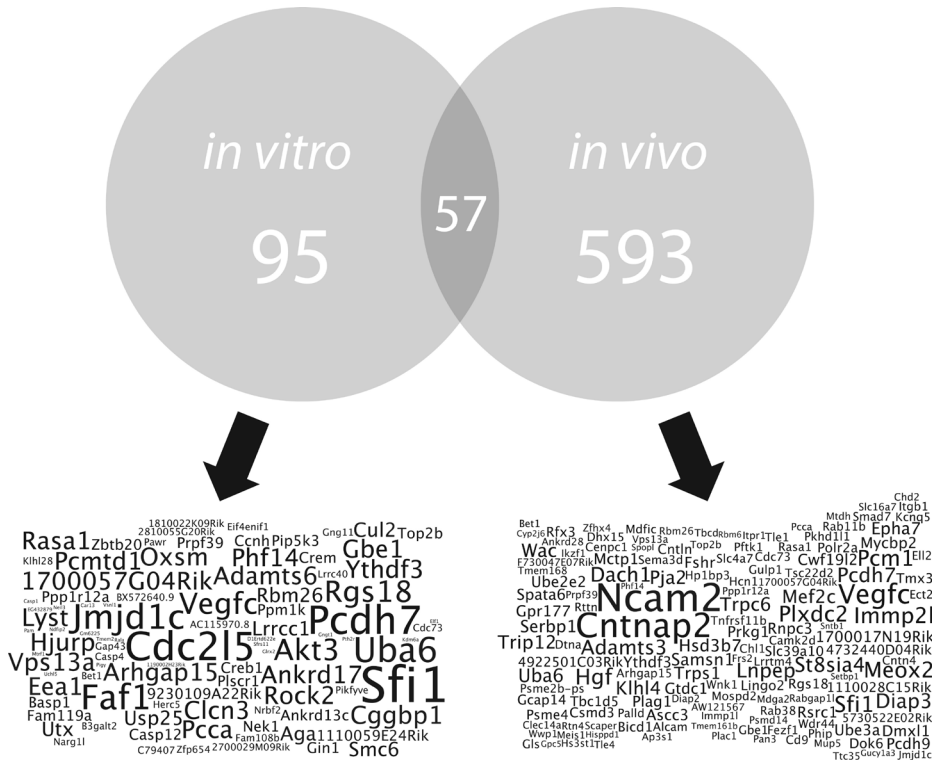


Figure 2: Common integration sites (CIS) of 5th order or higher in murine *in vitro* and *in vivo* datasets. Overlap among *in vitro* and *in vivo* murine common integration site genes (CIS). Genes flanking CIS of 5th order or higher are reported in the overlapping circles, with 57 CIS genes shared between the *in vitro* and *in vivo* samples. Word clouds below the circles correspond to the relative frequency of each named gene in the CIS pool (larger genes indicate more IS were found near that gene).

Due to the highly heterogeneous nature of the sample origins, the *in vivo* IS were further stratified based on the tissue (Table 4). In order to minimize the impact of variance, sample subsets containing less than 300 IS were excluded from the comparative analysis. Datasets with less than 300 IS included the spleen Pompe samples, thymic I12rg samples and the peripheral blood samples for all phenotypes.

Table 4: Integration sites recovered from various murine tissues

	BM				
	Il2rg	Rag1	Rag2	Pompe	Wild type
no. of unique IS	2442	649	465	755	1996
no. of unique genes	1617	539	425	677	1277
no. of IS within 10kb of TSS	326	127	79	156	213
no. of IS within 10kb of gene	1608	454	332	524	1168
no. of unique oncogenes	55	21	8	24	46
total no. of oncogenes	87	24	8	27	76
no. of CIS	333	73	23	47	224
CIS oncogenes	11	3	0	4	10
% within 10kb of TSS	13.3%	19.6%	17.0%	20.7%	10.7%
% within 10kb of gene	65.8%	70.0%	71.4%	69.4%	58.5%
% of IS that are CIS	13.6%	11.2%	4.9%	6.2%	11.2%
% of IS that are oncogenes	3.6%	3.7%	1.7%	3.6%	3.8%

	spleen			thymus		
	Il2rg	Rag1	Rag2	Rag1	Rag2	Pompe
no. of unique IS	576	728	1207	208	637	31
no. of unique genes	366	541	1007	181	589	31
no. of IS within 10kb of TSS	115	165	242	42	138	3
no. of IS within 10kb of gene	406	522	916	141	456	25
no. of unique oncogenes	13	25	38	7	19	0
total no. of oncogenes	20	37	45	8	21	0
no. of CIS	101	103	107	22	26	0
CIS oncogenes	5	6	1	1	1	0
% within 10kb of TSS	20.0%	22.7%	20.0%	20.2%	21.7%	9.7%
% within 10kb of gene	70.5%	71.7%	75.9%	67.8%	71.6%	80.6%
% of IS that are CIS	17.5%	14.1%	8.9%	10.6%	4.1%	0.0%
% of IS that are oncogenes	3.5%	5.1%	3.7%	3.8%	3.3%	0.0%

In order to determine if *in vivo* selection of clones in the bone marrow during or after engraftment could skew the integration profile of vector-treated animals, we first compared the IS profiles of the Lin- and BM samples for each of the four mutant mouse strains. The percentage of IS within 10kb of a gene was significantly lower in IS recovered from BM samples than in IS recovered from Lin- samples for all SCIDs, Pompe and wild type data-sets (p value <0.001). while the percentage of oncogenes was higher but not significantly

so (averages of 3.5% for BM vs 3.2% for Lin- datasets). The highest frequency of oncogenes amongst the mutant mouse strains was found in the *Rag1* samples for For the BM datasets, the following oncogenes were found in CIS of 3rd order or higher: *Mef2c* (Il2rg, two 2nd order CIS 117kb apart), *Rspo3* (Il2rg, 3rd order), *Arid1a* (Rag1, 3rd order), *Plag1* (wild type, 8th order), and *Lnpep* (wild type, 6th order).

Next, the integration profiles in spleen and BM were compared. Only the SCID datasets had sufficient spleen IS for comparative analysis. No significant differences were found when comparing spleen vs BM for frequency of IS near genes, gene TSS or oncogenes.

Comparing the Ingenuity Functional Pathways of the IS for BM and Lin- in wild type and Pompe datasets revealed a significant overrepresentation of genes linked to cancer in the Lin- datasets, but not in the BM (Figure 1a,b). This overrepresentation was also seen in the percentage of genes involved in cancer pathways in the wild type Lin- and Pompe BM datasets. The various SCID phenotypes showed an overrepresentation of genes involved in cancer in both the BM and Lin- Il2rg datasets and a high percentage of cancer-related genes in the Il2rg datasets as well as the Lin- datasets for Rag1 and Rag2. (Figure 1c,d).

Genes flanking IS were referenced against an expression level binning database generated via an Affymetrix mouse 430 2.0 gene array. Briefly, cDNA derived from mouse lineage negative Sca-1⁺cKit⁺ (LSK) cells after 2 days of stimulation with growth factors was measured on this whole genome array, and the resulting gene expression values were sorted into ten equally sized bins based on expression level. Genes flanking an IS were referenced against this database and sorted into the appropriate bin. The resulting binning chart (Figure 3) demonstrates that lentiviral vectors prefer to integrate into active genes in hematopoietic stem cells. A similar expression pattern is seen in genes flanking IS before (Figure 3a) and after (Figure 3b) transplantation, suggesting that IS near highly expressed genes do not undergo selective pressure.

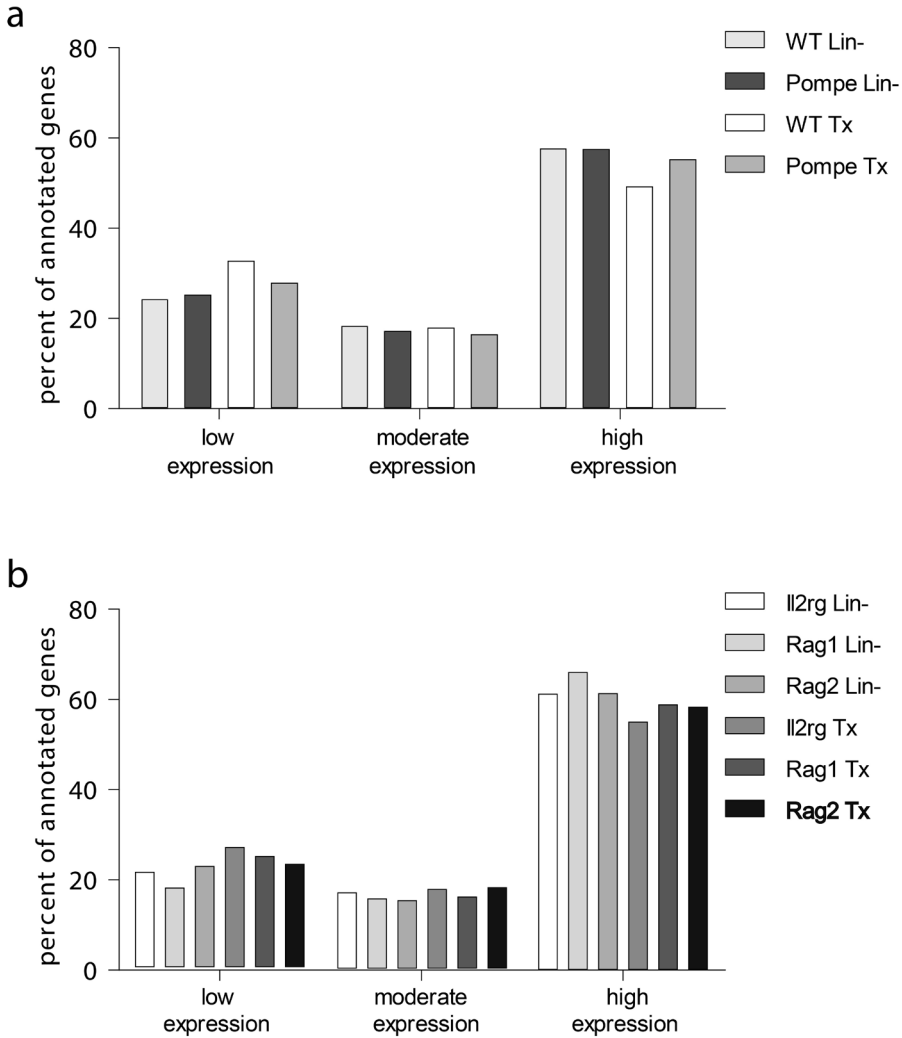


Figure 3: Gene expression binning data. Gene expression values of lineage negative Sca-1⁺ c-Kit⁺ cells after two days of stimulation, as measured on an Affymetix mouse 430 2.0 array, were sorted into 10 equally-sized bins based on their relative expression levels. Subsequently, the MAVRIC annotation tool uses genes nearest to an integration site in one of the lineage negative (Lin-) or transplanted (Tx) datasets to generate a histogram based on this binning profile. Bins 1-4, 5-6 and 7-10 were grouped into three categories dubbed low expression, moderate expression and high expression, respectively. Graphing these categories for Pompe and wild type (a) or SCID (b) integrations gives an overview of the relation between relative gene expression level and integration frequency. Highly expressed genes are much more likely to contain nearby integration sites.

Comparison of integration profiles: GFP vs IL2RG or GAA

In order to determine if insertion of a therapeutic transgene could result in a selective advantage and IS bias after engraftment, the integration profile of samples in *Il2rg*^{-/-} mice transplanted with *Il2rg*^{-/-} Lin⁻ cells transduced with an *IL2RG* vector was compared with the profile of mice transplanted after transduction with a GFP vector (Table 5). The number of integration sites for both subgroups was similar (1506 IS for *IL2RG* vectors vs 1567 for GFP vectors), allowing direct comparison of the integration site profiles. No significant differences in the frequency of IS near genes, TSS or oncogenes were seen between the *GFP* and *IL2RG* datasets. Noteworthy CIS oncogenes included *Mef2c* for the *IL2RG* vector dataset (two 2nd order CIS 112kb distant). Ingenuity Pathways Analysis revealed an overrepresentation of genes related to functional categories cancer, organismal survival and gene expression for both datasets as well as the *GAA* data (Figure 4). The *IL2RG* data had an increased frequency of IS near cancer-related genes, although this difference was not significant. The same *GFP* vs *IL2RG* vector analysis was performed using the Lin⁻ IS, and the same overrepresentation of genes linked to the cancer pathway was observed (data not shown) but no significant differences between the two transgenes were noted.

Table5: *in vivo* integrations from GFP vectors vs therapeutic vectors

	therapeutic		GFP	
	Il2rg	Pompe	Il2rg	Pompe
no. of unique IS	1506	376	1567	424
no. of unique genes	1071	350	1099	388
no. of IS within 10kb of TSS	200	70	245	91
no. of IS within 10kb of gene	996	253	1049	304
no. of unique oncogenes	33	12	39	12
total no. of oncogenes	51	14	58	13
no. of CIS	187	17	206	25
CIS oncogenes	5	2	7	1
% within 10kb of gene	66.1%	67.3%	66.9%	71.7%
% of IS that are CIS	12.4%	4.5%	13.1%	5.9%
% of IS that are oncogenes	3.4%	3.7%	3.7%	3.1%

This analysis was repeated for the Pompe IS data. The number of *in vivo* integrations available for analysis was much lower than in the *IL2RG* data (376 unique IS for Pompe with *GAA* vectors and 424 unique IS for Pompe GFP vectors), so higher variance in the data is expected. The GFP vector IS had a higher percentage of IS that were nearby a gene region and a higher percentage of CIS, but a lower oncogene frequency (3.1% vs 3.7% for the *in vivo* integrations). No Pompe Lin⁻ samples transduced with GFP were available, so the *in vitro* samples were not used for comparative analysis.

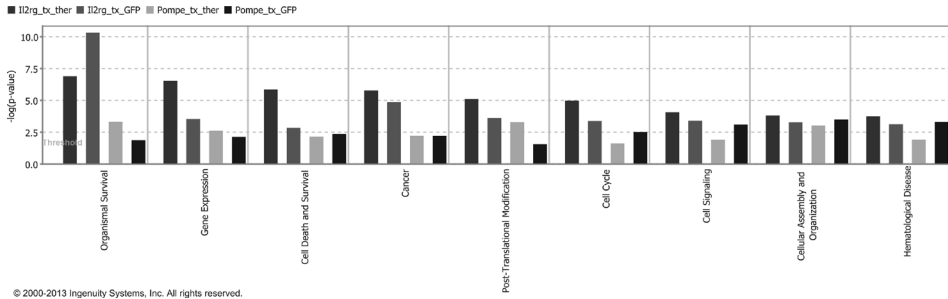


Figure 4: Top Ingenuity Functional Pathways linked to *in vivo* samples containing IL2G, GAA or GFP transgenes. Top Ingenuity functional pathways for genes flanking integrations in IL2rg^{-/-} or Pompe mice transduced with a therapeutic vector (ther) or a GFP control vector. Functional pathways are sorted based on their ranking in the IL2rg therapeutic dataset.

DISCUSSION

Lentiviral vectors are a viable alternative to gammaretroviral vectors for clinical gene therapy trials, offering a distinct integration pattern which favors integrations in or near active genes rather than near gene transcription start sites. While many different groups are investigating lentiviral gene therapy vectors for treatment of various diseases [7, 10, 26-29], assessing lentiviral integration patterns is complicated by systematic differences in the methodology of animal experiments in various labs. Vector design, transduction protocols, and differences in integration site annotation methods all complicate direct comparison of lentiviral integration profiles derived from different sources. In order to enable large-scale comparisons of lentiviral integration profiles under a wide range of physiological settings, the total IS annotations from a number of *in vitro* and *in vivo* experiments conducted in the same lab were pooled. Over 16,000 lentiviral integration sites have been annotated and analyzed, including murine IS collected *in vitro*, *in vivo*, from SCID, lysosomal storage disorder and wild type mouse phenotypes, and from human and rhesus CD34⁺ cells. By stratifying the pooled IS based on various parameters, some of the selective forces which act on gene therapy-treated HSCs after transplantation and engraftment in the bone marrow are revealed.

Comparison of LV integration sites generated *in vitro* in mouse, rhesus and human hematopoietic stem and progenitor cells confirms that the vector's integration preferences are not species specific. No significant differences were found between the species when comparing the frequency of IS near genes, gene TSS, or oncogenes. The three species also had a similar profile of overrepresented Ingenuity functional pathways, although the mouse HSPC having a prominent overrepresentation of genes involved in cancer compared to the rhesus and human datasets.

It has been hypothesized that viral integrations which improve the relative fitness of the transduced cell in the competitive environment of the bone marrow niche or thymus will be over-represented relative to those integrations which do not grant a competitive advantage. In this case, integration sites near proto-oncogenes or genes involved in cell proliferation, and which can influence the regulation of these genes, would be more likely to be found *in vivo*. However, the integration site profiles observed in this analysis suggest that this is not the case. Comparing (pre-transplant) Lin⁻ samples and samples collected from bone marrow, spleen, peripheral blood and lymph nodes post-transplant revealed that the frequency of an IS landing near an oncogene in pre-transplant samples was similar or lower than in post-transplant samples for the mouse SCID phenotypes studied here, while higher in transplanted samples than in Lin⁻ cells for the Pompe and wild type groups. At the same time, Ingenuity Pathways Analysis revealed a highly significant overrepresentation of genes related to cancer in the Lin⁻ samples for wild type and Pompe mice, while the transplanted samples had far fewer genes linked to cancer. Ingenuity analysis applied to SCID mice revealed a significant overrepresentation of genes linked to cancer in both

the Lin⁻ and transplanted samples, with very little difference between the two groups. Comparing the IS profiles of therapeutic vectors to GFP control vectors revealed no evidence of integration site skewing, suggesting that the transgene has little or no effect on the LV integration profile.

One feature of LV integration site profiles observed in our analysis was that the relative frequency of IS within 10kb of a gene was lower in transplanted cells than in Lin⁻ cells. A similar effect was seen when comparing the Lin⁻ cells of wild type and SCID animals, with the wild type Lin⁻ cells having significantly fewer integrations near or in Refseq genes. These findings suggests that there may be some sort of negative selection against cells containing integration sites near cancer or proliferative genes taking place in the bone marrow, and that this effect may be reduced in the immunodeficient SCID mice. Alternatively, the reduction in IS near or in genes could be the result of some of the integrations disrupting gene function and reducing the viability of the transduced cell, an effect which would not be seen in Lin⁻ cells cultured for only a week.

The prevalence of common integration sites (CIS) varied based on the size of the dataset analyzed, but in all cases was much higher than the expected frequency of CIS in datasets of comparable size based on mathematical models of random integration or on *in silico* data [30]. This further emphasizes that lentiviral integration patterns are only semi-random.

Prominent hot spots containing multiple CIS of high order (5th order or greater) were located in megabase ranges on most chromosomes, with chromosomes 1, 6, 8, and 11 having three or more CIS clusters within 20 megabases. Out of 1639 CIS identified by pooling all of the transplanted murine samples, only 57 were in or near known proto-oncogenes, for a frequency of 3.4%. This is very similar to the frequency of non-CIS oncogenes in the data as a whole, suggesting that no clustering or positive selection for IS near oncogenes takes place.

Overall, these data suggest that that lentiviral integration profiles are stable across a wide range of phenotypes, and that LV vector hematopoietic stem cell gene therapy has a relatively low risk of integrating near oncogenes.

MATERIALS AND METHODS

Mice

Wild-type BALB/c- C57BL-6, and C57BL-6-*Rag1*^{-/-}, BALB/c-*Rag2*^{-/-}[21], BALB/c-*Il2rg*^{-/-}[20] and FVB/N-*Gaa*^{-/-}[22, 31] mice were bred in the Experimental Animal Center of Erasmus MC. Congenic *Rag1*^{-/-} mice were derived from a C57BL-6 background as previously described[32]. All experiments were approved by the institutional Animal Ethical Committee of Erasmus MC in accordance with legislation in the Netherlands.

Lentiviral vectors

The third generation self-inactivating (SIN) lentiviral vector (LV) pRRL.PPT.SF.GFP.bPRE4*.SIN[24, 25] and derivatives were constructed using the HIV-1 vector backbone as previously described[20-22].

Therapeutic promoter cassettes were divided in ubiquitous promoters (UP) and cellular restricted promoters (CP). UP included the house keeping promoter human elongation factor 1 α short (EFS)[33], phosphoglycerate kinase (PGK)[34], and a 2.6 kb ubiquitous chromatin opening element (UCOE)[35, 36]. The CP used were the native human RAG1 or RAG2 promoters (RAG1p and RAG2p, respectively)[7], the TCR V β 6.7 gene promoter (TCRV β p)[10], and the γ chain promoter (γ Pr)[37]. The CP promoters were made by PCR amplification of human genomic DNA; excepting the RAG2p which was generated from the pGL2-RAG2-LUC plasmid, ligated into pcR-TOPO (Invitrogen) and verified by sequencing (primer sequences provided in table E1 in the Online Repository).

The following lentiviral vector constructs were used: *SF.IL2RG*, *SF.IL2RGco*, *PGK.IL2RGco*, *γ Pr.IL2RGco*, *EFS.IL2RGco*, *SF.GFP*, *PGK.GFP* and *γ Pr.GFP* for *Il2rg*^{-/-}; *SF.RAG1co*, *TCRBpr.RAG1co*, *γ Pr.RAG1co*, *RAG1pr.RAG1co* and *SF.GFP* for *Rag1*^{-/-}; *SF.RAG2*, *SF.RAG2co*, and *UCOE.RAG2co* for *Rag2*^{-/-}; *SF.GAAco*, *PGK.GAAco* and *SF.GFP* for *Gaa*^{-/-}; and *SF.GFP* for *in vitro* transduction of murine, rhesus and human HSPCs and *in vivo* studies in wild type BALB/c mice. Nomenclature “co” refers to codon optimization of transgene cDNA to improve transcription and translation[28], performed by GeneOptimizer[®] software (GeneArt AG, Regensburg, Germany).

Transduction and transplantation of lineage negative BM cells

BM cells from male mice were purified by lineage depletion (Lin-) (BD Biosciences, Santa Clara, CA). Lin- BM cells were transduced overnight at various cell densities in the presence of murine stem cell factor (mSCF) 100 ng/mL, human FMS-like tyrosine kinase 3-ligand (hFlt3-L) 50ng/ml and murine thrombopoietin (mTPO) 10 ng/ml. Vector dose varied based on the experiment, with a range of HeLa MOI 20 to 1. Subsequently, 5 \times 10⁵ to 1 \times 10³ Lin- cells were injected into the tail vein of 6 Gy or 2 Gy irradiated female *Il2rg*^{-/-} recipients or without conditioning. Mice were followed for six to twelve months and subsequently DNA was purified from peripheral blood, bone marrow, spleen, thymus and lymph nodes.

LAM-PCR and integration site analysis workflow

High-resolution insertion-site analysis by linear amplification–mediated PCR (LAM-PCR) of vector-genome border regions was performed as previously described [23]. Restriction enzymes Tsp509I or MseI were used with the lentiviral (HIV) primer set. LAM-PCR products were pooled and sequenced at GATC Biotech (Konstanz, Germany) and subsequently annotated using the MAVRIC integration site analysis tool [38]. Briefly, viral sequences were trimmed and the remaining genomic sequences aligned to the appropriate genome via BLAST (maximum e-value = 0.01, integrations must have one unique best e-value score). The genomic regions around aligned sequences were scanned using the Ensembl genome browser (maximum distance = 100kb). Each IS was annotated with a single, nearest Refseq gene[39], which was subsequently used for downstream analysis. Genes nearest to an integration site were flagged as proto-oncogenes based on a list of 528 mouse common integration site genes from the Retrovirus and transposon Tagged Cancer Gene Database (RTC GD). When pooling IS for large scale analysis, integrations within 10 bases of another integration site, on either end, were considered redundant and removed from the analysis. This step was included to prevent instances of an identical integration site derived repeatedly from multiple tissues in a single mouse from skewing the analysis. Common integration sites (CIS) were determined based on established methods[40].

Statistics

P values for comparisons of the frequency of oncogenes, IS near TSS, and IS within 10kb of a gene were calculated using Fisher's Exact test (2-sided), with Bonferroni's correction used to adjust the threshold p value (α) where applicable. P-values for overrepresented functional pathways calculated automatically using Ingenuity's software (Fisher's Exact test) and adjusted for multiple comparison analysis using Bonferroni's correction (calculated on the basis of 75 functional groups for each analysis).

ACKNOWLEDGEMENTS

The authors acknowledge Drs. F.J.T. Staal and K. Pike-Overzet, Dept. Immunology, Erasmus University Medical Center, for assistance in the microarray analyses.

REFERENCES

1. Gaspar, H.B., et al., *Long-term persistence of a polyclonal T cell repertoire after gene therapy for X-linked severe combined immunodeficiency*. *Sci Transl Med*, 2011. **3**(97): p. 97ra79.
2. Hacein-Bey-Abina, S., et al., *Efficacy of gene therapy for X-linked severe combined immunodeficiency*. *N Engl J Med*, 2010. **363**(4): p. 355-64.
3. Aiuti, A., et al., *Gene therapy for immunodeficiency due to adenosine deaminase deficiency*. *N Engl J Med*, 2009. **360**(5): p. 447-58.
4. Gaspar, H.B., et al., *Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction*. *Sci Transl Med*, 2011. **3**(97): p. 97ra80.
5. Ott, M.G., et al., *Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVII, PRDM16 or SETBP1*. *Nat Med*, 2006. **12**(4): p. 401-9.
6. Boztug, K., et al., *Stem-cell gene therapy for the Wiskott-Aldrich syndrome*. *N Engl J Med*, 2010. **363**(20): p. 1918-27.
7. Zarrin, A.A., et al., *Cloning and characterization of the human recombination activating gene 1 (RAG1) and RAG2 promoter regions*. *J Immunol*, 1997. **159**(9): p. 4382-94.
8. Cavazzana-Calvo, M., et al., *Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia*. *Nature*, 2010. **467**(7313): p. 318-22.
9. Cartier, N., et al., *Lentiviral hematopoietic cell gene therapy for X-linked adrenoleukodystrophy*. *Methods Enzymol*, 2012. **507**: p. 187-98.
10. Deng, X., et al., *Characterization of human TCR Vbeta gene promoter. Role of the dodecamer motif in promoter activity*. *J Biol Chem*, 1998. **273**(37): p. 23709-15.
11. Hacein-Bey-Abina, S., et al., *Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1*. *J Clin Invest*, 2008. **118**(9): p. 3132-42.
12. Howe, S.J., et al., *Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients*. *J Clin Invest*, 2008. **118**(9): p. 3143-50.
13. Stein, S., et al., *Genomic instability and myelodysplasia with monosomy 7 consequent to EVII activation after gene therapy for chronic granulomatous disease*. *Nat Med*, 2010. **16**(2): p. 198-204.
14. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration*. *Nat Biotechnol*, 2006. **24**(6): p. 687-96.
15. Montini, E., et al., *The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy*. *J Clin Invest*, 2009. **119**(4): p. 964-75.
16. Modlich, U., et al., *Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity*. *Blood*, 2006. **108**(8): p. 2545-53.
17. Woods, N.B., et al., *Gene therapy: therapeutic gene causing lymphoma*. *Nature*, 2006. **440**(7088): p. 1123.
18. Schwarzwaelder, K., et al., *Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo*. *J Clin Invest*, 2007. **117**(8): p. 2241-9.
19. Schroder, A.R., et al., *HIV-1 integration in the human genome favors active genes and local hotspots*. *Cell*, 2002. **110**(4): p. 521-9.
20. Huston, M.W., et al., *Correction of murine SCID-X1 by lentiviral gene therapy using a codon-optimized IL2RG gene and minimal pretransplant conditioning*. *Mol Ther*, 2011. **19**(10): p. 1867-77.
21. van Til, N.P., et al., *Correction of murine Rag2 severe combined immunodeficiency by lentiviral gene therapy using a codon-optimized RAG2 therapeutic transgene*. *Mol Ther*, 2012. **20**(10): p. 1968-80.
22. van Til, N.P., et al., *Lentiviral gene therapy of murine hematopoietic stem cells ameliorates the Pompe disease phenotype*. *Blood*, 2010. **115**(26): p. 5329-37.
23. Schmidt, M., et al., *High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR)*. *Nat Methods*, 2007. **4**(12): p. 1051-7.

24. Follenzi, A. and L. Naldini, *Generation of HIV-1 derived lentiviral vectors*. *Methods Enzymol*, 2002. **346**: p. 454-65.
25. Follenzi, A. and L. Naldini, *HIV-based vectors. Preparation and use*. *Methods Mol Med*, 2002. **69**: p. 259-74.
26. Yannaki, E. and G. Stamatoyannopoulos, *Hematopoietic stem cell mobilization strategies for gene therapy of beta thalassemia and sickle cell disease*. *Ann N Y Acad Sci*, 2010. **1202**: p. 59-63.
27. Tschernutter, M., et al., *Long-term preservation of retinal function in the RCS rat model of retinitis pigmentosa following lentivirus-mediated gene therapy*. *Gene Ther*, 2005. **12**(8): p. 694-701.
28. Pike-Overzet, K., et al., *Correction of murine Rag1 deficiency by self-inactivating lentiviral vector-mediated gene transfer*. *Leukemia*, 2011. **25**(9): p. 1471-83.
29. Shi, Q., et al., *Lentivirus-mediated platelet-derived factor VIII gene therapy in murine haemophilia A*. *J Thromb Haemost*, 2007. **5**(2): p. 352-61.
30. Candotti, F., et al., *Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans*. *Blood*, 2012. **120**(18): p. 3635-46.
31. Bijvoet, A.G., et al., *Generalized glycogen storage and cardiomegaly in a knockout mouse model of Pompe disease*. *Hum Mol Genet*, 1998. **7**(1): p. 53-62.
32. Mombaerts, P., et al., *RAG-1-deficient mice have no mature B and T lymphocytes*. *Cell*, 1992. **68**(5): p. 869-77.
33. Thornhill, S.I., et al., *Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency*. *Mol Ther*, 2008. **16**(3): p. 590-8.
34. van Til, N.P., et al., *Kupffer cells and not liver sinusoidal endothelial cells prevent lentiviral transduction of hepatocytes*. *Mol Ther*, 2005. **11**(1): p. 26-34.
35. Zhang, F., et al., *A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors*. *Mol Ther*, 2010. **18**(9): p. 1640-9.
36. Zhang, F., et al., *Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells*. *Blood*, 2007. **110**(5): p. 1448-57.
37. Markiewicz, S., et al., *Tissue-specific activity of the gammac chain gene promoter depends upon an Ets binding site and is regulated by GA-binding protein*. *J Biol Chem*, 1996. **271**(25): p. 14849-55.
38. Huston, M.W., et al., *Comprehensive Investigation of Parameter Choice in Viral Integration Site Analysis and Its Effects on the Gene Annotations Produced*. *Hum Gene Ther*, 2012.
39. Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. *Nucleic Acids Res*, 2007. **35**(Database issue): p. D61-5.
40. Deichmann, A., et al., *Insertion sites in engrafted cells cluster within a limited repertoire of genomic areas after gammaretroviral vector gene therapy*. *Mol Ther*, 2011. **19**(11): p. 2031-9.

Supplemental Table 1: CIS of order 5+ for all *in vitro* and *in vivo* datasets

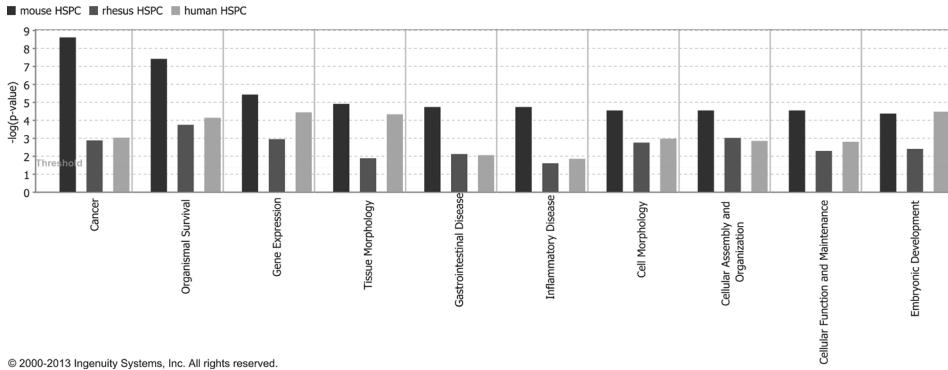
CIS order	nearest gene(s)	locus (chrom:bp)	dataset
5th	B3galt2;Cdc73;Glrx2;Uchl5	1:145500259-145632404	WT mouse <i>in vitro</i>
5th	Akt3	1:179046612-179183780	WT mouse <i>in vitro</i>
5th	2700029M09Rik;Clcn3	8:63392605-63453218	WT mouse <i>in vitro</i>
5th	LOC722601	2:1954335-2135540	rhesus <i>in vitro</i>
7th	BX572640.9;Sfi1;Eif4enif1	11:3030445-3110058	Il2rg <i>in vitro</i>
5th	Smc6;Vsnl1	12:11276072-11390057	Il2rg <i>in vitro</i>
8th	Plag1	4:3858379-3862732	WT <i>in vivo</i>
6th	Trip12	1:84725207-84766670	WT <i>in vivo</i>
6th	Uba2	7:34930305-35091597	WT <i>in vivo</i>
6th	Vegfc	8:55188886-55262672	WT <i>in vivo</i>
6th	Ube2e2	14:19550698-19721219	WT <i>in vivo</i>
5th	Nqo2;Ripk1;Bphl	13:34082481-34155957	WT <i>in vivo</i>
5th	Pkhd11	15:44325577-44393622	WT <i>in vivo</i>
5th	Fshr	17:89494826-89573458	WT <i>in vivo</i>
5th	Ap3s1	18:46919713-46942581	WT <i>in vivo</i>
9th	Ngly1;Oxsm;Top2B	14:17081150-17239702	Rag2 <i>in vivo</i>
8th	Sfi1;BX572640.9;Eif4enif1	11:3031603-3110289	Rag2 <i>in vivo</i>
6th	Mospd2	X:161374599-161414861	Rag2 <i>in vivo</i>
6th	4732440D04Rik	1:6206629-6244805	Rag2 <i>in vivo</i>
6th	Vegfc	8:55199088-55366242	Rag2 <i>in vivo</i>
6th	Dmxl1;Tnfaip8	18:50000245-50123411	Rag2 <i>in vivo</i>
6th	4930503L19Rik;Stard6;Poli;Mbd	18:70626903-70740285	Rag2 <i>in vivo</i>
6th	Rgs18	1:146599298-146617984	Rag2 <i>in vivo</i>
5th	Gtdc1	2:44703455-44758157	Rag2 <i>in vivo</i>
5th	Slit2;Pacrgl	5:48593553-48710476	Rag2 <i>in vivo</i>
5th	Glcci1	6:8226905-8359267	Rag2 <i>in vivo</i>
5th	Itp1	6:108189526-108250583	Rag2 <i>in vivo</i>
5th	Mir706;Wnk1	6:119958946-119968079	Rag2 <i>in vivo</i>
5th	1700017N19Rik	10:100157015-100167934	Rag2 <i>in vivo</i>
5th	4933413L06Rik;Hen1	13:118601875-118645154	Rag2 <i>in vivo</i>
5th	Nsun3;Stx19;Arl13b	16:62763719-62913276	Rag2 <i>in vivo</i>
5th	AC139157.4;Mpp7	18:7571900-7670499	Rag2 <i>in vivo</i>
5th	5730494M16Rik;AW554918	18:25298264-25453383	Rag2 <i>in vivo</i>
9th	Serbp1	6:67219389-67241340	Rag1 <i>in vivo</i>
7th	Prkg1	19:31604719-31613286	Rag1 <i>in vivo</i>
6th	Cd9	6:125433313-125433956	Rag1 <i>in vivo</i>

CIS order	nearest gene(s)	locus (chrom:bp)	dataset
6th	Kcnmb4;Cnot2	10:115927351-115996941	Rag1 <i>in vivo</i>
5th	Lingo2	4:36315548-36315795	Rag1 <i>in vivo</i>
5th	Mup5	4:61493189-61493599	Rag1 <i>in vivo</i>
5th	Trpc6	9:8548164-8548537	Rag1 <i>in vivo</i>
5th	Pja2	17:64672665-64673226	Rag1 <i>in vivo</i>
11th	Cenpc1;Stap1;Uba6	5:86459481-86582241	Il2rg <i>in vivo</i>
11th	Hsd3b7;Zfp668	7:134941128-135003319	Il2rg <i>in vivo</i>
10th	Adamts3	5:90215910-90284089	Il2rg <i>in vivo</i>
9th	Psme2b-ps;Olf1r56	11:48758858-48940405	Il2rg <i>in vivo</i>
9th	Meox2	12:37821771-37822579	Il2rg <i>in vivo</i>
8th	Fezf1;Cadps2	6:23232110-23268216	Il2rg <i>in vivo</i>
8th	4922501C03Rik	9:87107746-87246456	Il2rg <i>in vivo</i>
8th	Polr2a	11:69550965-69551788	Il2rg <i>in vivo</i>
7th	Spata6	4:111457160-111458241	Il2rg <i>in vivo</i>
7th	Hp1bp3	4:137704040-137705056	Il2rg <i>in vivo</i>
7th	Sfi1;Eif4enif1;Patz1	11:3031770-3202635	Il2rg <i>in vivo</i>
7th	5730522E02Rik	11:25601161-25602362	Il2rg <i>in vivo</i>
7th	Smad7	18:75574839-75575388	Il2rg <i>in vivo</i>
7th	Rfx3	19:27936866-28082262	Il2rg <i>in vivo</i>
6th	Klhl4	X:111641012-111641577	Il2rg <i>in vivo</i>
6th	Hisppd1;Pam	1:99749247-99908290	Il2rg <i>in vivo</i>
6th	Plxdc2	2:16422537-16565516	Il2rg <i>in vivo</i>
6th	Tbcd	11:121316541-121317930	Il2rg <i>in vivo</i>
6th	Coch;Strn3	12:52693480-52762846	Il2rg <i>in vivo</i>
6th	Mycbp2	14:103625972-103705779	Il2rg <i>in vivo</i>
6th	Csm3	15:48374807-48565562	Il2rg <i>in vivo</i>
6th	Rab11b	17:33891591-33892146	Il2rg <i>in vivo</i>
6th	Wac	18:7887637-7908433	Il2rg <i>in vivo</i>
5th	Plac1	X:50535886-50536046	Il2rg <i>in vivo</i>
5th	Sdccag8	1:178831743-178888431	Il2rg <i>in vivo</i>
5th	Rnpc3	3:113318930-113329219	Il2rg <i>in vivo</i>
5th	Iffo2;Ubr4	4:139002575-139003340	Il2rg <i>in vivo</i>
5th	Hgf	5:16083233-16147351	Il2rg <i>in vivo</i>
5th	Phtf2;Rsb1l1	5:20378460-20419684	Il2rg <i>in vivo</i>
5th	Pcm1	8:42346466-42421556	Il2rg <i>in vivo</i>
5th	Rbm6;Ip6k1	9:107737849-107928163	Il2rg <i>in vivo</i>
5th	Mnat1	12:74265405-74266301	Il2rg <i>in vivo</i>

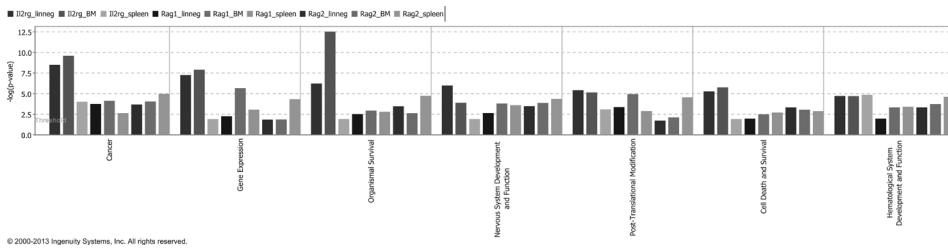
Supplemental Table 2: CIS oncogenes for all *in vitro* and *in vivo* datasets

CIS order	oncogene	locus (chrom:bp)	dataset
2nd	AHI1	6:135747276-135754727	human <i>in vitro</i>
2nd	RAP1GDS1	4:99302155-99320768	human <i>in vitro</i>
2nd	RREB1	6:7154873-7156445	human <i>in vitro</i>
2nd	Akap12	7:82778136-82791917	WT mouse <i>in vitro</i>
3rd	Mef2c	13:83673980-83688671	WT mouse <i>in vitro</i>
2nd	5730419I09Rik	6:143015786-143028748	Il2rg <i>in vitro</i>
2nd	Il2rg	X:98463949-98464344	Il2rg <i>in vitro</i>
2nd	Map3k8	18:4341613-4368948	Il2rg <i>in vitro</i>
2nd	Mef2c	13:83674833-83679611	Il2rg <i>in vitro</i>
2nd	Mid1	X:166412406-166414614	Il2rg <i>in vitro</i>
2nd	Rere	4:149745522-149771559	Il2rg <i>in vitro</i>
2nd	Rnf216	5:143825346-143841534	Il2rg <i>in vitro</i>
2nd	Runx2	17:44822168-44846482	Il2rg <i>in vitro</i>
2nd	Slc19a1	10:76503383-76503387	Il2rg <i>in vitro</i>
2nd	Zdhhc7	8:122600710-122600719	Rag1 <i>in vitro</i>
2nd	Bcl11a	11:23990060-23990061	WT <i>in vivo</i>
2nd	Camk2d	3:126374077-126385393	WT <i>in vivo</i>
2nd	Cr2	1:196978154-196978155	WT <i>in vivo</i>
2nd	Evi1	3:30013260-30033466	WT <i>in vivo</i>
2nd	Gtf2a1	12:92818318-92824334	WT <i>in vivo</i>
2nd	Lnpep	17:17680467-17684517	WT <i>in vivo</i>
4th	Lnpep	17:17733747-17764326	WT <i>in vivo</i>
8th	Plagl1	4:3858379-3862732	WT <i>in vivo</i>
2nd	Ppfibp1	6:146848693-146849124	WT <i>in vivo</i>
2nd	Rap1gds1	3:138661473-138662870	WT <i>in vivo</i>
2nd	Tnfrsf8	4:144959073-144959074	WT <i>in vivo</i>
2nd	Cr2	1:197025311-197025462	Pompe <i>in vivo</i>
2nd	Eps15l1	8:74919901-74919904	Pompe <i>in vivo</i>
2nd	Evi1	3:29951810-29976592	Pompe <i>in vivo</i>
2nd	Zfp521	18:13872964-13873209	Pompe <i>in vivo</i>
3rd	Chd2	7:80689961-80698119	Rag2 <i>in vivo</i>
2nd	Dhrs3	4:144479703-144480327	Rag2 <i>in vivo</i>
2nd	Eml4	17:83779316-83789841	Rag2 <i>in vivo</i>
2nd	Fli1	9:32333271-32333366	Rag2 <i>in vivo</i>

CIS order	oncogene	locus (chrom:bp)	dataset
2nd	Igf1r	7:75255470-75255471	Rag2 <i>in vivo</i>
2nd	Ikzf1	11:11604680-11618953	Rag2 <i>in vivo</i>
4th	Lnpep	17:17659083-17685294	Rag2 <i>in vivo</i>
2nd	Lrmp	6:145076139-145076149	Rag2 <i>in vivo</i>
3rd	Pde4d	13:109689194-109689256	Rag2 <i>in vivo</i>
2nd	Runx2	17:44956558-44956654	Rag2 <i>in vivo</i>
4th	Zeb2	2:44949600-44997523	Rag2 <i>in vivo</i>
4th	Zfp608	18:55072950-55123377	Rag2 <i>in vivo</i>
3rd	Arid1a	4:133273234-133273420	Rag1 <i>in vivo</i>
2nd	Bach2	4:32587042-32587094	Rag1 <i>in vivo</i>
3rd	Cdk6	5:3375852-3381855	Rag1 <i>in vivo</i>
3rd	Coro1a	7:133862044-133863162	Rag1 <i>in vivo</i>
4th	Gimap7	6:48659062-48679534	Rag1 <i>in vivo</i>
2nd	Ikzf1	11:11604706-11624778	Rag1 <i>in vivo</i>
3rd	Plag1	4:3820950-3860577	Rag1 <i>in vivo</i>
2nd	Rbm39	2:155999074-155999343	Rag1 <i>in vivo</i>
2nd	Zeb2	2:45085683-45085819	Rag1 <i>in vivo</i>
5th	Sdccag8	1:178831743-178888431	Il2rg <i>in vivo</i>
3rd	Gimap6	6:48657067-48680056	Il2rg <i>in vivo</i>
3rd	Jak1	4:100885754-100885930	Il2rg <i>in vivo</i>
3rd	Myb	10:20948239-20965158	Il2rg <i>in vivo</i>
3rd	Rspo3	10:29300524-29338034	Il2rg <i>in vivo</i>
2nd	Cdk6	5:3388366-3414078	Il2rg <i>in vivo</i>
2nd	Ghr	15:3424982-3425026	Il2rg <i>in vivo</i>
2nd	Ikzf1	11:11619194-11639027	Il2rg <i>in vivo</i>
2nd	Mef2c	13:83640034-83644074	Il2rg <i>in vivo</i>
2nd	Mef2c	13:83756876-83757178	Il2rg <i>in vivo</i>
2nd	Meis1	11:18802925-18831440	Il2rg <i>in vivo</i>
2nd	Pde4d	13:109588147-109589229	Il2rg <i>in vivo</i>
2nd	sep_09	11:117184972-117185475	Il2rg <i>in vivo</i>
2nd	Slc38a1	15:96453460-96453465	Il2rg <i>in vivo</i>
2nd	Smg6	11:74771873-74772308	Il2rg <i>in vivo</i>
2nd	Sox4	13:28882951-28893782	Il2rg <i>in vivo</i>
2nd	Stag1	9:100652120-100652147	Il2rg <i>in vivo</i>



Supplemental Figure 1: Top Ingenuity functional pathways for integrations recovered from cultured HSPC. Top functional pathways associated with genes flanking integration sites in cultured hematopoietic stem and progenitor cells (HSPC), as assessed by Ingenuity Pathways Analysis software (IPA). Functions sorted based on the order of significance in mouse HSPC.



Supplemental Figure 2: Top Ingenuity functional pathways recovered from SCID integrations. Top IPA functional pathways associated with genes flanking integration sites in *Il2rg*^{-/-}, *Rag1*^{-/-} and *Rag2*^{-/-} mice. Functions sorted based on the order of significance in lineage negative (lin neg) *Il2rg*^{-/-} cells dataset.

Met Tyr Xle Ala Cys
 Ala Tyr Xle Ala Cys
 Cys Xle Ala Cys Asn Pyl Thr
 Ala Lys Ser Cys Pyl Thr
 Xle Ser Pyl Lys SecSer
 PylAsn Arg Glu Asn Sar
 Ser Leu SerThr Pyl LeuLeulle IleAsaAsn His
 Lys Leu Glu LeuGlu Asn TrpIle LeuLeu His Sec
 Cys Glu Ser GGAC GCC GGCATGCC
 Ala Xle Thr AGCAGGTACGCGAGA His Asn Ser
 GluArg Thr AGGCTAGCTAG PylThr
 CAGGCTCGA Arg His
 CAGTGGGTAT PylAsn Arg AlaArg ArgIle
 GTGAGCTTGCTC Thr Ser Ile Asn Pyl Glu
 GATAGCGGGTA Sec Thr Pyl Thr His
 TGCTAGCCCGCC His Thr His
 GGGAAATGCAGG SerSec
 TACGCTTGCTAGA
 CCGCAGGCATGC
 CTGTATGCTTGCTC
 01001101011000010111
 00100 1110011011010000
 01100 001011011000110
 1100 0010000001010111
 011 010010110110001101
 10 0011010010110000101
 101101001 00000010010
 00011101 01011100110111010
 0011011 11011011100010110000
 100000 01010
 00001 1010
 000 100100

General Discussion

EFFICACY OF LENTIVIRAL GENE THERAPY FOR SCID-X1

The ideal candidate disease for gene therapy intervention would have the following characteristics: a chronic condition, caused by mutations in a single gene, with the deleterious effects localized to a single tissue which is easy for physicians to access. Several inherited immunodeficiency diseases fit all of these criteria, so it was natural that the first gene therapy trials would be conducted on SCID patients. The SCID-X1 and ADA-SCID trials initiated in 2000 featured MLV-based gammaretroviral vectors driven by the viral long terminal repeat (LTR) [1-3], which had strong expression of the therapeutic *IL2RG* and *ADA* transgenes and had been reported as safe and effective in pre-clinical studies[4, 5]. These gene therapy trials demonstrated that gammaretroviral gene therapy was capable of curing SCID with a single administration, but also revealed that the treatment could have severe adverse effects that were not foreseen in the pre-clinical animal models. The onset of oncogenesis in 5 of the 20 SCID-X1 gene therapy patients was a serious setback for the field and overshadowed the successful immune system reconstitution seen in most of the treated SCID-X1 and ADA-SCID patients[6, 7]. These leukemic events led to a flurry of scientific research focused on determining why they occurred and how to improve vector safety for future trials. This thesis contains research related to these goals. To address the risk of gammaretroviral vector-mediated insertional mutagenesis in gene therapy of SCID-X1, we generated self-inactivating SIN HIV-1 derived lentiviral vectors. Lentiviral vectors have several properties which make them attractive gene therapy vectors. They are able to integrate with high efficiency into hematopoietic stem cells following overnight transduction and without pre-stimulation with growth factors. They also have a reduced likelihood of integrating near transcription start sites compared to gammaretroviruses and improved resistance to silencing. Lentiviruses have no preference to integrate into the promoter regions of active genes or into genes involved in cell growth and cancer[8].

The initial SCID-X1 clinical trials used the viral long terminal repeat to control *IL2RG* transgene expression. This LTR was later implicated as potential risk factor for the type of oncogenic events that occurred in the London and Paris trials. The leukemia cases that occurred in these trials were caused by integrations in vulnerable regions near proto-oncogenes resulting in subsequent upregulation and constitutive expression[9]. Other strong, constitutively active viral promoters, such as the SF promoter used in the X-linked chronic granulomatous disease (CGD) trial[10], have been shown to modify gene expression in regions greater than 100 kb, resulting in undesired upregulation of proto-oncogenes. However, it should be noted that the same SF promoter was used in the recent London ADA-SCID trial, and to date no adverse events have been seen (48 to 72 months after treatment)[11]. Thus it is likely that the disease background plays an important role in the relative vulnerability to vector-related leukemic events. Nevertheless, the use of strong retroviral enhancer/promoter sequences is a known risk factor for triggering

adverse events and is not desirable for clinical gene therapy applications. Another potential drawback of using viral promoters is that many of these promoter elements are at risk of being silenced by methylation in mammalian cells.[12-14]

The lentiviral vectors described in this thesis made use of eukaryotic promoters, including the phosphoglycerate kinase promoter (PGK) or a 1.1kb section of the native *IL2RG* promoter region (γ Pr) to drive transgene expression. These promoters lead to a weaker level expression which is sufficient for correction of the SCID phenotype but should reduce the risks of unintended upregulation of genes flanking a vector integration site. Most eukaryotic promoters are less prone to methylation compared to viral promoters and so can be expected to retain their function long-term. Codon optimization of the *IL2RG* gene (dubbed *coIL2RG*) proved essential for improved viral vector titers, transgene transcript levels and protein production, allowing immune reconstitution when using eukaryotic promoters with modest expression levels such as those listed above.

Transduction of lineage negative (Lin-) *Il2rg*^{-/-} bone marrow cells with these lentiviral vectors and subsequent transplantation into *Il2rg*^{-/-} recipient mice led to a reversal of the SCID phenotype. T cell numbers, polyclonal TCR β gene rearrangements, concanavalin A/IL-2 stimulated T cell proliferation, T cell dependent and independent B cell antibody responses and levels of natural antibodies all improved over the course of six months as the transduced HSCs reconstituted the immune system. The *coIL2RG* vectors proved effective at restoring T cell populations and immune system functionality to wild type levels even at a low vector MOI, provided a minimum level of pre-transplant conditioning was applied. A series of GFP vectors demonstrated improved selectivity of the γ Pr promoter for expression in white blood cells. Overall, these vectors performed similarly to published results of gammaretroviral[15] and lentiviral[16] gene therapy for murine SCID-X1. The γ Pr-*coIL2RG* and PGK-*coIL2RG* vectors fully restored T lymphocyte populations in the periphery, BM and spleen, but were less effective in restoring B cell populations. This reduction in circulating B cell populations did not seem to affect plasma levels of IgM and IgG1, and the mice had fully normalized antibody responses to Pneumococcal polysaccharides and tetanus toxoid.

Titration of the viral vectors, particularly the LV- γ Pr-*coIL2RG* vector, demonstrated convincingly that lowering the vector dose to reach the preferred 1 copy per cell for clinical application had no significant effect on the reconstitution of T and B cell lineages. Likewise, reduction of the Lin- cell number transplanted to 10⁵/mouse (4x10⁶/kg, equivalent to 4x10⁷/kg unfractionated BM cells or roughly 10⁶/kg CD34⁺ cells in humans) did not noticeably influence immune reconstitution.

The importance of recipient BM conditioning for lentiviral gene therapy protocols

Reduction of the pre-transplant conditioning of recipient mice from 6 Gy to 2 Gy had no significant effect on the efficacy of the gene therapy treatment. However, eliminating

the pre-transplant conditioning altogether severely reduced the efficacy of the treatment in mice transplanted with either lentiviral vector treated or wild type cells, with a corresponding drop in donor cell BM chimerism, percentage of lymphocytes present in spleen and BM, and failures to produce antibodies upon immunization. This effect of reduced lymphocyte reconstitution was consistent across multiple experiments and positively correlated with a reduced cell dose (less than 10^5 Lin⁻ cells per mouse). Based on these observations, it is likely that applying mild conditioning to human SCID-X1 gene therapy protocols could result in improved immune reconstitution, particularly in when a low number of transduced cells is available. The critical role of cytoreductive conditioning has already been confirmed in murine and clinical studies of ADA-SCID gammaretroviral gene therapy [17, 18].

Unfortunately, the type of myeloablative and myelosuppressive conditioning typically used in the clinic prior to allogeneic bone marrow transplantation is associated with significant morbidity and mortality [19]. By the time they are approved for gene therapy intervention, SCID-X1 patients are often in such poor health due to recurrent infections that even mild myelosuppressive regimens could be fatal [20]. Development of a less toxic way to improve donor cell engraftment may provide a safer way to treat these patients.

Granulocyte colony forming factor (G-CSF) is a cytokine which disrupts the interaction between SDF-1 in stromal cells and CXCR4 in granulocyte precursors, resulting in a myeloid hyperplasia in the bone marrow and release of quiescent HSCs into circulation [21, 22]. Repeated treatments of G-CSF over a period of 4-5 days results in the release of a considerable number of CD34⁺ cells into the peripheral blood, which can be harvested and subsequently administered into patients requiring bone marrow transplants [23]. For this reason, G-CSF is routinely used in the clinic to collect donor cells for bone marrow transplantation without resorting to more invasive procedures. Since engraftment of hematopoietic stem cells in the bone marrow is a competitive process, with donor HSC engraftment increasing based on the number of cells transplanted [24], G-CSF mobilization of recipient HSC has the potential to improve donor cell engraftment and immune system reconstitution.

A G-CSF mobilization protocol was tested in immunodeficient *Il2rg*^{-/-} mice prior to lentiviral gene therapy treatment and the efficacy compared to mild 2Gy irradiation conditioning. After seven months, T and B cell numbers in the spleen, BM and peripheral blood and T-cell dependent antibody responses to tetanus toxoid were similar for mice transplanted with *Il2rg*^{-/-} Lin⁻ cells after transduction with a *SF.coIL2RG* lentiviral vector following G-CSF mobilization, mice transplanted with wild type BALB/c Lin⁻ cells following G-CSF mobilization, and mice transplanted with *Il2rg*^{-/-} Lin⁻ cells transduced by *SF.coIL2RG* vector following 2 Gy irradiation. Repeating the G-CSF mobilization and lentiviral gene therapy experiment using the γ Pr.*coIL2RG* vector at minimal copies per cell resulted in modest levels of lymphocyte reconstitution, similar to the results seen when using this vector after 2 Gy irradiation. Donor cell chimerism levels in the BM and

spleen were lower in mice treated with G-CSF than in mice given 2Gy irradiation, an observation that was consistent regardless of the cell type or vector used. At the same time, levels of chimerism and lymphocyte reconstitution in mice treated with G-CSF were higher than in mice transplanted without any conditioning. The modest efficacy of G-CSF mobilization in these experiments is offset by the greatly reduced toxicity of G-CSF treatment compared to the myelosuppressive conditioning used in the clinical settings, suggesting that mobilization of recipient bone marrow may be a viable strategy for future SCID-X1 clinical trials.

Risk assessment of IL2RG gene therapy

After the leukemias were discovered in the initial SCID-X1 clinical trials, the *IL2RG* gene was implicated as potentially having a role in oncogenesis[25]. Further investigation in animal models has since revealed that *LMO2*, the principle oncogene in the clinical SCID-X1 trials, and *IL2RG* are often deregulated together in leukemias, and that there is strong selection for *IL2RG* deregulation in *LMO2*-initiated instances of T cell leukemia [26]. However, deregulation of these two genes is insufficient to trigger oncogenesis: leukemia onset requires other contributing mutations including overexpression of other oncogenes and downregulation of tumor suppressor genes. In no cases was activation or overexpression of the *IL2RG* alone sufficient for leukemogenesis. Over the course of five years and 11 long term *in vivo* repopulation assay experiments, a total of 239 *Il2rg*^{-/-} mice were transplanted with LV vector transduced Lin⁻ cells or untreated wild type Lin⁻ cells and followed for at least 6 months. An additional 118 mice were followed for at least eight months as secondary recipients after receiving total BM cells recovered from the mice transplanted with transduced Lin⁻ cells. Over the course of these experiments, the overall rates of oncogenesis were very similar for vectors containing the *IL2RG* transgene, or a neutral transgene such as GFP (6.0% and 7.4%, respectively) and similar to the background rate of leukemiagenesis in animals transplanted with wild type cells without viral transduction (11.1%).

The murine leukemia virus (MLV) promoters used in the SCID-X1 trials have also been implicated as having a role in oncogenesis, both by activating nearby genes and influencing the integration site profile of the vector[27, 28]. In addition, study of the SF promoter revealed that it does have increase potential to generate immortal clones *in vitro*[29]. However, *in vivo* studies in a tumor-prone mouse model suggest that an internal SF promoter in a SIN vector has no increased risk of oncogenesis compared to eukaryotic promoters[30]. The differences in integration profiles seen with the MLV SF vector, including the increased tendency to integrate near proto oncogenes, were shown to be due to genetic selection of clones post-integration, and these differences disappeared once the SF vector was switched to a SIN configuration. In our experiments with SIN lentiviral vector gene therapy for murine SCID-X1, the choice of promoter seemed to have no effect on the rate of leukemogenesis.

The presence of additional safety elements in our 3rd generation SIN lentiviral vectors that were not present in the gammaretroviral vector used in the initial clinical trials, including internal promoters and the deletion of enhancers elements, gives the SIN LV vectors an improved safety profile and makes them an attractive alternative for future clinical trials of SCID-X1. Previous studies conducted with gammaretroviral and lentiviral vectors demonstrated that transcriptionally active LTRs, such as those used in the initial SCID-X1 clinical trials, are major determinants of genotoxicity[30]. The same studies suggested that lentiviral vectors require a 10-fold higher viral load to trigger leukemogenesis than gammaretroviral vectors, likely due to gammaretroviruses' preference to integrate into the promoter regions of active genes and into genes involved in cell growth and cancer[8].

Low risk of oncogenesis is not the same as no risk, and a total of 17 leukemic events were observed in experimental mice. The majority of these leukemias (12 out of 17) were CD4⁺CD8⁺, the same phenotype as the leukemias in the SCID-X1 clinical trials. There were also single instances of CD4⁺, CD8⁺, CD19⁺, CD11b⁺ and Sca-1⁺ leukemias. Fourteen leukemias were found in mice treated with lentiviral vectors, with all but one of these containing vector integrations in the leukemic clones. In most cases, the dominant vector integration sites in the leukemic clones were within 50kb of either a known oncogene or a tumor suppressor gene: prominent oncogenes with flanking integration sites included *Mef2c* and *Ccnd2*. Three leukemias were found in mice transplanted with untransduced wild type Lin⁻ cells. Two of these leukemias were donor-derived, suggesting that the proliferative stress associated with lymphocyte reconstitution was a major factor in leukemia onset. The potential for HSC transplantation to trigger leukemogenesis in *Il2rg*^{-/-} mice makes it difficult to state with surety that insertional mutagenesis is responsible for all the cases of leukemogenesis observed in these mice, even in those mice which have vector insertions in the leukemic clones. Some of these leukemic clones may have been caused by mutations generated during proliferative stress without vector involvement.

Overall, these experiments suggest that the relative risk of developing leukemia in gene therapy treated *Il2rg*^{-/-} mice using 3rd generation SIN lentiviral vectors is unaffected by the use of the SF promoter at low vector copy numbers, and further argues against the idea that the *IL2RG* gene is oncogenic. The use of third generation lentiviral vectors for treatment of SCID-X1 is relatively benign and does not significantly induce insertional mutagenesis when used at low copy number and transplanting a modest number of lineage negative cells. These vectors should be considered for future SCID clinical trials as they may offer a significant benefit to safety without reducing the efficacy of the gene therapy treatment.

Integration site analysis

Thorough analysis of vector integration sites in pre-clinical models would not be possible without automated analysis tools to rapidly and efficiently annotate high numbers of integration site sequences. Automated tools also have the benefits of having a predefined

set of criteria to ensure analysis consistency between datasets. The MAVRIC web tool was designed to automatically validate integration site sequences, align those using BLAST and annotate them using data retrieved from the Ensembl databases. MAVRIC is capable of performing vital QC steps, including removing external sequences such as viral LTRs and clustering redundant sequences to both speed up the BLAST process and keep track of the relative frequency of integration sites.

During analysis of the many lentiviral integration sites generated during *in vitro* and *in vivo* gene therapy experiments, several key factors emerged that should be considered during any analysis project. The first of these is linking IS to nearby genes. Linking each integration to a single gene is a common method of analysis, which allows each IS to have the same impact on secondary analyses such as gene expression arrays, etc. However, it is important to realize that there may be other nearby genes which could be affected by the vector integration, particularly if the vector integrates into an exon or important regulator sequence. For this reason, the genomic region around each integration site up to a certain distance should be considered.

Another factor to consider is the annotation differences generated by changes in NCBI builds, choosing to filter based RefSeq genes, changing the E-value threshold, etc. Analysis results can be significantly altered by minor changes to annotation settings. For clarity and consistency, it is necessary to very explicitly state all parameters used in IS analysis. Without a thorough detailing of the analysis parameters, reproduction of alignments and annotations will prove difficult even when using identical sequence datasets.

MAVRIC was used to annotate over 16,000 lentiviral integration sites from gene therapy experiments for various SCIDs as well as glycogen storage disease type II (Pompe disease). Comparative analysis of the various vector constructs, tissue types, disease phenotypes and species demonstrated that the lentiviral vector's integration profile is stable across a wide range of transduced cells and that the risk of integrating near proto-oncogenes is not affected by promoter or transgene.

The future of gene therapy

Following the discovery of the first leukemias in the SCID-X1 clinical trial, analysis of the vector integration sites revealed integrations proximal to oncogenes LMO2 and CCND4. Investigating the mechanisms behind these adverse events led to the development of sophisticated new methods of facilitating vector integration site analysis. Innovations in vector technology, sequencing and bioinformatics have greatly improved the understanding of viral vector integration patterns and led to more confident safety assessments of proposed gene therapy vectors, pushing gene therapy to the forefront of cutting-edge medicine.

The top priority of current vector development for SCID-X1 is to improve the viral vector safety profile compared the first generation gammaretroviral vectors while retaining good efficacy. To this end, numerous modifications have been implemented, including

switching from strong, constitutively active viral promoters to use of internal promoters and eukaryotic or cell specific promoters, the removal of viral LTRs and codon optimization of the *IL2RG* transgene. In addition to the MLV-derived gammaretroviral and HIV-1 derived lentiviral vector types such as alpharetroviral vectors[31] are being developed. Internal modifications to vector constructs are also being explored, including the addition of miRNAs to improve cell type specificity, adjustments to the viral envelope to improve tropism, and the addition of inducible “kill switches” to deactivate transgene expression in case of adverse events. Non-viral means of delivering nucleic acids are being developed, including zinc finger nucleases[32], transposons[33], mageselectofection[34] and nanoparticles[35]. Rapid technological innovations suggest that many of these genetic modification tools will be tested in the clinic in the near future. In addition, an assortment of next generation sequencing projects and assays necessary for the introduction of personalized, targeted gene therapy treatment strategies are being developed. These projects will greatly benefit from bioinformatics resources such as MAVRIC to study viral integration in regulatory elements such as enhancers regulating gene expression *in vivo*.

In 2012, a new international phase I/II clinical trial for SCID-X1 using an updated gammaretroviral vector was initiated. This new trial, a collaboration of groups in Paris, London and Boston, marks the first new gene therapy trial for SCID-X1 since 2005. The redesigned gammaretroviral vector is self-inactivating, lacking all enhancer-promoter activities of the LTR U3 region. The *IL2RG* gene is under the control of an intron-less version of the human elongation factor 1- α (EFS) promoter. Pre-clinical studies with this vector demonstrated that it is able to produce high levels of *IL2RG* expression while having an improved safety profile over the vector used in the previous trials[36]. Enrollment in this trial is ongoing, and the preliminary results from the first treated patients are promising.

Recently, two reports on ongoing lentiviral hematopoietic stem cell gene therapy trials have been published. These trials for the treatment of metachromatic leukodystrophy[37] and Wiskott-Aldrich syndrome[38], respectively, demonstrate that lentiviral gene therapy has high efficacy and is capable of sustained expression of therapeutic transgenes. Three patients were involved in each trial, and all of them have shown clinical benefit. No adverse events have been reported in any of the patients in these trials, with a follow-up time ranging from 7 to 32 months.

In the past five years, more than 200 clinical gene therapy trials have been initiated, using gammaretroviral lentiviral, adenoviral, and adeno-associated viral vectors to treat diseases ranging from acute lymphoblastic leukemia to hemophilia to Leber’s congenital amaurosis. Many of these trials are reporting marked improvements in the quality of life of patients who had no viable treatment options prior to the advent of gene therapy. Nevertheless, gene therapy can still be considered to be in its infancy. The initiation of new clinical trials, combined with lifelong follow-up of those patients already treated with gene therapy, will further improve our understanding of gene therapy vectors and lead to even more effective and safer treatments in the future.

REFERENCES

1. Cavazzana-Calvo, M., et al., *Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease*. Science, 2000. **288**(5466): p. 669-72.
2. Gaspar, H.B., et al., *Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector*. Lancet, 2004. **364**(9452): p. 2181-7.
3. Aiuti, A., et al., *Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning*. Science, 2002. **296**(5577): p. 2410-3.
4. Bodine, D.M., et al., *Long-term in vivo expression of a murine adenosine deaminase gene in rhesus monkey hematopoietic cells of multiple lineages after retroviral mediated gene transfer into CD34+ bone marrow cells*. Blood, 1993. **82**(7): p. 1975-80.
5. Candotti, F., et al., *Retroviral-mediated gene correction for X-linked severe combined immunodeficiency*. Blood, 1996. **87**(8): p. 3097-102.
6. Hacein-Bey-Abina, S., et al., *Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1*. J Clin Invest, 2008. **118**(9): p. 3132-42.
7. Howe, S.J., et al., *Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients*. J Clin Invest, 2008. **118**(9): p. 3143-50.
8. Schwarzmaelker, K., et al., *Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo*. J Clin Invest, 2007. **117**(8): p. 2241-9.
9. Bushman, F.D., *Retroviral integration and human gene therapy*. J Clin Invest, 2007. **117**(8): p. 2083-6.
10. Ott, M.G., et al., *Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVII, PRDM16 or SETBP1*. Nat Med, 2006. **12**(4): p. 401-9.
11. Gaspar, H.B., et al., *Hematopoietic stem cell gene therapy for adenosine deaminase-deficient severe combined immunodeficiency leads to long-term immunological recovery and metabolic correction*. Sci Transl Med, 2011. **3**(97): p. 97ra80.
12. Challita, P.M. and D.B. Kohn, *Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo*. Proc Natl Acad Sci U S A, 1994. **91**(7): p. 2567-71.
13. Klug, C.A., S. Cheshier, and I.L. Weissman, *Inactivation of a GFP retrovirus occurs at multiple levels in long-term repopulating stem cells and their differentiated progeny*. Blood, 2000. **96**(3): p. 894-901.
14. Pikaart, M.J., F. Recillas-Targa, and G. Felsenfeld, *Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators*. Genes Dev, 1998. **12**(18): p. 2852-62.
15. Lo, M., et al., *Restoration of lymphoid populations in a murine model of X-linked severe combined immunodeficiency by a gene-therapy approach*. Blood, 1999. **94**(9): p. 3027-36.
16. Zhang, F., et al., *Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells*. Blood, 2007. **110**(5): p. 1448-57.
17. Candotti, F., et al., *Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: clinical comparison of retroviral vectors and treatment plans*. Blood, 2012. **120**(18): p. 3635-46.
18. Carbonaro, D.A., et al., *Gene therapy/bone marrow transplantation in ADA-deficient mice: roles of enzyme-replacement therapy and cytoreduction*. Blood, 2012. **120**(18): p. 3677-87.
19. Armitage, J.O., *Bone marrow transplantation*. N Engl J Med, 1994. **330**(12): p. 827-38.
20. Cavazzana-Calvo, M., et al., *Gene therapy for severe combined immunodeficiency*. Annu Rev Med, 2005. **56**: p. 585-602.
21. Aiuti, A., et al., *The chemokine SDF-1 is a chemoattractant for human CD34+ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34+ progenitors to peripheral blood*. J Exp Med, 1997. **185**(1): p. 111-20.
22. Whetton, A.D. and G.J. Graham, *Homing and mobilization in the stem cell niche*. Trends Cell Biol, 1999.

- 9(6): p. 233-8.
23. Elfenbein, G.J. and R. Sackstein, *Primed marrow for autologous and allogeneic transplantation: a review comparing primed marrow to mobilized blood and steady-state marrow*. *Exp Hematol*, 2004. **32**(4): p. 327-39.
 24. Stewart, F.M., et al., *Lymphohematopoietic engraftment in minimally myeloablated hosts*. *Blood*, 1998. **91**(10): p. 3681-7.
 25. Woods, N.B., et al., *Gene therapy: therapeutic gene causing lymphoma*. *Nature*, 2006. **440**(7088): p. 1123.
 26. Dave, U.P., et al., *Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy*. *PLoS Genet*, 2009. **5**(5): p. e1000491.
 27. Modlich, U., et al., *Cell-culture assays reveal the importance of retroviral vector design for insertional genotoxicity*. *Blood*, 2006. **108**(8): p. 2545-53.
 28. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration*. *Nat Biotechnol*, 2006. **24**(6): p. 687-96.
 29. Zychlinski, D., et al., *Physiological promoters reduce the genotoxic risk of integrating gene vectors*. *Mol Ther*, 2008. **16**(4): p. 718-25.
 30. Montini, E., et al., *The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy*. *J Clin Invest*, 2009. **119**(4): p. 964-75.
 31. Suerth, J.D., et al., *Alpharetroviral self-inactivating vectors: long-term transgene expression in murine hematopoietic cells and low genotoxicity*. *Mol Ther*, 2012. **20**(5): p. 1022-32.
 32. Lombardo, A., et al., *Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery*. *Nat Biotechnol*, 2007. **25**(11): p. 1298-306.
 33. Ivics, Z. and Z. Izsvak, *Nonviral gene delivery with the sleeping beauty transposon system*. *Hum Gene Ther*, 2011. **22**(9): p. 1043-51.
 34. Sanchez-Antequera, Y., et al., *Magslectofection: an integrated method of nanomagnetic separation and genetic modification of target cells*. *Blood*, 2011. **117**(16): p. e171-81.
 35. Conley, S.M. and M.I. Naash, *Nanoparticles for retinal gene therapy*. *Prog Retin Eye Res*, 2010. **29**(5): p. 376-97.
 36. Thornhill, S.I., et al., *Self-inactivating gammaretroviral vectors for gene therapy of X-linked severe combined immunodeficiency*. *Mol Ther*, 2008. **16**(3): p. 590-8.
 37. Biffi, A., et al., *Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy*. *Science*, 2013.
 38. Aiuti, A., et al., *Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome*. *Science*, 2013.

SUMMARY

Hematopoietic stem cell (HSC) retroviral gene therapy for treatment of monogenetic diseases has been demonstrated to have high efficacy in clinical trials for X-linked severe combined immunodeficiency (SCID-X1), adenosine deaminase (ADA)-SCID, Wiscott-Aldrich syndrome (WAS), and metachromatic leukodystrophy (MLD). By using retroviral vectors to deliver a healthy copy of the mutated gene to HSCs *ex vivo* and subsequently returning the treated cells to the patient where they engraft in the bone marrow, the majority of patients obtained long-term clinical benefits and in some cases complete remission of the disease phenotype. HSC gene therapy is becoming attractive alternative to bone marrow transplantation for these types of diseases, particularly if no suitable human leukocyte antigen (HLA)-matched donor is available. However, the semi-random nature of retroviral integrations is associated with the risk of triggering insertional mutagenesis. This risk was clearly seen in the 2000 SCID-X1 clinical trials, where 5 out of 20 patients treated with an MLV-derived gammaretroviral vector developed leukemia-like symptoms due to vector integrations flanking, and deregulating, proto-oncogenes. Four of the affected patients were successfully treated by chemotherapy without reducing the effectiveness of the gene therapy, while one patient succumbed in spite of chemotherapy treatment. Leukemias or other hematopoietic abnormalities due to insertional mutagenesis were also seen in clinical trials for WAS and chronic granulomatous disease (X-CGD). In response to these severe adverse events, the scientific community involved in gene therapy focused their efforts on developing new gene therapy vectors and new methods to predict with greater accuracy the relative safety of those vectors prior to their implementation in clinical trials. The contents of this thesis, developing and assessing new lentiviral gene therapy vectors for the treatment of SCID-X1, are part of that communal effort.

Chapter 2 focuses on developing new gene therapy vectors for treatment of SCID-X1 by using new 3rd generation self-inactivating (SIN) lentiviral vectors containing one of two physiological promoters: the cellular phosphoglycerate kinase (PGK) promoter or a 1.1kb region of the *IL2RG* gene promoter (γ cPr), to drive *IL2RG* expression. Physiological promoters are much weaker than the viral promoters used in the 2000 SCID clinical trials and thought to be less likely to deregulate genes flanking integration sites. To ensure that sufficient γ c protein expression was obtained to reverse the SCID phenotype, a codon-optimized *IL2RG* gene (co*IL2RG*) was used. A vector containing the strong viral SF promoter was included for comparison. Vectors were tested *in vivo* by transduction and transplantation of lineage negative (Lin-) bone marrow cells from *Il2rg*^{-/-} mice into *Il2rg*^{-/-} recipients. The efficacy of these vectors was compared to that of wild type Lin- cell transplantation under a range of pre-transplant conditioning regimens. These experiments demonstrated that LV gene therapy using eukaryotic promoters and a codon optimized *IL2RG* gene could reverse the SCID phenotype, provided that a modest level of pre-transplant conditioning was used.

Mice treated with HSC gene therapy are typically subjected to total body irradiation prior to transplantation, in order to create open niches in the bone marrow for donor HSCs to engraft. However, this sort of cytoreductive conditioning is not plausible in severely immunocompromised SCID-X1 patients, who often suffer from multiple recurrent infections at the time gene therapy is initiated. In Chapter 3, a pre-transplantation protocol using granulocyte colony stimulating factor (G-CSF) was initiated to determine if a non-cytoreductive method of creating open bone marrow niches could improve engraftment of donor HSCs. Lentiviral gene therapy experiments in *Il2rg*^{-/-} mice utilizing the G-CSF protocol demonstrated that recipient HSC mobilization improves engraftment of donor HSCs relative to no conditioning regimen, suggesting that G-CSF may be a viable alternative to cytoreductive conditioning methods.

Chapter 4 consists of an analysis of the leukemias which developed in *Il2rg*^{-/-} mice during LV gene therapy experiments. A total of 17 confirmed leukemias were observed, including 7 in mice treated with therapeutic vectors, 7 in mice treated with GFP vectors, and 3 in mice transplanted with untransduced wild type Lin⁻ cells. The rates of leukemia onset were similar for the three groups (6.0%, 7.4% and 11.1%, respectively). The large majority of leukemias found had a CD4⁺CD8⁺ phenotype, although . Prominent proto-oncogenes found flanking integration sites in dominant clones included *Mef2c*, *Mll3*, *Ccnd2* and *Pax5*. Other leukemias had no known oncogenes flanking the dominant integration sites, but did have integrations in or near tumor suppressor genes or transcriptional repressor genes.

Because analyzing integration sites by hand is both time consuming and introduces the possibility of human error, the analysis process was automated. Chapter 5 describes a web-based integration site analysis platform, Methods for Analyzing ViRAl Integration Collections (MAVRIC). MAVRIC is capable of quickly annotating trimmed genomic sequences and returns data on distribution of distances from nearest 5' gene end, nearest gene expression level binning, and Ensembl data on any gene within a user-defined distance from the integration site. MAVRIC was used to annotate integration sites from *IL2RG* gene therapy experiments, as well integration sites generated in other disease models for which lentiviral gene therapy is being evaluated, to generate safety profiles for the vectors under development.

Chapter 6 describes how a large pool of lentiviral integration sites from various *in vitro* and *in vivo* experiments were annotated using MAVRIC. The origins of these integrations included SCID mice, Pompe mice, wild type mice, and murine, human and rhesus HSCs. The integration sites were stratified based on species, disease phenotype, tissue, transgene and promoter to determine any differences in the integration patterns. The resulting comparative analyses demonstrated the stability of lentiviral integration site preferences across species and disease phenotypes, while uncovering some evidence of selective pressure acting on the LV integration profile *in vivo*: namely, that the frequency of integration sites in or near genes is reduced in transplanted cells compared to cultured HSCs.

The experiments described in this thesis demonstrate that lentiviral gene therapy vectors using a codon optimized IL2RG gene driven by select physiological promoters is effective for treatment of SCID-X1 when combined with some level of pre-transplant conditioning or HSC mobilization, even at one vector copy per cell. Analysis of vector integration patterns and leukemias observed in experimental animals demonstrates that lentiviral HSC gene therapy has a low risk of triggering oncogenic events *in vivo* and suggests that the promoter and transgene have little or no effect on the relative safety of self-inactivating lentiviral vectors. Combined, these results argue that SIN LV gene therapy is sufficiently safe for the treatment of life-threatening diseases such as SCIDs.

SAMENVATTING (DUTCH SUMMARY)

Retrovirale genterapie van hematopoïetische stamcellen (HSCs) voor de behandeling van monogene erfelijke ziekten is effectief gebleken in klinische trials voor *X-linked severe combined immunodeficiency* (SCID-X1), adenosine deaminase (ADA)-SCID, Wiscott-Aldrich syndroom (WAS), en metachromatische leukodystrofie (MLD). Met gebruik van retrovirale vectoren om een gezonde kopie van het gemuteerde gen *ex vivo* te integreren in het genoom van HSC's en de behandelde cellen terug te transplanteren naar het beenmerg van de patiënt werd bij het merendeel van de patiënten het fenotype partieel of volledig gecorrigeerd. Stamcelgenterapie werd daarmee een aantrekkelijk alternatief voor allogene beenmergtransplantatie voor dit type aandoeningen, in het bijzonder indien een geschikte HLA-identieke donor niet beschikbaar is. Echter, het semi-random integratiepatroon van de retrovirale vectoren is geassocieerd met het risico van insertiemutagenese. Dit risico bleek in de eerste klinische trials voor SCID-X1, waarin 5 van de 20 behandelde patiëntjes leukemie ontwikkelden als gevolg van deregulerende vector-integraties nabij proto-oncogenen. Vier patiëntjes werden daarvoor met succes behandeld met behoud van immuun competentie, één overleed als gevolg van complicaties van de nodig geachte allogene stamceltransplantatie. Leukemie en andere hematopoïetische abnormaliteiten ten gevolge van insertionele mutagenese werden ook waargenomen in de klinische trials voor WAS en chronische granulomateuze ziekte (CGD). In antwoord op deze *severe adverse events* worden in het kader van samenwerkingsprojecten gesubsidieerd door de Europese Commissie nieuwe gentransfer vectoren ontwikkeld en nieuwe methoden om accuraat de veiligheid van deze vectoren bij gebruik in klinische trials te voorspellen. Dit proefschrift, waarin de effectiviteit en veiligheidsstudies worden beschreven van lentivirale stamcelgenterapie maakt deel uit van deze inspanning.

Hoofdstuk 2 is gericht op de ontwikkeling van nieuwe gentransfer vectoren voor de behandeling van SCID-X1 door gebruik te maken van derde generatie zogenaamde self-inactivating (SIN) lentivirale vectoren met fysiologische promoters, i.e., de cellulaire phosphoglyceraat kinase (PGK) promotor of een 1.1kb deel van de *IL2RG* gen promotor (γ cPr) om expressie van het *IL2RG* gen te bewerkstelligen. Fysiologische promoters zijn veel zwakker dan de virale promoters die gebruikt werden in de eerste SCID-X1 klinische trials en daarmee minder waarschijnlijk in staat tot deregulering van flankerende genen. Om er zeker van te zijn dat het γ c eiwit voldoende tot expressie zou komen werd een *codon-optimized* *IL2RG* gen (co*IL2RG*) geconstrueerd. Een vector met de krachtige virale SF promotor diende als controle ter vergelijking van de effectiviteit. De vectoren werden *in vivo* getest door transductie en transplantatie van *lineage negative* (Lin-) beenmergcellen van *Il2rg*^{-/-} muizen naar *Il2rg*^{-/-} ontvangers. De effectiviteit werd vergeleken met die van normale Lin- cellen met gebruikmaking van verschillende pre-transplantatie conditioneringsregimes. De experimenten toonden aan dat LV genterapie met eukaryote promoters en een *codon optimized IL2RG* gen het SCID fenotype volledig corrigeert, mits een milde pre-transplantatie conditionering werd gebruikt.

Muizen die worden behandeld met HSC genterapie worden voorbehandeld met totale lichaamsbestraling voorafgaand aan de transplantatie teneinde open niches te creëren in het beenmerg voor de getransplanteerde stamcellen. Deze cyto-reductieve conditionering is echter niet geschikt voor SCID-X1 patiënten, die doorgaans bij diagnose al lijden aan meerdere chronische en levensbedreigende infecties. In hoofdstuk 3 wordt een pretransplantatie protocol met gebruik van *granulocyte colony stimulating factor* (G-CSF) beschreven om te onderzoeken of een niet-cyto-reductieve benadering om open beenmerg niches te creëren voldoende zou zijn voor transplantatie van autologe HSCs. Lentivirale genterapie experimenten met *Il2rg*^{-/-} muizen en het G-CSF protocol toonden aan dat mobilisatie van ontvanger de effectiviteit van de transplantatie ten opzichte van geen conditionering verbetert, wat er op wijst dat G-CSF conditionering een levensvatbaar alternatief zou kunnen zijn voor cyto-reductieve conditionering.

Hoofdstuk 4 geeft een analyse weer van gevallen van leukemie die tot ontwikkeling kwamen in *Il2rg*^{-/-} muizen gedurende de LV genterapie experimenten. In totaal werd in 17 muizen leukemie bevestigd, waaronder 7 in muizen behandeld met therapeutische vectoren, 7 in muizen behandeld met GFP vectoren en 3 in muizen die werden behandeld met ongetransduceerde normale stamcellen. De frequentie van het optreden van leukemie in de drie groepen was niet significant verschillend (6.0%, 7.4% en 11.1%, respectievelijk). De overgrote meerderheid van de gevallen van leukemie had een CD4⁺CD8⁺ fenotype. Onder proto-oncogenen die werden gevonden naast integraties in dominante klonen waren *Mef2c*, *Mill3*, *Cnd2* en *Pax5* prominent aanwezig. Bij andere gevallen werden echter geen oncogenen aangetroffen, maar integraties nabij tumor suppressor genen of transcriptie repressor genen.

Aangezien de handmatige analyse van integraties tijdrovend is en menselijke fouten niet uitsluit werd een geautomatiseerde analyse ontworpen. Hoofdstuk 5 beschrijft een web-based platform voor integratie-analyses, genaamd *Methods for Analyzing ViRal Integration Collections* (MAVRIC). Met MAVRIC kunnen gensequenties snel worden geannoteerd en gegevens worden verkregen over de afstand tot de dichtstbijzijnde 5' eind van een gen, de relatie tot het genexpressie niveau van het dichtstbijzijnde gen vastgesteld, en Ensembl gegevens verkregen voor elk gen met een door de gebruiker gedefinieerde afstand tot de integraties. MAVRIC is als onderdeel van het veiligheidsprofiel gebruikt voor de analyse van integraties in de *IL2RG* genterapie experimenten en in andere diermodellen waarin lentivirale genterapie wordt geëvalueerd.

In hoofdstuk 6 wordt beschreven hoe een reeks lentivirale integraties werd geanalyseerd met MAVRIC. De integraties waren afkomstig van SCID muizen, Pompe muizen, normale muizen en van stamcellen van de muis, de resusaap en de mens en gerangschikt naar diersoort, fenotype van de onderliggende aandoening, weefsel, transgen en promotor om potentiële verschillen in het integratiepatroon op het spoor te komen. De resultaten toonden de stabiliteit van het lentivirale integratiepatroon tussen diersoorten en ziekte-fenotypen aan. Tevens werd aanwijzing gevonden voor een zekere mate van selectiedruk

op het LV integratiepatroon *in vivo*, i.e., de frequentie van lentivirale integraties in of nabij genen neemt af na transplantatie in vergelijking tot *in vitro* doorgekweekte HSCs.

De experimenten beschreven in dit proefschrift tonen aan dat lentivirale gentransfer vectoren met een *codon-optimized* IL2RG gen onder geselecteerde fysiologische promoters effectief is voor de stamcelgentherapie van SCID-X1 in combinatie met een milde vorm van pre-transplantatie conditionering van de ontvangers, zelf bij 1 vectorkopie per cel. Analyse van het integratiepatroon en het optreden van leukemie liet zien dat het risico op oncogene transformatie *in vivo* na lentivirale stamcelgentherapie gering is, met weinig tot geen effect van promotor en transgen op de relatieve veiligheid van *self-inactivating* lentivirale vectoren. Concluderend pleiten deze gegevens ten gunste van SIN LV gentherapie als voldoende effectief en veilig voor de behandeling van levensbedreigende aandoeningen zoals SCID.

LIST OF TERMS AND ABBREVIATIONS

ADA	adenosine deaminase
APC	antigen presenting cell
BLAST	Basic Local Alignment Search Tool
BM	bone marrow
BMT	bone marrow transplantation
CCND2	cyclin D2 (proto-oncogene)
CD34	cluster of differentiation 34 protein (human HSC marker)
cDNA	complimentary DNA
CIS	common integration site
coIL2RG	codon optimized IL2RG gene
co γ c	codon optimized IL2RG gene
DNA	deoxyribonucleic acid
ELISA	enzyme linked immunosorbent assay
FACS	fluorescence activated cell sorting
FIt3-L	FMS-like tyrosine kinase 3-ligand
G-CSF	granulocyte colony stimulating factor
GFP	green fluorescent protein
GVHD	graft vs host disease
Gy	gray (irradiation dose)
HBSS	Hank's balanced salt solution
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HSC	hematopoietic stem cell
Ig	Immunoglobulins
IL7	Interleukin-7
IL2RG	Interleukin-2 receptor gamma
IS	integration site
ITAM	immunoreceptor tyrosine-based activation motifs
JAK	janus kinase
LAM-PCR	linear amplification-mediated PCR
Lin-	lineage negative
LMO2	LIM domain only 2 (proto-oncogene)
LM-PCR	ligation-mediated PCR
LSK	lineage negative, Sca-1+, C-kit+
LTR	long terminal repeat
LV	lentiviral
MAVRIC	Methods for Analyzing ViRal Integration Collections web tool
MOI	multiplicity of infection
NCBI	National Center for Biotechnology Information
NK	natural killer cell

PCR	polymerase chain reaction
PGK	phosphoglycerate kinase promoter
qPCR	quantitative PCR
RAG	recombination activating gene
RTCGD	Retroviral Tagged Cancer Gene Database
RV	gammaretroviral
SCF	stem cell factor
SCID	severe combined immunodeficiency
SEM	standard error of the mean
SF	spleen focus forming virus promoter
SIN	self-inactivating
STAT	signal transducer and activator of transcription
TCR	T cell receptor
TNP-KLH	trinitrophenol keyhole limpet hemocyanin
TPO	thrombopoietin
UCB	umbilical cord blood
UCOE	ubiquitous chromatin opening element
γ c	common gamma chain protein
γ cPr	1.1kb region of the <i>IL2RG</i> gene promoter

PHD PORTFOLIO

Summary of PhD training and teaching activities

Name PhD student: Marshall William Huston PhD period: Dec. 2006 – Dec.2011
Erasmus MC Department: Hematology Promotor(s): Prof. dr. Gerard Wagemaker
Research School: COEUR Supervisor: Dr. Niek van Til

1. PhD training	Year	Workload (Hours/ECTS)
Courses		
• Biomedical Research Techniques V	2006	1.5
• Laboratory animal science	2007	3.0
• Principles of research in medicine	2009	0.7
• Matlab fundamentals and statistical method	2009	0.5
• Basic course on 'R'	2010	1.0
Presentations		
• Oral national conference (NVGCT, Dutch Hematology conference) (2x)	2009-2010	3.0
• Oral international conference (ESGCT, ASGCT, CONSERT) (6x)	2007-2011	9.0
• Poster international conference (ESGCT, ASGCT, CONSERT/PERSIST) (7x)	2007-2011	10.5
International conferences		
• European Society of Cell and Gene Therapy (2x)	2007-2008	2.0
• American Society of Cell and Gene Therapy (4x)	2008-2011	4.0
• CONSERT/PERSIST annual meeting (5x)	2007-	5.0
Seminars and workshops		
• Applied bioinformatics	2007	0.25
• Browsing genes and genomes with Ensembl	200	0.3
• Basic data analysis on gene expression arrays (BAGE)	2010	0.75
• Photoshop and Illustrator C5 workshop for PhD student	2011	0.3
2. Teaching activities		
Lecturing		
• Insertional mutagenesis and LMO2: consequences for gene therapy (medical student course "Keuzeonderwijs genterapie")	2008-2010	1.5
Supervising, practicals and excursions		
• Supervising medical students during their research internships (2 students, each 1 year internship)	2010-2011	4.0

PUBLICATIONS

Huston MW, van Til NP, Visser TP, Arshad S, Brugman MH, Cattoglio C, Nowrouzi A, Li Y, Schambach A, Schmidt M, Baum C, von Kalle C, Mavilio F, Zhang F, Blundell MP, Thrasher AJ, Verstegen MM and Wagemaker G. *Correction of murine SCID-X1 by lentiviral gene therapy using a codon optimized IL2RG gene and minimal pre-transplant conditioning*. Molecular Therapy (2011) vol. 19 no. 10, 1867–77.

Huston MW, Brugman MH, Horsman S, Stubbs A, van der Spek P and Wagemaker G. *Comprehensive investigation of the parameters of viral integration site analysis and their effects on the gene annotations produced*. Human Gene Therapy (2012) vol. 23 no. 11, 1209-19.

van Til NP, de Boer H, Mashamba N, Wabik A, Huston MW, Visser TP, Fontana E, Poliani PL, Cassani B, Zhang F, Thrasher AJ, Villa A and Wagemaker G. *Correction of Immunity in Murine Rag2 Severe Combined Immunodeficiency by Lentiviral Gene Therapy Using a Codon-Optimized RAG2 Therapeutic Transgene*. Molecular Therapy (2012) vol. 20 no. 10, 1968-80.

Sanchez-Antequera Y, Mykhaylyk O, van Til NP, Cengizeroglu A, de Jong JH, Huston MW, Anton M, Johnston IC, Pojda Z, Wagemaker G and Plank C. *Magslectofection: an integrated method of nanomagnetic separation and genetic modification of target cells*. Blood (2011) vol. 117 no. 16 e171-e181.

CURRICULUM VITAE

Name Marshall William Huston
Date of birth September 20, 1979
Place of birth Happy Camp, California, USA
Nationality American

Positions

June 2013- Present Postdoc in the group of Dr. I.F.M. de Coo, dept. of Neurology. Erasmus University Medical Center, Rotterdam, the Netherlands.

Dec. 2011- May 2013 PhD position in the group of Prof. G. Wagemaker, dept. of Hematology. Erasmus University Medical Center, Rotterdam, the Netherlands.

May 2004- Aug. 2006 Research Scientist in the group of Dr. X. Wu, dept. of mRNA profiling. Roche Molecular Systems, Pleasanton, California, USA.

Education

Dec. 2006- Nov. 2011 PhD student, lab of Prof. G. Wagemaker, dept. of Hematology. Erasmus University Medical Center, Rotterdam, the Netherlands.

Aug. 2002- Dec. 2003 MSc in Bioinformatics, supervisor: S. Nilsson. Chalmers University of Technology, Göteborg, Sweden.

Aug. 1997- Dec. 2001 BA in Molecular and Cell Biology, University of California. Berkeley, California, USA.

Grants

April 2013 Energy4Kids/Ride4Kids: awarded a 2 year grant for the development of mitochondrial-targeted AAV gene therapy

Jan. 2013 Sophia Foundation Stichting: awarded a 2 year grant for the development of mitochondrial-targeted AAV gene therapy for Leber's Hereditary Optic Neuropathy (LHON)

ACKNOWLEDGEMENTS

First and foremost, I give thanks to my wonderful family. Mom and dad, from the moment of my birth you've given me nothing but love, comfort and good counsel. Even when I said I was going to quit my job and move thousands of miles away to tinker with weird science - for the second time in 5 years! - you were 100% supportive. If it wasn't for you I'd probably be working in a circus somewhere. Thanks also to my sibs, Kyle and Lindsey, for always having my back. You know I'll always have yours. I still owe you reparations for all the Indian rug burns, purple nurple and assorted browbeatings I administered when we were kids. I promise I'll pay you back as soon as I make my second million dollars (the first million is reserved for buying a circus). A big thank you to my uncle Ralph, who helped me get my first job in science and also assured me that the Netherlands was a nice place to live in spite of the weather. He was right, as usual. To Janet and Gail and Alexandra and Sherry and Frank and Angela and Sara and grandpa & grandma J and grandma H and grandpa H (RIP): thanks for being there, I love you all and I wish I got to see you more than once or twice a year. Blood really is thicker than water: I have the lab results to prove it.

Next, I'd like to thank my promoter and co-promoter. Gerard, you hired me based on a 30-minute interview over a cup of coffee at SFO. Thanks for believing in me, and for all of the guidance and financial/logistical support along the way. This has truly been a life-changing experience for me. Niek, thanks for being my co-promoter and an endless fountain of knowledge about lentiviral vectors. I'll miss all of the travel adventures we had to the various conferences, and of course the hot springs at Leukerbad! I look forward to collaborating with you again in the future, be it in the Netherlands, the USA, or someplace even further out.

Heartfelt thanks to my inner doctoral committee, esteemed professors Bobby Gaspar, Jan Cornelissen and Peter van der Spek, for making sure this thesis was up to snuff. I really appreciate you taking the time out of your busy schedules to comment on my research, and I'm honored to have had the chance to present my work to you in person.

To Merel, my partner in the three-legged race to finish our PhDs: we did it! Thanks for all the supporting words and encouragement during the grueling push to get this thing done, and thanks for all the silly pranks to keep us both sane during the long hours in the lab. I hope you get your very own office soon, so that I can cover it in tinfoil. Martijn Brugman, you were my first mentor in the Netherlands. Thanks for showing me around Rotterdam and for setting the groundwork for MAVRIC. Monique Versteegen, thanks for telling me that "you'll make it" at the end of my first year as a PhD I'm glad I proved you right. To all the other PhD students I met in the Wagemaker group, I say thanks for

sharing your journey with me. Fatima and Elnaz, I'm so happy to join the "I survived my PhD studies" club: thanks for paving the way for me. Rana and Qiushi, hang in there! No PhD work is ever easy, but I know you girls have what it takes to make it to the finish line. Leonie, thanks for always pushing me to branch out into new projects and experiences in gene therapy. I wouldn't be in my current position without your endless enthusiasm and support. And special thanks to Egied for doing a great job on the layout of this thesis, and for helping me optimize many figures for publication over the years.

Dank u wel to all the postdocs, technicians and students (both master and medical) in the Wagemaker group who've helped me with my experiments over the years. Helen, thanks for being a great friend as well as a great technician...and I swear, I *will* jump in the fountain at Weena just as soon as Holland wins the World Cup in 2014. Yvette and Guus, I'm looking forward to working/joking/playing MtG with you in the coming year. Henk and Daan: best of luck with your medial careers, you will make fine doctors. Amir, Aya, Carla, Claudia, Nora, Roya, Shazia, Wendy and Yuedan: thanks a bunch for giving me a hand. And of course, Trui: one part technician, one part opera singer, five parts enthusiasm. Madame Visser, you will always be the first lady of the lab.

Andrew and Bas, thanks a ton for all your help with the Bioinformatics work in my thesis. At best, I'm a tinkerer in computer science, coding at around a 5th grade level...without your support, MAVRIC never would have gotten off the ground. Hope we get to work together on another great project soon!

To all my other Hema friends: thanks for all the good times! Beer clubs, dance nights, Halloween parties, that amazing trip to France...you guys and gals kept me happy and productive during the long (long, loooong) Dutch winters and made every summer day the best it could be. Onno, Joyce, Nikken, Si, Kasia, Jana, Adrian and Amiet, you guys rock. And to Eric V: hurry back, buddy! Nobody else here in Rdam likes to play video games or drink whiskey nearly enough to keep up with me.

A big shout out to my Berkeley boys: Davis, Sean, Pouya and Carrey. Still friends after all these years, and still kicking ass in our respective fields. I learned just as much from hanging out with you guys as I did in my college courses, and what you taught me has had a much broader effect on my life since. All hail Castle Haystack! Lonny and Jamie, you guys are my oldest friends, and the only ones who can really appreciate how just far away Seiad and Happy Camp are, physically and spiritually, from the 'real world.' Here's to another 20 years of friendship. I look forward to making fun of your white hair (Lonny) or complete lack of hair (Jamie) in 2033. Oh, and Lonny...see you on B-net.

Grazie a mi amore, Noemi Angela Zambetti. Pasticcino, you keep me sane when things are getting too crazy, and help me go crazy when things are in danger of becoming too sane. We've been together almost three years now (time flies!) and I enjoy your company more every day. You are the light of my life and the reason I smile in my sleep. I'm looking forward to all the nice things you write about me in your own thesis acknowledgements (coming next year, fingers crossed).

And finally, thanks to everyone else who helped me out along the way, whether it was in the Netherlands, California, Sweden, Singapore, or parts in between. I wish I had time to thank you all by name here, but I'm out of room. Instead, I'll do it in person, over a beer (or coffee, or tomato juice... whatever's clever). Cheers!

