Molecular Therapy
Original Article

Lentiviral Vectors with Cellular Promoters Correct Anemia and Lethal Bone Marrow Failure in a Mouse Model for Diamond-Blackfan Anemia

Shubhranshu Debnath,[1] Pekka Jaako,[1] Kavitha Siva,[1] Michael Rothe,[2] Jun Chen,[1] Maria Dahl,[1] H. Bobby Gaspar,[3] Johan Flygare,[1] Axel Schambach,[2,4] and Stefan Karlsson[1]

[1]Molecular Medicine and Gene Therapy, Lund Strategic Center for Stem Cell Biology, Lund University, Lund 22184, Sweden; [2]Institute of Experimental Hematology, Hannover Medical School, Hannover 30625, Germany; [3]Molecular Immunology Unit, Institute of Child Health, University College London, London WC1N 1EH, UK; [4]Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

Diamond-Blackfan anemia is a congenital erythroid hypoplasia and is associated with physical malformations and a predisposition to cancer. Twenty-five percent of patients with Diamond-Blackfan anemia have mutations in a gene encoding ribosomal protein S19 (RPS19). Through overexpression of RPS19 using a lentiviral vector with the spleen focus-forming virus promoter, we demonstrated that the Diamond-Blackfan anemia phenotype can be successfully treated in Rps19-deficient mice. In our present study, we assessed the efficacy of a clinically relevant promoter, the human elongation factor 1α short promoter, with or without the locus control region of the β-globin gene for treatment of RPS19-deficient Diamond-Blackfan anemia. The findings demonstrate that these vectors rescue the proliferation defect and improve erythroid development of transduced RPS19-deficient bone marrow cells. Remarkably, bone marrow failure and severe anemia in Rps19-deficient mice was cured with enforced expression of RPS19 driven by the elongation factor 1α short promoter. We also demonstrate that RPS19-deficient bone marrow cells can be transduced and these cells have the capacity to repopulate bone marrow in long-term reconstituted mice. Our results collectively demonstrate the feasibility to cure RPS19-deficient Diamond-Blackfan anemia using lentiviral vectors with cellular promoters that possess a reduced risk of insertional mutagenesis.

INTRODUCTION

Diamond-Blackfan anemia (DBA) is a rare inherited bone marrow failure disorder with pure red blood cell aplasia manifesting early in life. The hematological profile of patients with DBA shows macrocytic anemia with reticulocytopenia, normal or decreased levels of neutrophils, and variable platelet counts.[1] Patients with DBA also exhibit various non-hematological manifestations such as physical abnormalities and cancer predisposition.[2,3]

In at least 60%–70% of cases, DBA is caused by functional haploinsufficiency of genes encoding for ribosomal proteins.[4–11] Recent studies have discovered that two genes, erythroid transcriptional factor *GATA1* and *TSR2* (a direct binding partner of RPS26), can cause the DBA phenotype.[12–15] Twenty-five percent of patients have mutations in a gene coding ribosomal protein S19 (RPS19).[4] For given mutations all reported patients are heterozygous. Furthermore, in most cases, the mutations are predicted to result in haploinsufficiency of the respective ribosomal protein.[16,17] Corticosteroids are the main therapeutic option in DBA.[3] Around 80% of patients initially respond to corticosteroids, but only 40% of patients sustain the therapeutic response and the remaining 40% need chronic blood transfusion. Twenty percent of patients go into spontaneous remission and maintain acceptable hemoglobin levels without therapeutic intervention. Allogeneic bone marrow transplantation is currently the only curative treatment available for patients with DBA.[18]

Our previous studies demonstrated that enforced expression of RPS19 improves the proliferation, erythroid colony-forming potential, and differentiation of patient-derived RPS19-deficient hematopoietic progenitor cells in vitro.[19,20] Moreover, RPS19 overexpression enhances the engraftment and erythroid differentiation of patient-derived hematopoietic stem cells and progenitor cells when transplanted into immune-compromised mice.[21] Collectively, these studies suggest that gene therapy may be a future therapeutic modality in the treatment of RPS19-deficient DBA. In our proof-of-principle study using lentiviral vectors harboring the spleen focus-forming virus (SFFV) promoter and a codon-optimized human RPS19 cDNA followed by the internal ribosomal entry site (IRES) and GFP (SFFV-RPS19), we showed that the DBA phenotype of Rps19-deficient mice can be successfully treated.[22]

In the current study, we assessed the efficacy of clinically relevant promoters to drive the therapeutic gene. To this effect, we designed

Received 29 November 2016; accepted 1 April 2017;
http://dx.doi.org/10.1016/j.ymthe.2017.04.002.

Correspondence: Stefan Karlsson, Division of Molecular Medicine and Gene Therapy, Molecular Medicine and Gene Therapy, Lund Strategic Center for Stem Cell Biology, Lund University, BMC A12, 22184 Lund, Sweden.
E-mail: stefan.karlsson@med.lu.se

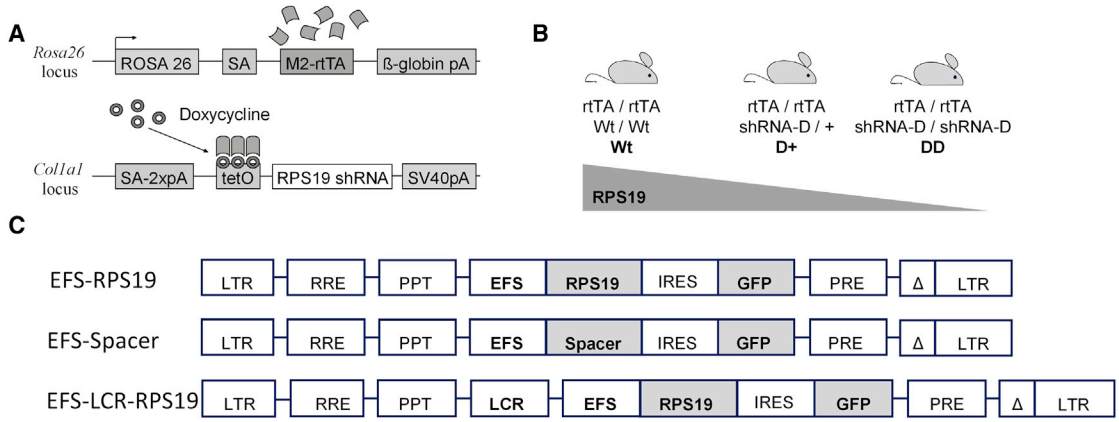


Figure 1. Mouse Model for RPS-Deficient DBA and SIN Lentiviral Vectors for DBA Gene Therapy

Transgenic mice containing a doxycycline-regulatable Rps19-targeting shRNA allow an inducible and graded downregulation of Rps19. (A) Overview of modified loci. Black arrowheads indicate TSSs. (B) Breeding strategy to adjust the level of Rps19 downregulation. (C) EFS-RPS19 vector, codon-optimized human *RPS19* cDNA was constructed under the control of the human elongation factor 1α short (EFS) promoter and inserted into a lentiviral vector. Following the *RPS19* cDNA, an internal ribosomal entry site (*IRES*), a GFP sequence, and improved post-transcriptional regulatory element (*Pre**) were inserted. EFS-Spacer vector, in which the *RPS19* cDNA was replaced with an equally long non-coding spacer sequence, was used as a control. The LCR-EFS-RPS19 vector, where in locus control region of β-globin gene was inserted before the EFS promoter. LTR, long terminal repeat; pA, polyadenylation signal; PPT, polypurine tract; RRE, Rev response element; SA, splice acceptor; Wt, wild-type.

lentiviral vectors harboring a codon-optimized human RPS19 cDNA driven by the shortened version of the human elongation factor 1α (EFS) promoter. Lentiviral vectors with the EFS promoter are shown to have a significantly decreased risk of insertional mutagenesis,[23,24] and no evidence of clonal dominance was reported during clinical trials of gene therapy for severe combined immunodeficiency X1 (SCID-X1) using the EFS promoter.[25]

The EFS promoter was followed by IRES and GFP (EFS-RPS19), while a vector without the RPS19 cDNA was used as a control (EFS-Spacer). To assess the therapeutic potential of the EFS-RPS19 vector in vivo, we transduced c-Kit-enriched bone marrow cells from control and uninduced small hairpin RNA (shRNA)-D mice and these were injected into lethally irradiated wild-type mice. The recipients transplanted with the EFS-Spacer transduced shRNA-D bone marrow showed a dramatic decrease in blood cellularity that led to death after a few weeks, while the recipients transduced with EFS-RPS19 shRNA-D bone marrow exhibited close to normal blood cellularity. These results demonstrate that EFS promoter-driven enforced expression of RPS19 can cure severe anemia and bone marrow failure in RPS19-deficient mice.

RESULTS

Enforced Expression of RPS19 by the EFS Promoter in Rps19-Deficient Bone Marrow Cells Improves Proliferation and Erythroid Development In Vitro

We have shown that enforced expression of RPS19 expands erythroid development in RPS19-deficient patients with DBA.[19–21] In our previous study using lentiviral vectors driven by the SFFV promoter, we showed that the DBA phenotype of mice can be successfully treated.[22] In this study, we assessed the efficacy of clinically relevant promoters like the EFS promoter in our mouse model of RPS19-deficient DBA. Briefly, this model contains an Rps19-targeting shRNA (shRNA-D) that is expressed under a doxycycline-responsive promoter located downstream of the collagen A1 gene (Figure 1A). Experimental animals were bred to be either heterozygous (D+) or homozygous (DD) for the shRNA in order to generate two models with intermediate or severe Rps19 deficiency, respectively (Figure 1B). To correct the Rps19 deficiency, we developed self-inactivating (SIN) lentiviral vectors harboring a codon-optimized human *RPS19* cDNA driven by the internal *EFS* promoter, followed by *IRES* and *GFP* (EFS-RPS19) with or without a β-globin locus control region (*LCR*) cassette (Figure 1C). The codon-optimized *RPS19* cDNA was further modified to prevent its recognition and downregulation by the *Rps19*-targeting shRNA used. A similar vector without the *RPS19* cDNA was used as a control vector (EFS-Spacer).[22,26–28]

To assess the functionality of these vectors, we cultured transduced c-Kit-enriched bone marrow (BM) cells from control and heterozygous RPS19 shRNA (D+) mice in liquid cultures in the presence of doxycycline (Figure 2A). Based on the percentage of GFP+ cells, the initial transduction efficiency was between 20% and 40% on average (Figure 3B). D+ cells transduced with the EFS-Spacer control vector failed to expand during 7 days of culture after transduction (Figure 2B). In contrast, the EFS-RPS19 and LCR-EFS-RPS19 vectors mediated a 2-fold increase in total cell number compared to the EFS-Spacer vector.

Next we quantified the erythroid colony-forming potential of transduced c-Kit-enriched BM cells from control and D+ mice in methyl cellulose cultures in the presence of doxycycline for 14 days (Figure 2C). The findings demonstrate that the EFS-RPS19 and LCR-EFS-RPS19 vectors mediated a 3-fold increase in the total number of erythroid colonies compared to the EFS-Spacer vector.

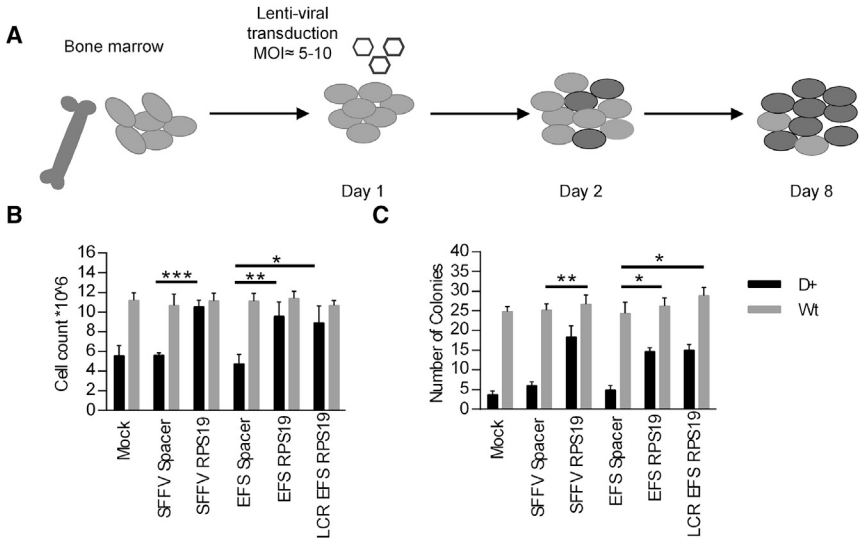


Figure 2. Enforced Expression of RPS19 Derived from the EFS Promoter Is Sufficient to Rescue the DBA Phenotype In Vitro

c-Kit-enriched hematopoietic progenitors (0.5×10^6) from the BM of uninduced mice were transduced and seeded in liquid culture or methyl cellulose in the presence of doxycycline. (A) Experimental design. (B) Total cell counts on day 8 after growth in liquid culture. (C) Total erythroid colony counts in methyl cellulose cultures (M3436) in the presence of doxycycline on day 14. Data shown in (B) and (C) represent the average of three independent experiments with three technical replicates. *$p < 0.05$; ***$p < 0.001$. Wt, wild-type.

Enforced Expression of RPS19 by the EFS Promoter Is Sufficient to Rescue the DBA Phenotype In Vivo

Subsequently, we probed whether EFS-RPS19 and LCR-EFS-RPS19 vectors generate a sufficient amount of RPS19 in vivo in order to assay the therapeutic efficacy. Doxycycline administration to transplanted recipients with the homozygous RPS19 shRNA (DD) genotype causes acute and lethal BM failure, while D+ recipients (one RPS19 shRNA allele) develop mild chronic anemia.[28] Since DD mice develop lethal BM failure shortly after doxycycline administration, we chose this model to test the efficacy of gene correction to rigorously assess whether the lethal phenotype could be rescued and the mice cured. Uninduced BM cells from the control and DD mice were transduced with the vectors, and the transduced cells were transplanted into wild-type recipient mice. Following engraftment and stable donor-derived regeneration of the hematopoietic system, recipient mice were administered doxycycline to downregulate endogenous *Rps19* in order to induce the disease (Figure 3A). Since we showed previously that the hematopoietic phenotype in Rps19-deficient mice is autonomous to the blood system, we decided to use lethally irradiated wild-type recipients.[28]

Before transplantation, initial transduction efficiencies with therapeutic and control vectors were measured based on the percentage of GFP+ cells and were between 20% and 40% on average (Figure 3B). After 2 weeks of doxycycline treatment, most of the mice receiving DD BM transduced with EFS-Spacer vector died due to dramatic BM failure (data not shown). At this time point, all groups showed high overall donor reconstitution, confirming the absence of recipient-derived hematopoiesis (Figures 3C and 3D). We demonstrated that the recipients transplanted with the EFS-RPS19 or LCR-EFS-RPS19 DD BM had normal blood cellularity (Figures 3E and 3F).

Doxycycline administration for 18 weeks was used as the time point to assess long-term efficacy (Figure 4A). Most recipients with DD BM transduced with EFS-Spacer vectors died, but the remaining surviving recipients exhibited macrocytic anemia and a decrease in erythrocyte numbers, hemoglobin value, and platelet counts (Figures 4B–4H). Remarkably, recipients transplanted with the EFS-RPS19 or LCR-EFS-RPS19 DD BM had normal blood cellularity and BM cellularity (Figures 4C–4H). Additionally, we analyzed the samples by fluorescence-activated cell sorting (FACS) to allow fractionation of the myeloid-erythroid compartment in the BM.[28,29] The mean percentage of GFP+ cells was substantially higher in recipients with EFS-RPS19 or LCR-EFS-RPS19 DD BM than in the other groups, indicating the competitive advantage of gene-corrected cells in the hematopoietic hierarchy (Figure 5).

RPS19-Deficient BM Cells Transduced with RPS19 Vectors Provide Long-Term Reconstitution

We asked whether doxycycline-induced, Rps19-defcient BM cells transduced with RPS19 lentiviral vectors can result in long-term engraftment in doxycycline-induced lethally irradiated wild-type recipient mice (Figure 6A). To this end, DD and control mice were induced with doxycycline for 1 week and erythrocyte numbers and hemoglobin levels were measured to confirm the DBA phenotype (Figures 6B and 6C). BM cells from induced mice were transduced and transplanted into doxycycline-induced lethally irradiated mice. Initial transduction efficiencies with therapeutic and control vectors were measured based on the percentage of GFP+ cells and were between 20% and 50% (Figure 6D). Most of the mice receiving DD BM transduced with EFS-Spacer failed to engraft and did not survive beyond 2–3 weeks after transplantation. Almost 60% of the mice receiving DD BM with corrected EFS-RPS19 vector survived and showed long-term engraftment (Figure 6E). We assessed long-term engraftment and the hematopoietic contribution of mice with gene-corrected DD BM at 16 weeks post-transplantation. At this point, these mice exhibited improved BM cellularity and recovery of erythrocyte numbers, hemoglobin levels, and platelet counts (Figures 6F–6K).

Gene-Corrected Rps19-Deficient Cells Show Polyclonal Hematopoiesis and Have a Typical Lentiviral Insertion Profile

The risk of insertional mutagenesis is a major concern regarding the future clinical use of lentiviral vectors. To assess the safety of

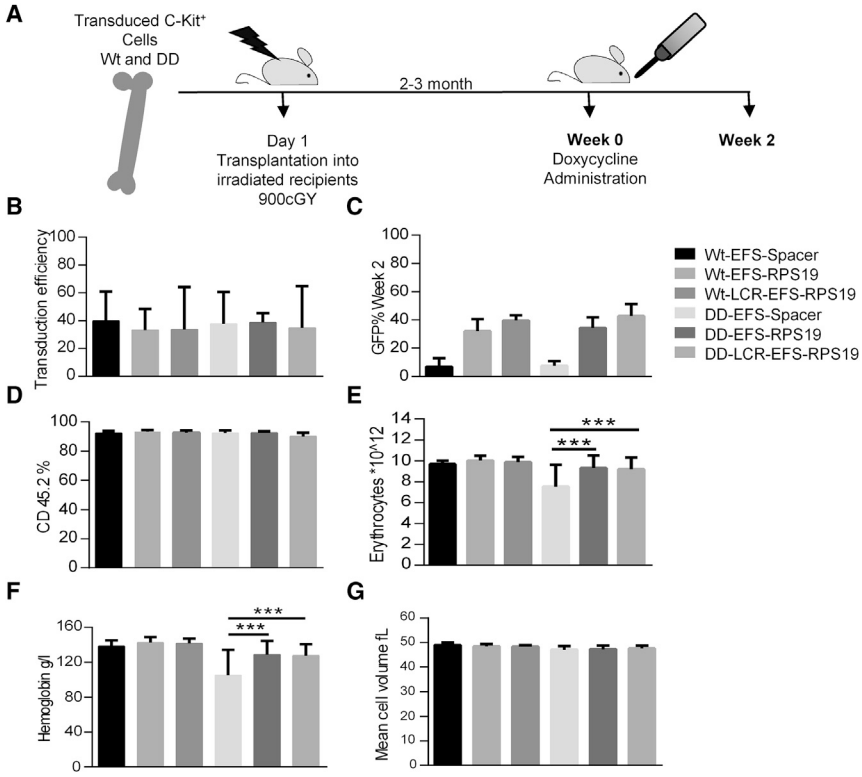


Figure 3. Enforced Expression of RPS19 Derived from the EFS Promoter Is Sufficient to Rescue the Acute DBA Phenotype In Vivo

Enforced expression of *RPS19* results in short-term rescue of the hematological defect in RPS19-deficient mice. (A) Experimental strategy to validate the short-term therapeutic potential of EFS-RPS19 and LCR-EFS-RPS19 vectors. (B) Transduction efficiency is shown. (C and D) GFP reconstitution and donor reconstitution are shown. (E–G) Erythrocyte number (E), hemoglobin concentration (F), and mean corpuscular value (G) (n = 20–21). Error bars represent the SD. ***p < 0.001. MCV, mean corpuscular value; Wt, wild-type.

the EFS-RPS19 vector integration profile as well as the clonal dynamics of the transduced cells, we performed insertion site analysis of DNA from BM cells of four mice per vector group obtained from recipients after 16–18 weeks of doxycycline administration. Integration sites per vector group (Wt-EFS-Spacer [WES], Wt-.EFS-RPS19 [WER], Wt-LCR.-EFS-RPS19 [WLER], DD-EFS-RPS19 [DER], and DD-LCR-EFS.RPS19 [DLER]) were analyzed by linear amplification-mediated (LAM) PCR followed by Ion Torrent sequencing. A total of 2.88×10^6 sequences were processed, clustered for homology (increasing the read count of individual insertions), trimmed for the remaining vector sequences, and aligned to the murine genome. The 2.18×10^5 sequence reads were assigned to 5,420 individual insertions. Despite the known limitations in terms of absolute quantification of amplicon sequencing in integration site analysis,[30] we use the read count as a surrogate marker for clonal abundance. We investigated the insertion profile in the different mice for the number of hits close to transcriptional start sites (TSSs) of genes, the clonal diversity,[31] common insertion sites (CISs), and overlaps with cancer gene databases. Detailed information for each mouse is provided in Table S1 and Figures S1–S3. We did not observe a tendency to preferentially integrate within a 10-kb window around the TSSs of genes (Figure 7A). The overlap of EFS-RP19 insertions with the Retroviral Tagged Cancer Gene Database (RTCGD)[32] or the All Onco cancer gene list[33] was not different from a randomized control dataset (Figure 7B). We did not observe a significant difference in clonal diversity between the vector groups. However, six mice had a lower sequence diversity (Figure S2F) compared to all other treated animals. For two of these mice (DER1 and DLER3), we observed a dominant insertion within genes (Malt1 and Cdh26) listed in the RTCGD database. Both genes were found only once in an artificial B-cell lymphoma mouse model during insertional mutagenesis screens.[34] From the overlap of gene symbols close to insertion sites and cancer gene databases alone, we cannot conclude a functional relationship exists between vector integration and increased clonal abundance. As we also cannot exclude a proliferation advantage due to insertional mutagenesis, we depict overlaps with four reference databases for those insertions with a read count above the 97.5 percentile of all reads (Table S2) and for all detected CISs (Table S3). A chi-square analysis revealed no statistical differences for the overlap with cancer gene databases between the vector groups. When we checked for common high-risk insertions in or near Prdm16, Mecom, Notch1, Lmo2, Setbp1, Ccnd2, Sox4, and Tal1, we found either no hits (Lmo2, Tal1) or only read contributions $\leq$ 0.58% (n = 19 of 5,420 sequences).

DISCUSSION

Here, we demonstrate the efficacy of RPS19 lentiviral vectors using clinically relevant promoters to correct lethal BM failure in Rps19-deficient mice. We show that the EFS promoter can express enough RPS19 to correct RPS19-deficient BM failure and the EFS-driven RPS19 single-gene vector can be used in a clinical gene therapy trial for RPS19-deficient DBA. Previously, we demonstrated that enforced expression of RPS19 improves the proliferation, erythroid colony-forming potential, and differentiation of patient-derived RPS19-deficient hematopoietic progenitor cells in vitro.[19,20] Using xenograft models, we have also shown that overexpression of RPS19 enhances the engraftment and erythroid differentiation of patient-derived hematopoietic stem and progenitor cells.[21] In our proof-of-principle study using lentiviral vectors driven by the SFFV promoter and harboring a codon-optimized human RPS19 cDNA followed by IRES and GFP, we showed that the DBA phenotype of mice can be successfully treated.[22]

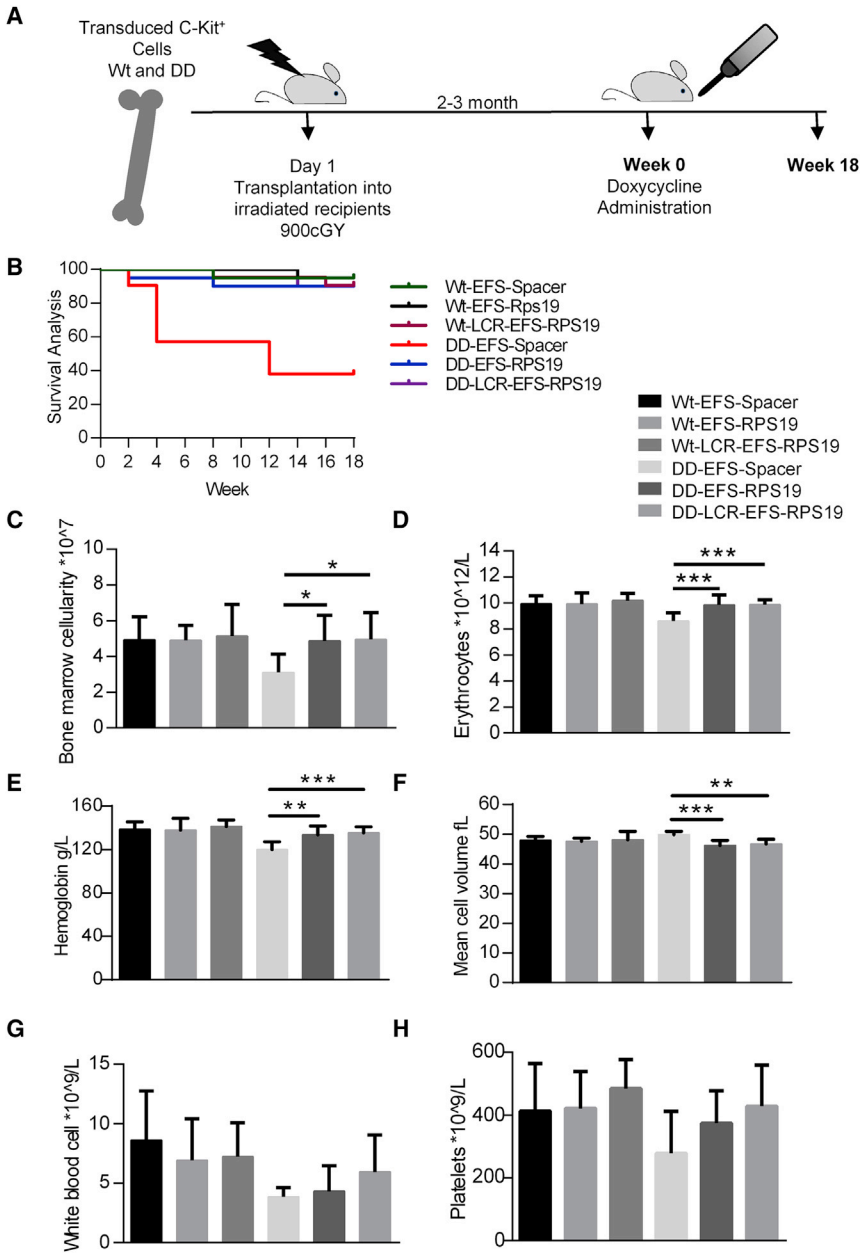


Figure 4. Enforced Expression of RPS19 Derived from the EFS Promoter Is Sufficient to Rescue the DBA Phenotype In Vivo

Enforced expression of *RPS19* results in long-term rescue of the hematological defect in RPS19-deficient mice. (A) Experimental strategy to validate the long-term therapeutic potential of EFS-RPS19 and LCR-EFS-RPS19 vectors. (B) Survival curve. (C) BM cellularity after 18 weeks of doxycycline induction. (D–H) Erythrocyte number (D), hemoglobin concentration (E), mean corpuscular value (F), white blood cell count (G), and platelet number (H) after 18 weeks of doxycycline induction (n = 20–21). Error bars represent the SD. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. MCV, mean corpuscular value; Wt, wild-type.

In the current study, we decided to utilize the ubiquitously expressed EFS promoter with or without the LCR of the β-globin gene for treatment of RPS19-deficient DBA. We have shown that these vectors rescue the proliferation defect and erythroid development of transduced c-Kit+ DD BM cells in vitro. The induction of Rps19 deficiency in recipient mice with DD BM generated lethal BM failure. Remarkably, the BM failure generated by DD BM was cured with EFS-RPS19. Since quite high levels of RPS19 are needed to correct the RPS19 deficiency by transgenesis, we were concerned that the EFS promoter might not generate sufficient levels of RPS19 in erythroid progenitors to correct the anemia. Therefore, we included vectors containing the EFS plus the β-globin LCR. However, the findings show that the EFS promoter without the β-globin LCR generates sufficient levels of RPS19 to cure anemia and BM failure in RPS19-deficient mice.

Additionally, we demonstrated that RPS19-deficient BM cells can be transduced and these cells survived the transduction procedure and had the capacity to repopulate the BM. However, most of the studies were performed with transduced shRNA-D/D BM transplanted into normal recipients. RPS19 deficiency was induced once the recipients had a stable graft. This is a justified, since we have previously shown that anemia and BM failure in the induced mice is due to the deficiency in the hematopoietic cells and not to a failure of the niche cells.[28] If the recipients have Rps19 deficiency in all cells before transplantation of the transduced cells, some of the Rps19-deficient mice will not tolerate the combined toxicity of doxycycline Rps19 downregulation and radiation. However, the majority of the Rps19-deficient mice survived this procedure, as mentioned above.

In this study, we have shown that our RPS19-deficient mouse model is valuable and suitable for testing gene therapy using viral vectors with the RPS19 gene. However, it should be emphasized that this model is different from haploinsufficiency in patients with DBA, which is based on mutations in the RPS19 gene (most often point mutations or small deletions). In the mice used here, haploinsufficiency is generated by RNAi that is induced postnatally. Haploinsufficiency generates most of the hematological symptoms found in DBA in mice but not the physical abnormalities found in a large fraction of patients. Haploinsufficiency in mice causes reduced proliferation and erythroid development, which can be corrected by overexpression of RPS19. A similar effect was seen in cells from patients with RPS19-deficient DBA. Upon

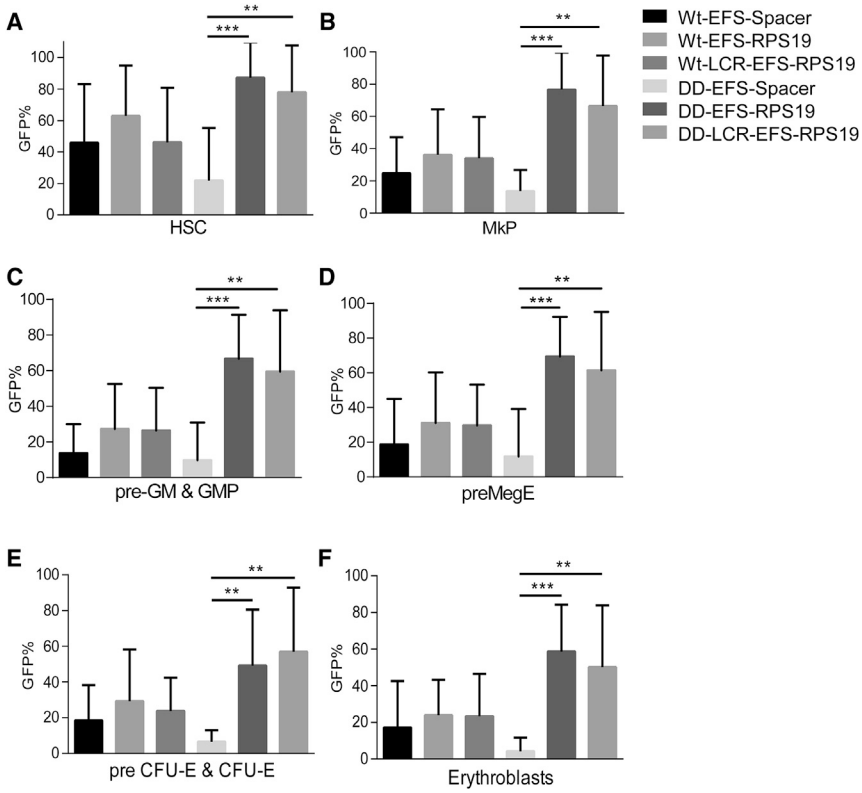


Figure 5. Gene-Corrected Rps19-Deficient Cells Gain a Competitive Advantage Resulting in Increased Contribution to Hematopoiesis In Vivo

(A–F) The percentage of transduced cells in the HSC (A), MkP (B), pre-GM (granulocyte macrophage) and granulocyte macrophage progenitors (GMP) (C), preMegE (D), and preCFU-E (E) and colony forming unit (CFU)-E erythroblast compartment (F) (n = 16–24 per group). Error bars represent the SD. *p < 0.05; **p < 0.01; ***p < 0.001. Wt, wild-type.

overexpression of RPS19 in the cells from patients, cellular proliferation and erythroid development were greatly improved.[19,20]

Of course, it is clear that RPS19 vectors can only be used to treat patients with RPS19-deficient DBA. Therefore, patients with mutations in other ribosomal protein genes or the GATA1 gene cannot be treated with RPS19 vectors. Recently, mutations in GATA1 were found in a few patients with DBA. The GATA1 gene in humans produces two mRNAs, one long and one short. Patients with DBA could not produce the long form of GATA1.[12,13] Mice produce only the long form of GATA1; therefore, it will be difficult to evaluate the possibility of using mice as experimental animals in the development of GATA1 gene therapy for human GATA1-deficient patients with DBA.

The data presented in Figure 5 show that the RPS19 vectors increase the production of hematopoietic stem cells (HSCs) and early progenitor cells after overexpression in RPS19-deficient hematopoietic cells. In competitive transplantation experiments, we showed previously that RPS19-deficient HSCs have a competitive disadvantage compared to normal HSCs.[28] Collectively, these data suggest that RPS19-deficient HSCs treated with RPS19 vectors may have a competitive advantage compared with untreated cells. It is therefore possible that gene therapy of RPS19-deficient DBA may be performed with little or no BM ablation before transplantation of the gene-corrected cells, due to the possible competitive advantage of these cells. However, the need for ablation in a clinical gene therapy setting must be investigated further in order to design clinical gene therapy trials with minimal risks for patients.

Significantly, by designing a codon-optimized RPS19 cDNA driven by the EFS promoter, we have succeeded in generating a clinically relevant vector system that allows high enough RPS19 expression for functional correction of anemia and BM failure in Rps19-deficient mice. Our studies assessing the efficacy of clinically relevant EFS promoters show a less likely risk of causing insertional oncogenesis.[23] Further, our studies using EFS-RPS19 or LCR.EFS-RPS19 vectors that are safer than vectors with viral promoters, but can nevertheless generate sufficient RPS19 expression to correct the pathophysiology of DBA. In normal cells, ribosomal protein production is tightly regulated and physiological and excess protein is subjected to proteasomal degradation.[35] Because of this mechanistic regulation of ribosomal protein, ectopic expression of *RPS19* possesses a very low risk of promoting uncontrolled growth. In our study, we did not observe any hematologic abnormalities due to enforced expression of *RPS19*. Our results collectively demonstrate the feasibility of clinical gene therapy to cure RPS19-deficient patients with DBA in the future using EFS promoter-driven enforced expression of RPS19.

MATERIALS AND METHODS

Lentiviral Vector Constructs

SIN lentiviral vectors used in this study were derived from pRRL.PPT.PGK.GFP pre-vector.[36] A codon-optimized human RPS19 cDNA was designed and inserted downstream of the EFS promoter with or without the LCR of the β-globin gene (LCR).[26,27,36] Following the RPS19 cDNA, the IRES, GFP, and improved post-transcriptional regulatory element (pre*) were inserted. Two vectors were obtained without the LCR pRRL.PPT.EFS.RPS19co.iresGFP.pre* vector (hereafter, EFS-RPS19) and with the LCR pRRL.PPT.LCR.EFS.RPS19co.iresGFP.pre* vector (hereafter, LCR-EFS-RPS19). Lentiviral vectors were produced by the Vector Unit at Lund University. Briefly, standard calcium phosphate transfection of 293T cells was used with the helper plasmid cytomegalovirus (pCMV)ΔR8.91 and pMD2.G—VSV-G envelop expressing plasmid (pMDG). The lentivirus-containing supernatant was harvested 24 hr after transfection, and the lentivirus was

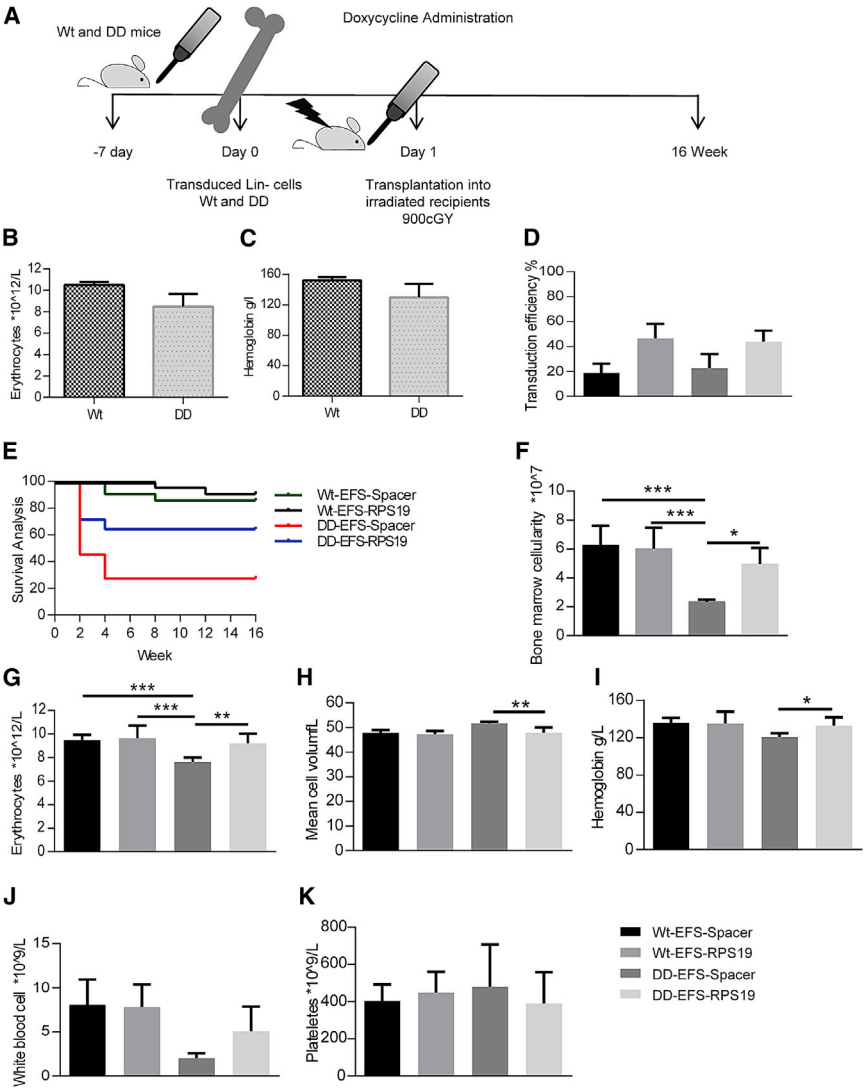


Figure 6. Rps19-Deficient BM Cells Can Be Transduced and the Transduced Cells Provide Long-Term Reconstitution

Rps19-deficient BM can be transduced; after genetic correction, these cells show long-term engraftment in lethally irradiated wild-type mice. (A) Experimental strategy to validate the long-term reconstitution capacity of corrected Rps19-deficient cells. (B–F) Pre-transplant Wt and DD mice erythrocyte numbers (B), hemoglobin concentration (C), transduction efficiency (D), survival curve (E), and BM cellularity (F) after 16 weeks of doxycycline induction. (G–K) Erythrocyte number (G), mean corpuscular volume (H), hemoglobin concentration (I), white blood cell count (J), and platelet number (K) after 16 weeks of doxycycline induction (n = 20–28). Error bars represent the SD. *p < 0.05; **p < 0.01; ***p < 0.001. MCV, mean corpuscular volume; Wt, wild-type.

concentrated by ultracentrifugation at 25,000 rpm (SW32 rotor, Beckman L-70 Ultracentrifuge; Beckman Coulter) for 90 min at 4°C. Pellets were resuspended in serum-free medium (StemSpan SFEM; Stemcell Technologies) and stored at −80°C. Lentivirus titer was assessed by FACS for the transfer of GFP to HT1080 cells.

Mice

The mouse models are engineered to contain a doxycycline-regulatable Rps19-targeting shRNA (shRNA-D) located downstream of the collagen A1 locus, allowing dose-dependent downregulation of Rps19 expression.[28] To generate two models with intermediate or severe Rps19 deficiency, transgenic animals were bred either heterozygous or homozygous for shRNA-D, respectively. RPS19 deficiency was induced by feeding the mice doxycycline in drinking water (1 mg/mL or 2 mg/mL doxycycline; Sigma-Aldrich) supplemented with 10 mg/mL sucrose (Sigma-Aldrich). Mice were maintained at the Lund University animal facility and all animal experiments were performed with consent from the Lund University animal ethics committee.

Blood and BM Analysis

Peripheral blood was collected from the tail vein into microvette tubes (Sarstedt) and was analyzed using sysmex XE-5000. Erythrocytes were lysed using ammonium chloride for 10 min at room temperature. To evaluate the contribution toward various blood lineages following BM transplantation, samples were stained with the following antibodies for 30 min on ice in the dark: CD45.1 (110730; Biolegend) CD45.2 (47-0454-82; eBioscience), B220 (103208; Biolegend), B220 (103212; Biolegend), CD3 (100312; Biolegend), CD11b (101208; Biolegend), and Gr1 (108408; Biolegend). Experiments were performed using a FACS Canto II cytometer (BD Biosciences) and were analyzed by FlowJo software (version 10.0.2; Tree Star). FACS analysis of the myeloerythroid compartment in BM was performed.[28,29] BM cells were isolated by crushing the femur and tibia in PBS containing 2% fetal bovine serum (FBS) (Gibco). Fresh cells were stained with the following antibodies: CD41 (12-0411-83; eBioscience), GR1 (115910; Biolegend), CD11b (101210; Biolegend), B220 (103210; Biolegend), CD3 (100310; Biolegend), c-Kit (47-1171-82; eBioscience), CD105 (120404; Biolegend), and Sca-1 (122520; Biolegend). Streptavidin was purchased from Life Technologies (Q10101MP). Propidium iodide (Life Technologies) was used to exclude dead cells. Experiments were performed using a FACS LSR II cytometer (Becton Dickinson) and were analyzed by FlowJo software (version 10.0.2).

Transduction and Transplantation of Hematopoietic Cells

c-Kit+-expressing cells were enriched from BM of transgenic mice (CD45.2) using CD117 microbeads and a magnetic-activated cell sorting (MACS) separation column (Miltenyi) and were pre-stimulated in

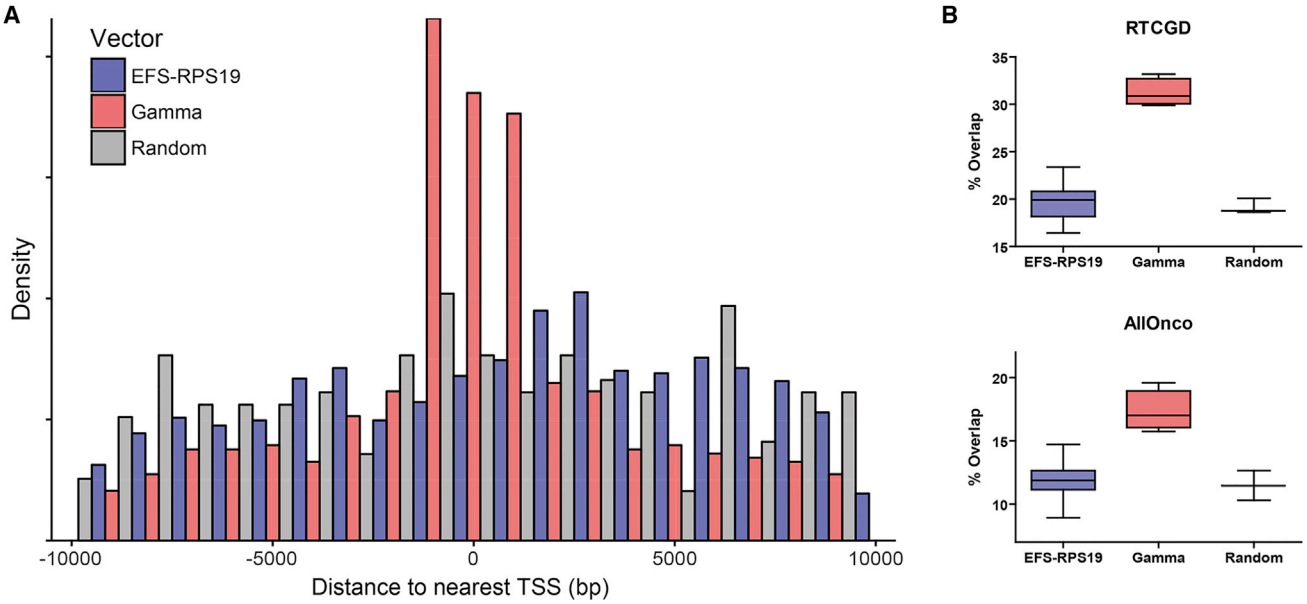


Figure 7. EFS-RPS19 Integrations Do Not Cluster around the TSS and Show No Increased Overlap with Cancer Gene Databases

(A) Density plot showing the frequency of integrations in a 10-kb window around the transcriptional start site. As we found no statistical differences between the different EFS-RPS19 vectors, we combined them in one group (blue) and compared them to a gamma retroviral integration profile (red) or a randomized dataset (gray). (B) The overlap of gene symbols closest to the insertion sites with either the Retroviral Tagged Cancer Gene Database (RTCGD) or the AllOnco cancer gene list of the EFS-RPS19 vectors was not different from that of a randomized dataset. The increased overlap of the gamma retroviral integration dataset is shown for comparison.

serum-free StemSpan serum-free expansion medium (SFEM) (Stemcell Technologies), supplemented with penicillin/streptomycin (P/S; Gibco), murine stem cell factor (mSCF) (100 ng/mL; PeproTech),and human thrombopoietin (hTPO; 50 ng/mL; PeproTech) in six-well plates (non-tissue culture treated; BD) for 1 day at 0.5×10^6 cells/mL. Retronectin-coated (20 ng/mL; Takara) six-well plates were preloaded with the viral vectors (100 μL/well corresponding to a MOI of 10–20) and 1 million cells were seeded into each well in 3 mL pre-stimulation medium. After incubation for 1 day, 0.5×10^6 bulk transduced cells were transplanted in 500 μL PBS into the tail vein of irradiated (900 cGy) wild-type mice recipients (CD45.1 or CD45.1/45.2).

Transduction and Transplantation of RPS19-Deficient Hematopoietic Cells

Lineage-negative (Lin−) cells were enriched from the BM of doxycycline-induced transgenic mice (CD45.2) using Lineage microbeads and a MACS separation column (Miltenyi). Retronectin-coated (20 ng/mL; Takara) six-well plates were preloaded with the viral vectors (100 μL/well corresponding to a MOI of 10–20) and 1 million cells were seeded into each well in 3 mL serum-free StemSpan SFEM medium supplemented with P/S, mSCF (100 ng/mL), hTPO (50 ng/mL), and doxycycline (1 μg/mL; Sigma-Aldrich) in six-well plates (non-tissue culture treated; BD) for 1 day 0.5×10^6 cells/mL. After incubation for 1 day, 0.5×10^6 bulk transduced cells and 1×10^6 untransduced Lin+ cells were transplanted in 300 μL PBS into the tail vein of lethally irradiated (900 cGy) wild-type mice recipients (CD45.1 or CD45.1/45.2).

Cell Culture

c-Kit+-expressing cells were enriched using CD117 microbeads and a MACS separation column (Miltenyi) and retronectin-coated (20 ng/mL; Takara), pre-stimulated in serum-free StemSpan SFEM medium supplemented with P/S, mSCF (100 ng/mL), and hTPO (50 ng/mL) in six-well plates (non-tissue culture treated; BD) for 1 day 0.5×10^6 cells/mL. Twelve-well plates were preloaded with the viral vectors (50 μL/well corresponding to a MOI of 10–20) and 0.5×10^6 cells were seeded into each well in 1 mL serum-free StemSpan SFEM medium supplemented with P/S, mSCF (100 ng/mL), murine interleukin (IL)-3 (mIL-3; 10 ng/mL; PeproTech), and erythropoietin (EPO; 2 U/mL; Janssen-Cilag) with or without doxycycline (1 μg/mL). Light microscopy was used to evaluate the proliferation of culture after 6 days. For the burst forming unit-erythroid (BFU-E) assay, 40×10^3 c-Kit+ transduced cells were seeded in 1.5 mL M3436 methylcellulose (Stemcell Technologies) with doxycycline (1 μg/mL) and colonies were scored on day 14.

Insertion Site Analysis

We used 300 ng genomic DNA of whole BM cells isolated 18 weeks after transplantation. Samples were processed by LAM PCR as described by Schmidt et al.[37] with modifications. For digestion, samples were split into three separate reactions with 5 U of CutSmart enzymes MlucI, MseI, and HindPI (the latter two with heat inactivation) from New England Biolabs (NEB). After digestion, samples were combined for nested PCR steps. The first nested PCR was performed with a forward primer binding to the SIN-LTR of the vectors

(IT-IS-FW-PCR1: 5′-*GTGGGTTTTCCAGTCACAC*TGCTCTTCCG ATCTTCCCTCAGACCCTTTTAGTCA-3′) and a reverse primer recognizing the linker cassette (IT-IS-RV-PCR1: 5′-*TTCGTTGGG AGTGAATTAGCC* AGTGGCACAGCAGTTAGG-3′). The vector- and linker-specific sequences are underlined, and the italic sequence represents a tail homologous to the primers used for Ion Torrent sequencing, as described previously.[38] Bioinformatics processing with custom Perl, R, and visual basic scripts involved barcode primer assignment, trimming, clustering, filtering, and Methods for Analyzing ViRal Integration Clusters analysis tool (MAVRIC) alignment.[39] For CIS analysis, we followed the suggestions by Wu et al.,[40] considering only five or more insertions in a 50-kb window. Distance to the TSSs was analyzed using the information from the MAVRIC alignments in combination with a customized R script (ggplot2; geom_histogram with the following parameters: aes y = density and binwidth = 1,000).

Control Datasets for Integration Site Analysis

The gamma retroviral integrations used for comparison originate from Lin− cell cultures (n = 4) transduced with RSF91.[41,42] DNA was harvested 4 days after transduction. The LAM-PCR procedure and next-generation sequencing were as described above. The randomized control datasets (n = 3) were produced by generating artificial chromosomal positions using the shuffle command (seed = 100, 101, and 102) of BEDtools.[43] The shuffled BED files contained 2,000 genomic positions (500-bp window size) randomly distributed among the murine genome (NCBI47/mm9) as a function of the chromosome size. The BED files were converted to FASTA format and processed by MAVRIC with parameters identical to the biological insertion site data of EFS-RPS19 or the gamma retroviral vector.

Statistical Analysis

One-way ANOVA with the Tukey multiple-comparison test was used to determine statistical significance using GraphPad Prism (version 6; GraphPad Software).

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.ymthe.2017.04.002.

AUTHOR CONTRIBUTIONS

S.K. conceptualized the project and directed the research; S.D., K.S., M.R., J.C., and M.D. performed the experiments; S.D., P.J., M.R., H.G.B., J.F., A.S., and S.K. analyzed the data; and S.D., P.J., M.R., A.S., and S.K. wrote the manuscript.

CONFLICTS OF INTEREST

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

The authors thank Beata Lindqvist for lentivirus production and Amol Ugale, Karolina Komorowska, and Abdul Ghani Alattar for technical assistance. This work was supported by a Hemato-Linne grant from the Swedish Research Council Linnaeus, grants from the Swedish Cancer Society and the Swedish Children's Cancer Society (to S.K.), the Tobias Prize awarded by the Royal Swedish Academy of Sciences financed by the Tobias Foundation, a clinical research grant from Lund University Hospital (to S.K.), and European Union (EU) project grants STEMEXPAND and PERSIST. J.F. was supported by the Diamond Blackfan Anemia Foundation. H.B.G. is supported by the Great Ormond Street Hospital Children's Charity and by the National Institute of Health Research Biomedical Research Centre at Great Ormond Street Hospital and University College London. A.S. and M.R. are supported by the Deutsche Forschungsgemeinschaft (DFG) (REBIRTH Cluster of Excellence and SFB738 projects).

REFERENCES

1. Willig, T.N., Niemeyer, C.M., Leblanc, T., Tiemann, C., Robert, A., Budde, J., Lambiliotte, A., Kohne, E., Souillet, G., Eber, S., et al. (1999). Identification of new prognosis factors from the clinical and epidemiologic analysis of a registry of 229 Diamond-Blackfan anemia patients. DBA group of Société d'Hématologie et d'Immunologie Pédiatrique (SHIP), Gesellshaft für Pädiatrische Onkologie und Hämatologie (GPOH), and the European Society for Pediatric Hematology and Immunology (ESPHI). Pediatr. Res. *46*, 553–561.

2. Orfali, K.A., Ohene-Abuakwa, Y., and Ball, S.E. (2004). Diamond Blackfan anaemia in the UK: clinical and genetic heterogeneity. Br. J. Haematol. *125*, 243–252.

3. Lipton, J.M., Atsidaftos, E., Zyskind, I., and Vlachos, A. (2006). Improving clinical care and elucidating the pathophysiology of Diamond Blackfan anemia: an update from the Diamond Blackfan Anemia Registry. Pediatr. Blood Cancer *46*, 558–564.

4. Draptchinskaia, N., Gustavsson, P., Andersson, B., Pettersson, M., Willig, T.N., Dianzani, I., Ball, S., Tchernia, G., Klar, J., Matsson, H., et al. (1999). The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. Nat. Genet. *21*, 169–175.

5. Gazda, H.T., Grabowska, A., Merida-Long, L.B., Latawiec, E., Schneider, H.E., Lipton, J.M., Vlachos, A., Atsidaftos, E., Ball, S.E., Orfali, K.A., et al. (2006). Ribosomal protein S24 gene is mutated in Diamond-Blackfan anemia. Am. J. Hum. Genet. *79*, 1110–1118.

6. Cmejla, R., Cmejlova, J., Handrkova, H., Petrak, J., and Pospisilova, D. (2007). Ribosomal protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. Hum. Mutat. *28*, 1178–1182.

7. Farrar, J.E., Nater, M., Caywood, E., McDevitt, M.A., Kowalski, J., Takemoto, C.M., Talbot, C.C., Jr., Meltzer, P., Esposito, D., Beggs, A.H., et al. (2008). Abnormalities of the large ribosomal subunit protein, Rpl35a, in Diamond-Blackfan anemia. Blood *112*, 1582–1592.

8. Gazda, H.T., Sheen, M.R., Vlachos, A., Choesmel, V., O'Donohue, M.F., Schneider, H., Darras, N., Hasman, C., Sieff, C.A., Newburger, P.E., et al. (2008). Ribosomal protein L5 and L11 mutations are associated with cleft palate and abnormal thumbs in Diamond-Blackfan anemia patients. Am. J. Hum. Genet. *83*, 769–780.

9. Doherty, L., Sheen, M.R., Vlachos, A., Choesmel, V., O'Donohue, M.F., Clinton, C., Schneider, H.E., Sieff, C.A., Newburger, P.E., Ball, S.E., et al. (2010). Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. Am. J. Hum. Genet. *86*, 222–228.

10. Gazda, H.T., Preti, M., Sheen, M.R., O'Donohue, M.F., Vlachos, A., Davies, S.M., Kattamis, A., Doherty, L., Landowski, M., Buros, C., et al. (2012). Frameshift mutation in p53 regulator RPL26 is associated with multiple physical abnormalities and a specific pre-ribosomal RNA processing defect in Diamond-Blackfan anemia. Hum. Mutat. *33*, 1037–1044.

11. Landowski, M., O'Donohue, M.F., Buros, C., Ghazvinian, R., Montel-Lehry, N., Vlachos, A., Sieff, C.A., Newburger, P.E., Niewiadomska, E., Matysiak, M., et al. (2013). Novel deletion of RPL15 identified by array-comparative genomic hybridization in Diamond-Blackfan anemia. Hum. Genet. *132*, 1265–1274.

12. Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., and Gazda, H.T. (2012). Exome

sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. J. Clin. Invest. *122*, 2439–2443.

13. Ludwig, L.S., Gazda, H.T., Eng, J.C., Eichhorn, S.W., Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., et al. (2014). Altered translation of GATA1 in Diamond-Blackfan anemia. Nat. Med. *20*, 748–753.
14. Klar, J., Khalfallah, A., Arzoo, P.S., Gazda, H.T., and Dahl, N. (2014). Recurrent GATA1 mutations in Diamond-Blackfan anaemia. Br. J. Haematol. *166*, 949–951.
15. Gripp, K.W., Curry, C., Olney, A.H., Sandoval, C., Fisher, J., Chong, J.X., Pilchman, L., Sahraoui, R., Stabley, D.L., and Sol-Church, K.; UW Center for Mendelian Genomics (2014). Diamond-Blackfan anemia with mandibulofacial dystostosis is heterogeneous, including the novel DBA genes TSR2 and RPS28. Am. J. Med. Genet. A. *164A*, 2240–2249.
16. Willig, T.N., Draptchinskaia, N., Dianzani, I., Ball, S., Niemeyer, C., Ramenghi, U., Orfali, K., Gustavsson, P., Garelli, E., Brusco, A., et al. (1999). Mutations in ribosomal protein S19 gene and Diamond Blackfan anemia: wide variations in phenotypic expression. Blood *94*, 4294–4306.
17. Angelini, M., Cannata, S., Mercaldo, V., Gibello, L., Santoro, C., Dianzani, I., and Loreni, F. (2007). Missense mutations associated with Diamond-Blackfan anemia affect the assembly of ribosomal protein S19 into the ribosome. Hum. Mol. Genet. *16*, 1720–1727.
18. Vlachos, A., Federman, N., Reyes-Haley, C., Abramson, J., and Lipton, J.M. (2001). Hematopoietic stem cell transplantation for Diamond Blackfan anemia: a report from the Diamond Blackfan Anemia Registry. Bone Marrow Transplant. *27*, 381–386.
19. Hamaguchi, I., Ooka, A., Brun, A., Richter, J., Dahl, N., and Karlsson, S. (2002). Gene transfer improves erythroid development in ribosomal protein S19-deficient Diamond-Blackfan anemia. Blood *100*, 2724–2731.
20. Hamaguchi, I., Flygare, J., Nishiura, H., Brun, A.C., Ooka, A., Kiefer, T., Ma, Z., Dahl, N., Richter, J., and Karlsson, S. (2003). Proliferation deficiency of multipotent hematopoietic progenitors in ribosomal protein S19 (RPS19)-deficient diamond-Blackfan anemia improves following RPS19 gene transfer. Mol. Ther. *7*, 613–622.
21. Flygare, J., Olsson, K., Richter, J., and Karlsson, S. (2008). Gene therapy of Diamond Blackfan anemia CD34(+) cells leads to improved erythroid development and engraftment following transplantation. Exp. Hematol. *36*, 1428–1435.
22. Jaako, P., Debnath, S., Olsson, K., Modlich, U., Rothe, M., Schambach, A., Flygare, J., and Karlsson, S. (2014). Gene therapy cures the anemia and lethal bone marrow failure in a mouse model of RPS19-deficient Diamond-Blackfan anemia. Haematologica *99*, 1792–1798.
23. 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. Mol. Ther. *16*, 718–725.
24. Zhang, F., Frost, A.R., Blundell, M.P., Bales, O., Antoniou, M.N., and Thrasher, A.J. (2010). A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. Mol. Ther. *18*, 1640–1649.
25. Hacein-Bey-Abina, S., Pai, S.Y., Gaspar, H.B., Armant, M., Berry, C.C., Blanche, S., Bleesing, J., Blondeau, J., de Boer, H., Buckland, K.F., et al. (2014). A modified γ-retrovirus vector for X-linked severe combined immunodeficiency. N. Engl. J. Med. *371*, 1407–1417.
26. Coci, E.G., Maetzig, T., Zychlinski, D., Rothe, M., Suerth, J.D., Klein, C., and Schambach, A. (2015). Novel self-inactivating vectors for reconstitution of Wiskott-Aldrich syndrome. Curr. Gene Ther. *15*, 245–254.
27. Montiel-Equihua, C.A., Zhang, L., Knight, S., Saadeh, H., Scholz, S., Carmo, M., Alonso-Ferrero, M.E., Blundell, M.P., Monkeviciute, A., Schulz, R., et al. (2012). The β-globin locus control region in combination with the EF1α short promoter allows enhanced lentiviral vector-mediated erythroid gene expression with conserved multilineage activity. Mol. Ther. *20*, 1400–1409.
28. Jaako, P., Flygare, J., Olsson, K., Quere, R., Ehinger, M., Henson, A., Ellis, S., Schambach, A., Baum, C., Richter, J., et al. (2011). Mice with ribosomal protein S19 deficiency develop bone marrow failure and symptoms like patients with Diamond-Blackfan anemia. Blood *118*, 6087–6096.
29. Jaako, P., Debnath, S., Olsson, K., Bryder, D., Flygare, J., and Karlsson, S. (2012). Dietary L-leucine improves the anemia in a mouse model for Diamond-Blackfan anemia. Blood *120*, 2225–2228.
30. Brugman, M.H., Suerth, J.D., Rothe, M., Suerbaum, S., Schambach, A., Modlich, U., Kustikova, O., and Baum, C. (2013). Evaluating a ligation-mediated PCR and pyrosequencing method for the detection of clonal contribution in polyclonal retrovirally transduced samples. Hum. Gene Ther. Methods *24*, 68–79.
31. Shannon, C.E. (1963). The mathematical theory of communication. MD Comput. *4*, 306–317.
32. Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A., and Copeland, N.G. (2004). RTCGD: Retroviral Tagged Cancer Gene Database. Nucleic Acids Res. *32*, D523–D527.
33. Sadelain, M., Papapetrou, E.P., and Bushman, F.D. (2011). Safe harbours for the integration of new DNA in the human genome. Nat. Rev. Cancer *12*, 51–58.
34. Bijl, J., Sauvageau, M., Thompson, A., and Sauvageau, G. (2005). High incidence of proviral integrations in the Hoxa locus in a new model of E2a-PBX1-induced B-cell leukemia. Genes Dev. *19*, 224–233.
35. Lam, Y.W., Lamond, A.I., Mann, M., and Andersen, J.S. (2007). Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. Curr. Biol. *17*, 749–760.
36. Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463–8471.
37. Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H., and von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat. Methods *4*, 1051–1057.
38. Selich, A., Daudert, J., Hass, R., Philipp, F., von Kaisenberg, C., Paul, G., Cornils, K., Fehse, B., Rittinghausen, S., Schambach, A., and Rothe, M. (2016). Massive clonal selection and transiently contributing clones during expansion of mesenchymal stem cell cultures revealed by lentiviral RGB-barcode technology. Stem Cells Transl. Med. *5*, 591–601.
39. Rittelmeyer, I., Rothe, M., Brugman, M.H., Iken, M., Schambach, A., Manns, M.P., Baum, C., Modlich, U., and Ott, M. (2013). Hepatic lentiviral gene transfer is associated with clonal selection, but not with tumor formation in serially transplanted rodents. Hepatology *58*, 397–408.
40. Wu, X., Luke, B.T., and Burgess, S.M. (2006). Redefining the common insertion site. Virology *344*, 292–295.
41. Hildinger, M., Abel, K.L., Ostertag, W., and Baum, C. (1999). Design of 5′ untranslated sequences in retroviral vectors developed for medical use. J. Virol. *73*, 4083–4089.
42. Schambach, A., Wodrich, H., Hildinger, M., Bohne, J., Kräusslich, H.G., and Baum, C. (2000). Context dependence of different modules for posttranscriptional enhancement of gene expression from retroviral vectors. Mol. Ther. *2*, 435–445.
43. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

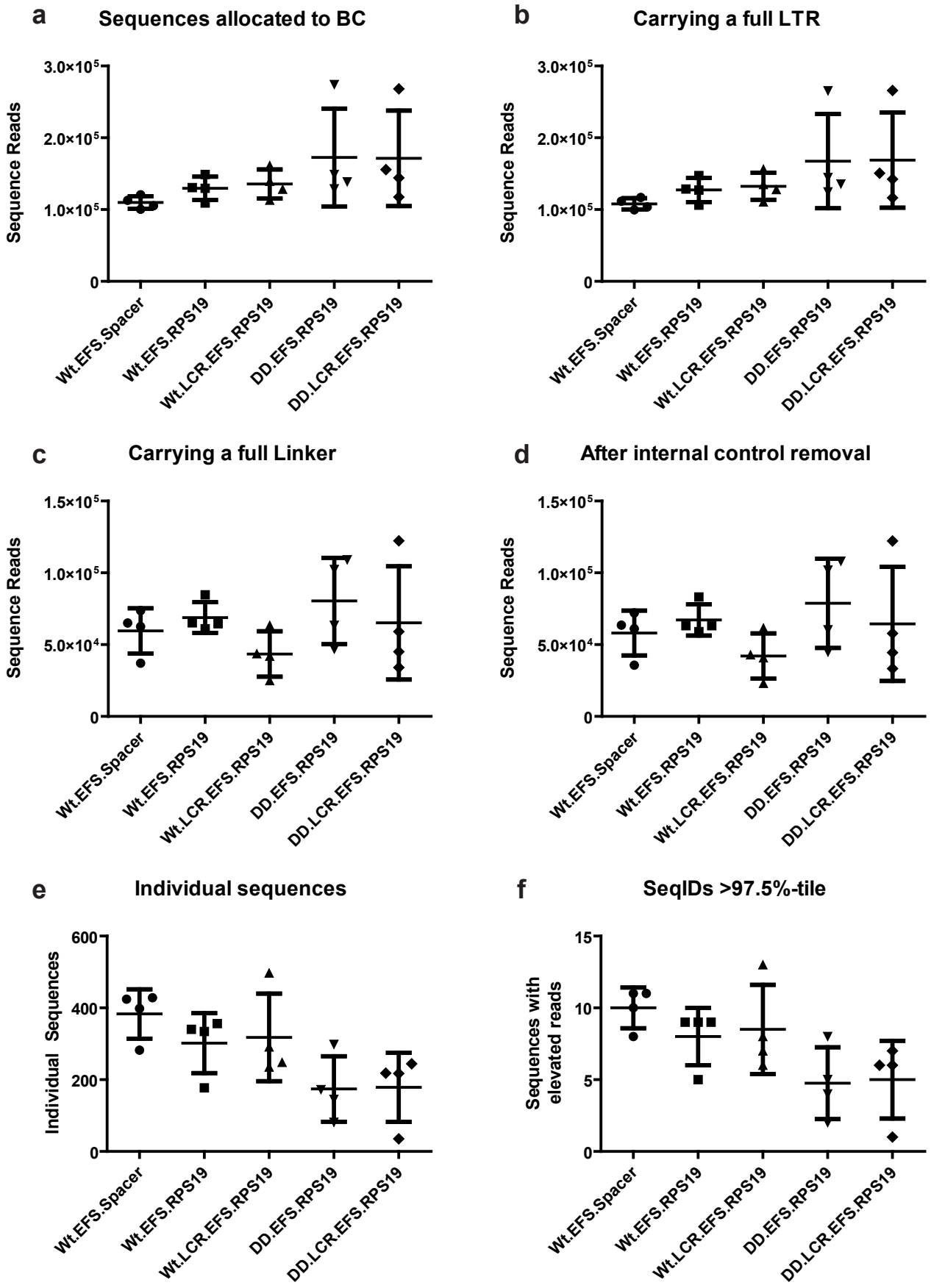
YMTHE, Volume *25*

Supplemental Information

Lentiviral Vectors with Cellular Promoters Correct Anemia and Lethal Bone Marrow Failure in a Mouse Model for Diamond-Blackfan Anemia

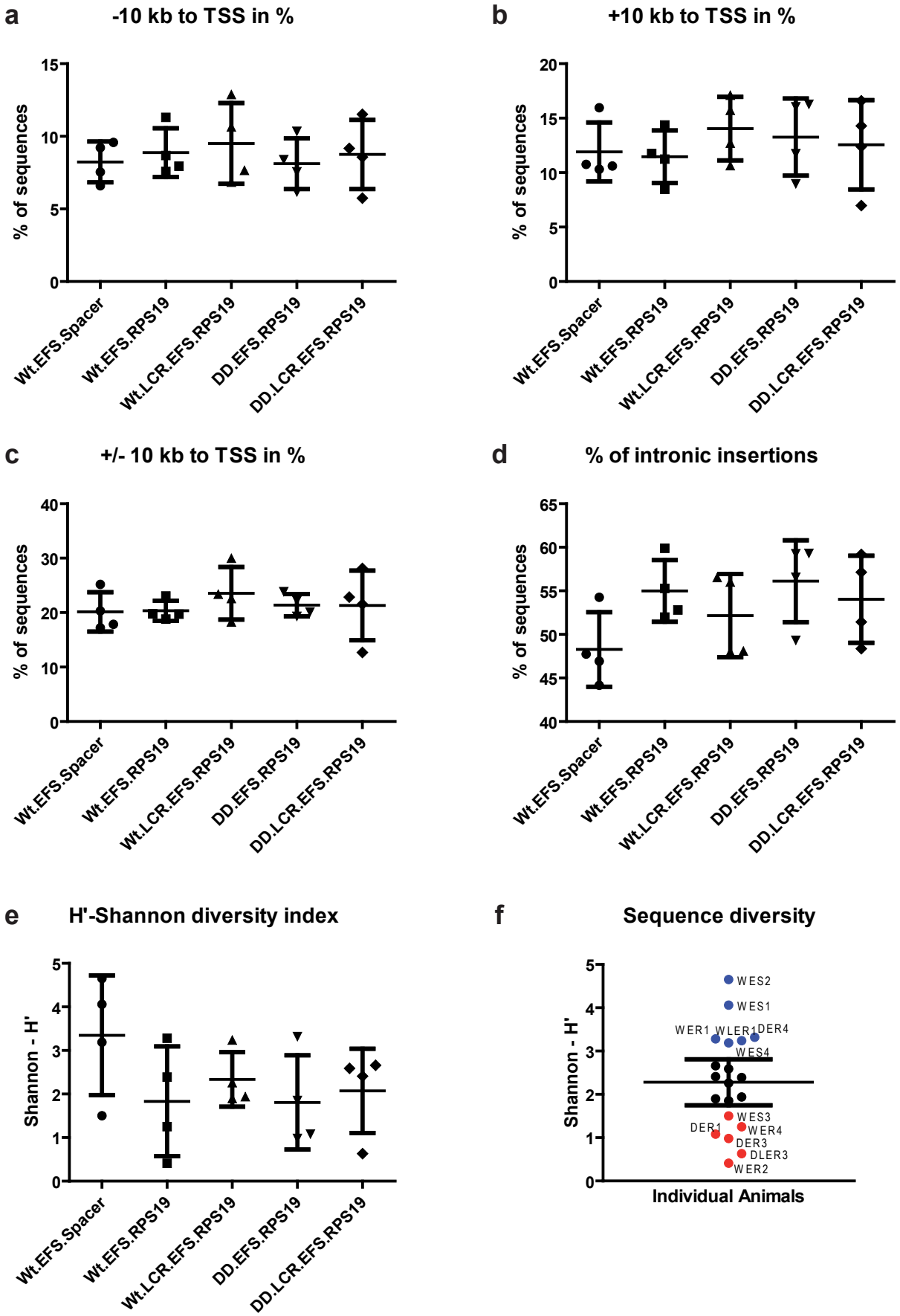
Shubhranshu Debnath, Pekka Jaako, Kavitha Siva, Michael Rothe, Jun Chen, Maria Dahl, H. Bobby Gaspar, Johan Flygare, Axel Schambach, and Stefan Karlsson

Supplementary Figure 1



Supplementary Figure 1
(a) The number of sequences allocated to each barcode primer (BC). Sequences carrying a full SIN-LTR region (b) and Linker (c). Number of sequences after internal control removal (d). Number of individual seuqneces in each vector goup (e). Number of sequences with a read count abover the 97.5%-tile of all sequences in one animal. Bars indicate means +/- SD.

Supplementary Figure 2

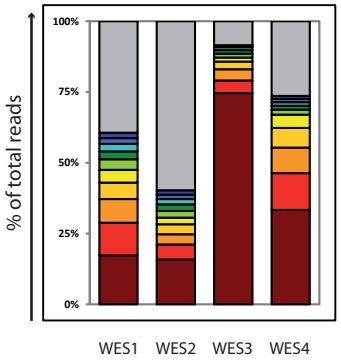


Supplementary Figure 2
Percent of sequences found in a 10 kb window upstream (a) or downstream (b) of the transcriptional start site (TSS) of a gene and the combined TSS information (c). Percent of intronic sequences found (d). The Shannon diversity index for the different vector groups (e). Shannon indices of all individual animals (f). Animals above the 95% confidence interval (CI) are marked in blue, those with a lower index in red. In a-e, bars indicate means +/- SD. In f, the bar indicates the mean together with the upper and lower 95%-CI.

Supplementary Figure 3

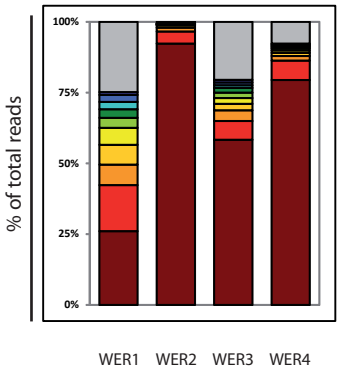
Wt.EFS.Spacer

	WES1		WES2		WES3		WES4	
#	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol
1	17.30%	Zfp869	15.84%	Klhl38	74.59%	Gm1574	33.39%	Thumpd3
2	11.55%	Gm13498	5.37%	Ankfn1	4.51%	Smpdl3a	12.98%	Tsr2
3	8.31%	Aicda	3.53%	Grem2	3.90%	Vstm2l	9.01%	Kcna10
4	5.85%	Golim4	3.53%	Usp24	2.77%	Hs3st1	6.92%	Herc3
5	4.49%	Herc6	2.36%	Olfr1211	1.41%	Oxr1	4.70%	Lcorl
6	3.72%	Slc41a2	2.36%	Tsr2	1.36%	4930402K13Rik	1.79%	Prss36
7	2.80%	Sgms1	2.36%	Slc41a2	1.23%	Pde10a	1.37%	4632404H12Rik
8	2.66%	Wnt2	1.83%	Sod3	1.05%	Gm8994	1.24%	Otoa
9	2.03%	Otoa	1.57%	Sh3bp4	0.35%	Fchsd2	1.20%	Fam5c
10	1.93%	Lepre1	1.57%	Adck1	0.32%	Nell1	1.07%	Aldh2



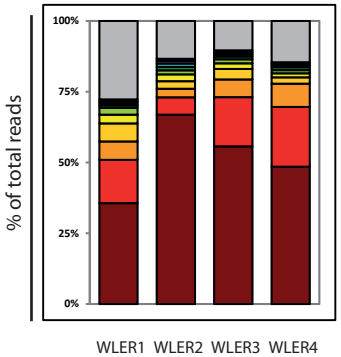
Wt.EFS.RPS19

	WER1		WER2		WER3		WER4	
#	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol
1	26.08%	Cdh23	92.28%	Clasp2	58.35%	Asb5	79.43%	Clasp2
2	16.24%	Wwox	4.29%	Plxdc2	6.65%	Prkca	6.88%	Shroom3
3	7.22%	Sult4a1	1.38%	Veph1	3.73%	Pik3c3	1.64%	Brpf1
4	7.03%	Zfp516	0.78%	Gm2382	2.33%	Foxi3	1.10%	Hspg2
5	5.98%	Dhx15	0.21%	Pdha2	2.03%	Palld	0.87%	Ptbp2
6	3.55%	Txlng	0.13%	Tbcd	1.89%	Hspg2	0.72%	Arhgap21
7	3.01%	Rpap3	0.11%	Supt3h	1.66%	Zik1	0.51%	Steap4
8	2.63%	Hmgcs2	0.05%	2810021B07Rik	0.97%	Satb2	0.45%	Pvrl1
9	2.47%	March1	0.04%	Cd9	0.97%	Otof	0.38%	Ahr
10	0.98%	Sgms1	0.03%	C130026I21Rik	0.92%	Prim2	0.32%	Timp2



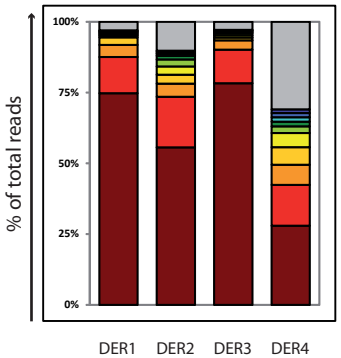
Wt.LCR.EFS.RPS19

	WLER1		WLER2		WLER3		WLER4	
#	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol
1	35.66%	Nedd4l	66.86%	Spag16	55.67%	Gm8910	48.54%	B3gnt2
2	15.28%	Spock1	6.15%	Tssc1	17.42%	Sfi1	21.08%	Arhgap21
3	6.48%	Prdm11	3.02%	Pecam1	6.22%	1810012P15Rik	8.21%	Nr2f1
4	6.38%	Gse1	2.67%	Cpeb3	3.71%	Atp2b2	2.23%	Fchsd2
5	3.06%	Smo	2.38%	Fggy	1.99%	Gm8910	1.44%	Pth1r
6	2.45%	Zhx2	1.42%	A330033J07Rik	1.42%	Cand1	1.08%	Slitrk3
7	0.83%	Map4k2	1.27%	Clstn2	0.98%	Gtf3c1	1.06%	Fos
8	0.72%	Ckap2l	1.15%	Tmem86a	0.73%	Cand1	0.62%	Dpyd
9	0.71%	Stk4	0.88%	Acyp2	0.72%	Aff3	0.58%	Fam171a1
10	0.69%	Ttc23l	0.82%	En1	0.71%	Gm5045	0.55%	Cdh13



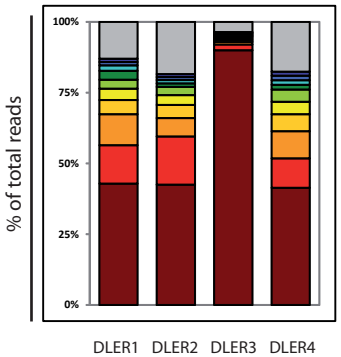
DD.EFS.RPS19

	DER1		DER2		DER3		DER4	
#	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol
1	74.72%	Malt1	55.65%	Ogfrl1	78.25%	Olig3	27.99%	Slc35b4
2	12.84%	Rbbp4	17.83%	Mta3	11.89%	Il21r	14.39%	Maml2
3	4.26%	Phip	4.66%	Rfk	3.31%	Olfr417	7.15%	Pcca
4	2.67%	2700078E11Rik	3.18%	Tbl1xr1	0.99%	Ephb2	6.14%	Ndrg4
5	0.57%	Tshz2	2.91%	Slc35b4	0.80%	Lao1	5.05%	Cd37
6	0.49%	Mixl1	2.45%	4833422C13Rik	0.74%	Hdac2	2.33%	Tkt
7	0.49%	Il21r	1.33%	Gcap14	0.39%	Cd36	1.63%	Cyp26b1
8	0.36%	Diap3	0.69%	Speer7-ps1	0.33%	Etl4	1.56%	1700018B08Rik
9	0.31%	Chuk	0.63%	Pnliprp2	0.26%	Ankrd28	1.48%	Sgk1
10	0.23%	5730507C01Rik	0.43%	Nkx1-1	0.20%	Tkt	1.32%	Kcnt2



DD.LCR.EFS.RPS19

	DLER1		DLER2		DLER3		DLER4	
#	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol	% Reads	Gene Symbol
1	42.88%	Cd81	42.53%	Plekhb1	89.99%	Cdh26	41.39%	Cebpe
2	13.53%	Cdh13	16.98%	Slco4c1	2.02%	Mier1	10.38%	4932443L11Rik
3	10.97%	Gtf3c1	6.53%	Pkd2l2	0.90%	C1galt1	9.62%	Cd93
4	5.01%	Stk40	4.63%	Tpm1	0.67%	Mrgprg	5.99%	Eya1
5	4.00%	Adam32	3.46%	Raph1	0.67%	Dok5	4.39%	Aadacl3
6	3.19%	1700030K09Rik	2.89%	Lsamp	0.56%	Eya1	4.31%	Ndufaf3
7	3.15%	4932443L11Rik	1.27%	Fzd6	0.45%	Pappa	1.70%	Pxn
8	1.91%	Bad	1.13%	Fhit	0.45%	Tbcb	1.61%	Glul
9	1.18%	Gm17019	1.09%	Fat3	0.34%	Itgb5	1.60%	Gal3st4
10	1.17%	Lamb1	1.06%	1700056E22Rik	0.34%	Grid1	1.40%	Mark2



The genes of the top 10 contributing sequences (in %) are shown for each animal. The bar graphs on the right display this composition as stacked frequencies. The color code of the bars corresponds to the color code of the numbers (#) in the respective tables (grey indicate the sum of all other sequences). Gene symbols are colored in case they were found more than once.

Supplementary Table 1 - Sequence statistics

Individual sequencing results:

Animal	Vector Group	BC	Sequences allocated to BC	Carrying a full LTR	Carrying a full Linker	After internal control removal	Aligned sequences using MAVRIC	Individual Sequences	SeqIDs >97.5%-tile	- 10 kb to TSS in %	+ 10 kb to TSS in %	+/- 10 kb to TSS in %	Mean TSS Distance (kb)	% of intronic insertions (closest)	% of intronic insertions	H'-Shannon diversity index	% GFP positive preTX	% GFP positive 18 weeks	RTCGD overlap closest	AllOnco overlap closest
WES1	Wt.EFS.Spacer	1	105494	103740	65022	63553	2069	398	10	7.5%	10.3%	17.8%	91	31.66%	47.74%	4.06	23.50%	5.14%	21.5%	11.8%
WES2	Wt.EFS.Spacer	2	100350	99502	37108	35724	764	282	8	9.2%	16.0%	25.2%	83	37.59%	54.26%	4.65	25.10%	13.10%	20.0%	14.7%
WES3	Wt.EFS.Spacer	3	120633	116954	73715	72065	13866	424	11	6.6%	10.6%	17.2%	103	32.78%	46.93%	1.50	25.10%	12.60%	19.4%	13.5%
WES4	Wt.EFS.Spacer	4	112974	111675	62595	61028	4684	428	11	9.6%	10.7%	20.3%	106	28.27%	44.16%	3.19	11.40%	11.50%	20.5%	12.1%
WER1	Wt.EFS.RPS19	5	129681	127154	64593	63197	3159	340	9	7.9%	11.8%	19.7%	72	33.53%	55.29%	3.28	45.40%	19.40%	16.7%	12.8%
WER2	Wt.EFS.RPS19	6	130555	128646	84691	83221	50704	177	5	11.3%	8.5%	19.8%	79	32.77%	51.98%	0.41	32.00%	32.30%	20.8%	11.9%
WER3	Wt.EFS.RPS19	7	109192	106162	61024	59096	6494	356	9	7.6%	11.2%	18.8%	109	33.71%	52.81%	2.39	38.30%	17.60%	21.0%	12.1%
WER4	Wt.EFS.RPS19	8	148872	147339	65164	63243	10748	334	9	8.7%	14.4%	23.1%	63	39.22%	59.88%	1.25	42.10%	4.90%	20.7%	11.0%
WLER1	Wt.LCR.EFS.RPS19	11	128438	127783	24884	23013	7224	497	13	12.9%	17.1%	30.0%	61	34.41%	56.54%	3.24	11.80%	15.20%	19.8%	11.6%
WLER2	Wt.LCR.EFS.RPS19	12	113195	110689	43684	42894	6843	235	6	7.7%	10.6%	18.3%	92	30.64%	48.09%	1.90	68.90%	15.20%	18.0%	13.4%
WLER3	Wt.LCR.EFS.RPS19	21	139277	134936	63442	61599	8197	248	7	6.9%	15.7%	22.6%	84	33.87%	47.98%	1.94	68.90%	59.30%	23.4%	11.3%
WLER4	Wt.LCR.EFS.RPS19	22	161499	156489	41953	40783	4520	291	8	10.7%	12.7%	23.4%	103	34.71%	56.01%	2.26	23.70%	17.40%	19.0%	8.9%
DER1	DD.EFS.RPS19	13	148907	144519	102312	101976	11484	145	4	10.3%	9.0%	19.3%	86	34.48%	55.86%	1.08	43.50%	76.20%	17.1%	11.6%
DER2	DD.EFS.RPS19	14	138631	135513	46789	44767	15308	81	2	6.2%	16.0%	22.2%	87	35.80%	59.26%	1.85	40.60%	71.60%	20.3%	12.2%
DER3	DD.EFS.RPS19	23	128768	124575	63485	60339	41511	298	8	8.4%	11.7%	20.1%	85	29.87%	49.33%	0.98	40.60%	60.80%	16.4%	11.6%
DER4	DD.EFS.RPS19	24	274024	265305	109004	108042	1286	172	5	7.6%	16.3%	23.8%	78	37.79%	59.30%	3.32	40.20%	71.40%	20.9%	12.9%
DLER1	DD.LCR.EFS.RPS19	15	268150	265875	59078	57761	16021	218	6	9.2%	12.4%	21.6%	72	40.37%	59.17%	2.41	10.00%	68.90%	19.1%	8.8%
DLER2	DD.LCR.EFS.RPS19	16	144241	142428	45056	44413	4423	244	7	5.7%	7.0%	12.7%	116	36.07%	48.36%	2.59	68.10%	63.80%	18.8%	11.9%
DLER3	DD.LCR.EFS.RPS19	17	155660	150682	122317	122181	889	35	1	8.6%	14.3%	22.9%	106	28.57%	51.43%	0.63	68.10%	88.90%	17.6%	2.9%
DLER4	DD.LCR.EFS.RPS19	18	117431	116347	34094	33351	8137	217	6	11.5%	16.6%	28.1%	53	31.34%	57.14%	2.66	22.90%	56.90%	22.7%	11.1%

Mean values for the individual groups:

Animal	Vector Group	BC	Sequences allocated to BC	Carrying a full LTR	Carrying a full Linker	After internal control removal	Aligned sequences using MAVRIC	Individual Sequences	SeqIDs >97.5%-tile	- 10 kb to TSS in %	+ 10 kb to TSS in %	+/- 10 kb to TSS in %	Mean TSS Distance (kb)	% of intronic insertions (closest)	% of intronic insertions	H'-Shannon diversity index	% GFP positive preTX	% GFP positive 18 weeks	RTCGD overlap % closest	AllOnco overlap % closest
WES1-4	Wt.EFS.Spacer	-	1.10E+05	1.08E+05	5.96E+04	5.81E+04	5346	383	10	8.2%	11.9%	20.1%	96	32.6%	48.3%	3.35	21.3%	10.6%	20.3%	13.0%
WER1-4	Wt.EFS.RPS19	-	1.30E+05	1.27E+05	6.89E+04	6.72E+04	17776	302	8	8.9%	11.5%	20.3%	91	34.8%	55.0%	1.83	39.5%	18.6%	19.8%	12.0%
WLER1-4	Wt.LCR.EFS.RPS19	-	1.36E+05	1.32E+05	4.35E+04	4.21E+04	6696	318	9	9.5%	14.0%	23.6%	90	33.4%	52.2%	2.34	43.3%	26.8%	20.0%	11.3%
DER1-4	DD.EFS.RPS19	-	1.73E+05	1.67E+05	8.04E+04	7.88E+04	17397	174	5	8.1%	13.3%	21.4%	91	34.5%	55.9%	1.81	41.2%	70.0%	18.7%	12.1%
DLER1-4	DD.LCR.EFS.RPS19	-	1.71E+05	1.69E+05	6.51E+04	6.44E+04	7368	179	5	8.8%	12.6%	21.3%	81	34.1%	54.0%	2.08	42.3%	69.6%	19.6%	8.7%

Abbreviations and explanation of special terms:
Animals 1-4 = **WES** for **W**t.**E**FS.**S**pacer; **WER** = **W**t.**E**FS.**R**PS19; **WLER** = **W**t.**L**CR.**E**FS.**R**PS19; **DER** = **D**D.**E**FS.**R**PS19; **DLER** = **D**D.**L**CR.**E**FS.**R**PS19.
BC = Barcode primer ID; LTR = Long terminal repeat; internal control = vector specific amplicon generated for all samples due to primer extension from the LTR into the vector; MAVRIC = alignment tool available at http://mavric.erasmusmc.nl; SeqIDs >97.5%-tile = The read count for all individual sequences which belonged to a specific barcode primer were used determine the 97.5% as a cutoff to identify statisitcally dominant sequences; kb = kilo base pairs; TSS = transcriptional start site.

Supplementary Table 2 - High read database overlaps

Sequence	BC	Gene Symbol	Reads	Reads %	Chr.	Raw Distance	TSS Distance	RTCGD	NCG	Bushman	Deichmann	Vector
SeqID586509	13	Malt1	8201	74.73%	18	intronic	11974	1				DD.EFS.RPS19
SeqID4205569	13	Rbbp4	314	12.85%	4	intronic	7647					DD.EFS.RPS19
SeqID2713415	13	2700078E11R	290	2.67%	19	intronic	31910					DD.EFS.RPS19
SeqID4252778	13	Phip	256	4.26%	9	intronic	27925					DD.EFS.RPS19
SeqID3565639	14	Ogfrl1	8181	55.65%	1	-14065	-14113					DD.EFS.RPS19
SeqID43762	14	Mta3	2032	17.83%	17	intronic	29885	3				DD.EFS.RPS19
SeqID1949000	23	Olig3	32483	78.25%	10	25772	-25822					DD.EFS.RPS19
SeqID1108949	23	Il21r	4385	11.89%	7	intronic	19608					DD.EFS.RPS19
SeqID5573409	23	Olfr417	1290	3.31%	1	12308	-12488					DD.EFS.RPS19
SeqID1184833	23	4930549C01R	409	0.99%	4	102741	105531					DD.EFS.RPS19
SeqID4697585	23	Lao1	327	0.80%	4	intronic	6412					DD.EFS.RPS19
SeqID4176892	23	Hdac2	307	0.74%	10	17354	-17383					DD.EFS.RPS19
SeqID6111976	23	Arhgap21	135	0.33%	2	-221609	342572	1				DD.EFS.RPS19
SeqID1486607	23	Cd36	120	0.39%	5	intronic	55201					DD.EFS.RPS19
SeqID4384655	24	Slc35b4	342	28.02%	6	-5909	-5970					DD.EFS.RPS19
SeqID4894480	24	Maml2	168	14.40%	9	21345	-21409					DD.EFS.RPS19
SeqID174802	24	4930594M22R	91	7.16%	14	185058	-185224					DD.EFS.RPS19
SeqID1545991	24	Ndrg4	78	6.15%	8	intronic	22298					DD.EFS.RPS19
SeqID4687807	24	Cd37	17	5.06%	7	-9747	-9785	3				DD.EFS.RPS19
SeqID4473143	15	Cd81	6693	42.88%	7	intronic	8627					DD.LCR.EFS.RPS19
SeqID4473143	15	Kcnq1	6693	42.88%	7	45823	-45888					DD.LCR.EFS.RPS19
SeqID1076133	15	Hsbp1	2160	13.53%	8	30307	-30347					DD.LCR.EFS.RPS19
SeqID164608	15	Gtf3c1	1758	10.97%	7	intronic	1501					DD.LCR.EFS.RPS19
SeqID6687017	15	Stk40	719	5.01%	4	intronic	26671					DD.LCR.EFS.RPS19
SeqID4031797	15	Adam32	641	4.00%	8	intronic	19399					DD.LCR.EFS.RPS19
SeqID2354018	15	1700030K09R	509	3.19%	8	intronic	14892					DD.LCR.EFS.RPS19
SeqID2354018	15	Eps15l1	509	3.19%	8	-37297	-37359	5				DD.LCR.EFS.RPS19
SeqID3289489	16	Plekhb1	1830	42.54%	7	-1214	-1271	3				DD.LCR.EFS.RPS19
SeqID1061386	16	Slco4c1	677	16.98%	1	-83376	136764					DD.LCR.EFS.RPS19
SeqID4056926	16	Pkd2l2	278	6.54%	18	intronic	23860	1				DD.LCR.EFS.RPS19
SeqID4885305	16	Tpm1	205	4.64%	9	intronic	7085					DD.LCR.EFS.RPS19
SeqID4920862	16	Raph1	147	3.46%	1	-35284	-35319					DD.LCR.EFS.RPS19
SeqID4120201	16	Gap43	115	2.89%	16	-239253	331463					DD.LCR.EFS.RPS19
SeqID3430700	16	Fzd6	56	1.27%	15	intronic	13874					DD.LCR.EFS.RPS19
SeqID3281443	17	Cdh26	800	90.09%	2	intronic	30063	1				DD.LCR.EFS.RPS19
SeqID6464552	18	Cebpe	3272	41.40%	14	-4779	6591	2				DD.LCR.EFS.RPS19
SeqID6464552	18	Acin1	3272	41.40%	14	-18568	-18652					DD.LCR.EFS.RPS19
SeqID4445526	18	4932443L11R	845	10.39%	8	intronic	13914					DD.LCR.EFS.RPS19
SeqID4445526	18	Evi5l	845	10.39%	8	intronic	26328					DD.LCR.EFS.RPS19
SeqID4445526	18	Lrrc8e	845	10.39%	8	33858	-33932					DD.LCR.EFS.RPS19
SeqID4445526	18	Map2k7	845	10.39%	8	45771	-45845					DD.LCR.EFS.RPS19
SeqID4258572	18	Cd93	739	9.62%	2	intronic	6570					DD.LCR.EFS.RPS19
SeqID4258572	18	Sstr4	739	9.62%	2	36305	41432	1				DD.LCR.EFS.RPS19
SeqID4956240	18	Eya1	430	5.99%	1	intronic	9140	1				DD.LCR.EFS.RPS19
SeqID6124805	18	Ndufaf3	345	4.31%	9	overlapping	1371					DD.LCR.EFS.RPS19
SeqID6124805	18	Dalrd3	345	4.31%	9	3920	-4053	2				DD.LCR.EFS.RPS19
SeqID573864	18	Aadacl3	325	4.39%	4	intronic	8110					DD.LCR.EFS.RPS19
SeqID3132051	5	Gm17455	824	26.09%	10	79572	83403					Wt.EFS.RPS19
SeqID4932493	5	Wwox	394	16.24%	8	intronic	131044	1				Wt.EFS.RPS19
SeqID1029546	5	Sult4a1	227	7.22%	15	intronic	8265					Wt.EFS.RPS19
SeqID4376068	5	Zfp516	222	7.03%	18	intronic	39408	1				Wt.EFS.RPS19
SeqID2637860	5	Dhx15	189	5.98%	5	intronic	36506					Wt.EFS.RPS19
SeqID5006758	5	Txlng	111	3.55%	X	intronic	8747					Wt.EFS.RPS19
SeqID4077725	5	Rpap3	95	3.01%	15	-77810	108528					Wt.EFS.RPS19
SeqID5028364	5	Hmgcs2	83	2.63%	3	intronic	15678					Wt.EFS.RPS19
SeqID2028512	5	March1	78	2.47%	8	203902	-203928	2				Wt.EFS.RPS19
SeqID1374495	6	Clasp2	31008	92.28%	9	intronic	29869					Wt.EFS.RPS19
SeqID6200104	6	Plxdc2	2176	4.29%	2	intronic	301285	1				Wt.EFS.RPS19
SeqID1350306	6	Veph1	504	1.38%	3	-8438	-8528	1				Wt.EFS.RPS19
SeqID3286496	6	Gm2382	396	0.78%	9	intronic	7111					Wt.EFS.RPS19
SeqID4583289	6	Pdha2	104	0.21%	3	-4512	6864					Wt.EFS.RPS19
SeqID1853950	7	Asb5	3330	58.36%	8	206019	-206043					Wt.EFS.RPS19
SeqID5725894	7	Prkca	409	6.65%	11	intronic	13088	1				Wt.EFS.RPS19
SeqID3892214	7	Pik3c3	223	3.73%	18	intronic	71794					Wt.EFS.RPS19
SeqID2917855	7	Foxi3	136	2.33%	6	69285	73821					Wt.EFS.RPS19
SeqID615467	7	Ldlrad2	123	1.89%	4	-20192	23960					Wt.EFS.RPS19
SeqID615467	7	Usp48	123	1.89%	4	43071	-43115					Wt.EFS.RPS19
SeqID3108392	7	4930512H18R	120	2.03%	8	7101	-7232					Wt.EFS.RPS19
SeqID4207183	7	Satb2	63	0.97%	1	intronic	13					Wt.EFS.RPS19
SeqID6083448	7	Otof	63	0.97%	5	intronic	61137					Wt.EFS.RPS19
SeqID5142031	7	Zik1	57	1.66%	7	intronic	5762					Wt.EFS.RPS19
SeqID2937544	8	Clasp2	8537	79.44%	9	intronic	29959					Wt.EFS.RPS19
SeqID3393560	8	Shroom3	739	6.88%	5	intronic	13551					Wt.EFS.RPS19
SeqID3306329	8	Brpf1	176	1.64%	6	intronic	996					Wt.EFS.RPS19
SeqID3020552	8	Ldlrad2	118	1.10%	4	-20152	23920					Wt.EFS.RPS19
SeqID3020552	8	Usp48	118	1.10%	4	43031	-43138					Wt.EFS.RPS19
SeqID2939831	8	Ptbp2	94	0.87%	3	-185892	-186085					Wt.EFS.RPS19
SeqID5421272	8	Steap4	51	0.51%	5	109606	-109690					Wt.EFS.RPS19
SeqID5721195	8	Pvrl1	47	0.45%	9	25889	-25920					Wt.EFS.RPS19
SeqID4441213	8	Arhgap21	43	0.72%	2	-19542	140505	1				Wt.EFS.RPS19
SeqID4985045	8	Ahr	36	0.38%	12	-159796	-159889					Wt.EFS.RPS19
SeqID2421234	1	Zfp869	358	17.31%	8	-13765	-13798					Wt.EFS.Spacer
SeqID1000377	1	Gm13498	175	11.56%	2	1925	4091					Wt.EFS.Spacer
SeqID1211330	1	Aicda	168	8.32%	6	11908	22288					Wt.EFS.Spacer
SeqID1211330	1	Apobec1	168	8.32%	6	-1629	26282					Wt.EFS.Spacer
SeqID2326942	1	Pigy	89	4.50%	6	-39374	44877					Wt.EFS.Spacer
SeqID6572120	1	Slc41a2	77	3.72%	10	intronic	38093					Wt.EFS.Spacer
SeqID6322744	1	Golim4	61	5.85%	3	-7726	-7765					Wt.EFS.Spacer
SeqID3991195	1	Wnt2	55	2.66%	6	-41048	82694					Wt.EFS.Spacer
SeqID3340846	1	Lepre1	40	1.93%	4	intronic	1615					Wt.EFS.Spacer
SeqID2492575	1	2700046G09R	31	2.80%	19	151477	-151558					Wt.EFS.Spacer
SeqID1707480	1	Otoa	26	2.03%	7	intronic	16849					Wt.EFS.Spacer
SeqID6466903	2	Klhl38	83	15.86%	15	intronic	3791					Wt.EFS.Spacer
SeqID2504271	2	Ankfn1	41	5.37%	11	intronic	99709					Wt.EFS.Spacer
SeqID4110068	2	Grem2	27	3.54%	1	-2698	90733					Wt.EFS.Spacer
SeqID4467672	2	Usp24	27	3.54%	4	284393	-284424	2				Wt.EFS.Spacer
SeqID11533	2	Olfr1211	18	2.36%	2	-1488	2424					Wt.EFS.Spacer
SeqID3928647	2	Tsr2	18	2.36%	X	intronic	6731					Wt.EFS.Spacer
SeqID3928647	2	Fgd1	18	2.36%	X	228	43600					Wt.EFS.Spacer
SeqID4836488	2	Slc41a2	18	2.36%	10	intronic	38093					Wt.EFS.Spacer
SeqID1675022	2	Sod3	14	1.83%	5	intronic	5296					Wt.EFS.Spacer
SeqID4384147	3	Gm1574	10343	74.60%	13	19063	-19156					Wt.EFS.Spacer
SeqID2337964	3	Smpdl3a	625	4.51%	10	intronic	17039					Wt.EFS.Spacer
SeqID6527559	3	Vstm2l	375	3.90%	2	intronic	10662					Wt.EFS.Spacer
SeqID2619595	3	Hs3st1	299	2.77%	5	-74773	216314					Wt.EFS.Spacer
SeqID3572513	3	Oxr1	196	1.41%	15	93769	-93848					Wt.EFS.Spacer
SeqID3642672	3	4930402K13R	187	1.36%	X	12279	14060					Wt.EFS.Spacer
SeqID4598140	3	1700010I14Ri	162	1.23%	17	217316	-217341	1				Wt.EFS.Spacer

SeqID53789	3	Gm8994	145	1.05%	6	52683	-52763					Wt.EFS.Spacer
SeqID4597121	3	Fchsd2	48	0.35%	7	intronic	48599	4				Wt.EFS.Spacer
SeqID2721847	3	Nell1	44	0.32%	7	144467	1034859					Wt.EFS.Spacer
SeqID6629317	3	Tex2	43	0.31%	11	-68347	-68384	1				Wt.EFS.Spacer
SeqID569910	4	Thumpd3	1564	33.40%	6	37331	-37411	2				Wt.EFS.Spacer
SeqID3138597	4	Tsr2	604	12.98%	X	intronic	6731					Wt.EFS.Spacer
SeqID3138597	4	Fgd1	604	12.98%	X	232	43604					Wt.EFS.Spacer
SeqID3112660	4	Kcna10	411	9.01%	3	2172	-2270					Wt.EFS.Spacer
SeqID6459712	4	Nap1l5	190	6.92%	6	-19260	21153					Wt.EFS.Spacer
SeqID5298349	4	Lcorl	179	4.70%	5	-47105	-47136	1				Wt.EFS.Spacer
SeqID58947	4	Prss36	80	1.79%	7	-7287	-7384					Wt.EFS.Spacer
SeqID58947	4	Fus	80	1.79%	7	13347	-13444					Wt.EFS.Spacer
SeqID58947	4	Myst1	80	1.79%	7	28180	41496					Wt.EFS.Spacer
SeqID4917017	4	4632404H12R	62	1.37%	3	intronic	4429					Wt.EFS.Spacer
SeqID1852058	4	Otoa	56	1.24%	7	intronic	16870					Wt.EFS.Spacer
SeqID3249122	4	Fgfbp1	46	0.98%	5	-15631	18529					Wt.EFS.Spacer
SeqID6400041	4	Aldh2	43	1.07%	5	intronic	19266					Wt.EFS.Spacer
SeqID3612652	4	Fam5c	28	1.20%	1	143219	-143242					Wt.EFS.Spacer
SeqID3329693	11	Mir122a	2469	35.66%	18	33455	-33551					Wt.LCR.EFS.RPS19
SeqID1720765	11	Spock1	1095	15.28%	13	intronic	285168					Wt.LCR.EFS.RPS19
SeqID6418278	11	Syt13	466	6.48%	2	13367	54328					Wt.LCR.EFS.RPS19
SeqID2944184	11	Gse1	456	6.38%	8	38190	-38243	6				Wt.LCR.EFS.RPS19
SeqID3029594	11	Smo	221	3.06%	6	intronic	9439					Wt.LCR.EFS.RPS19
SeqID5015274	11	9330154K18R	98	2.45%	15	-59002	-59078					Wt.LCR.EFS.RPS19
SeqID1182811	11	Map4k2	58	0.83%	19	intronic	10045					Wt.LCR.EFS.RPS19
SeqID1182811	11	Sf1	58	0.83%	19	12450	-12510					Wt.LCR.EFS.RPS19
SeqID1182811	11	Men1	58	0.83%	19	10288	16201					Wt.LCR.EFS.RPS19
SeqID1182811	11	Rasgrp2	58	0.83%	19	48100	-48160	4				Wt.LCR.EFS.RPS19
SeqID3113489	11	Ckap2l	52	0.72%	2	intronic	287					Wt.LCR.EFS.RPS19
SeqID3113489	11	Il1a	52	0.72%	2	-2684	13047					Wt.LCR.EFS.RPS19
SeqID1164543	11	Sort1	47	0.65%	3	intronic	44728					Wt.LCR.EFS.RPS19
SeqID5147166	11	Ttc23l	47	0.69%	15	-1151	-1230					Wt.LCR.EFS.RPS19
SeqID3641389	11	Kcns1	39	0.71%	2	-28837	36332	1				Wt.LCR.EFS.RPS19
SeqID5265118	11	Ttc1	29	0.51%	11	intronic	16637					Wt.LCR.EFS.RPS19
SeqID5153003	11	Mgll	26	0.53%	6	intronic	37742	1				Wt.LCR.EFS.RPS19
SeqID5403262	12	Spag16	4401	66.87%	1	intronic	229362					Wt.LCR.EFS.RPS19
SeqID5801105	12	Tssc1	417	6.15%	12	272564	388228					Wt.LCR.EFS.RPS19
SeqID5439955	12	Tex2	181	3.03%	11	-68381	-68410	1				Wt.LCR.EFS.RPS19
SeqID145626	12	Fggy	163	2.38%	4	intronic	86125					Wt.LCR.EFS.RPS19
SeqID3632824	12	A330032B11R	137	2.67%	19	103340	-103416					Wt.LCR.EFS.RPS19
SeqID3513274	12	Id4	82	1.42%	13	76167	-76193					Wt.LCR.EFS.RPS19
SeqID2573674	21	Gm8910	3981	55.67%	3	-15914	-15954					Wt.LCR.EFS.RPS19
SeqID4578111	21	Sfi1	1334	17.42%	11	-2411	64025					Wt.LCR.EFS.RPS19
SeqID6032595	21	1810012P15R	510	6.22%	11	20115	40402					Wt.LCR.EFS.RPS19
SeqID2190941	21	Sec13	304	3.71%	6	-138251	-138303					Wt.LCR.EFS.RPS19
SeqID4887823	21	Cand1	116	1.42%	10	intronic	38024					Wt.LCR.EFS.RPS19
SeqID3908770	21	Gm8910	86	1.99%	3	-15860	-15888					Wt.LCR.EFS.RPS19
SeqID5816349	21	Gtf3c1	74	0.98%	7	intronic	1501					Wt.LCR.EFS.RPS19
SeqID5961090	22	B3gnt2	2194	48.55%	11	-26770	52985	2				Wt.LCR.EFS.RPS19
SeqID1379082	22	Arhgap21	854	21.09%	2	-19558	140521	1				Wt.LCR.EFS.RPS19
SeqID1695057	22	Nr2f1	320	8.21%	13	-458602	-458640					Wt.LCR.EFS.RPS19
SeqID4302263	22	Fchsd2	100	2.24%	7	intronic	33260	4				Wt.LCR.EFS.RPS19
SeqID4146056	22	Pth1r	65	1.44%	9	intronic	20669					Wt.LCR.EFS.RPS19
SeqID1010799	22	Fos	47	1.06%	12	36438	39822	6				Wt.LCR.EFS.RPS19
SeqID1204764	22	Slitrk3	46	1.08%	3	-24849	-24973					Wt.LCR.EFS.RPS19

Legend:
BC = Barcode of index primer; Chr. = Chromosome; TSS = Transcriptional Start Site; CIS = common integration site; RTCGD = Retroviral Tagged Cancer Gene Database (Akagi et al., Nucleic Acids Res. 2004); NCG = Network of Cancer Genes - Version 5 (http://ncg.kcl.ac.uk/); Bushman = Bushman Cancer Gene List (http://www.bushmanlab.org/links/genelists); Deichmann = referring to Deichmann et al., Mol Ther. 2011, as a list of insertions found in clinical and preclinical insertion site screens (this is not a cancer database *per se*).

SeqID4473143	The light blue color indicates that the same SeqID is present more than once in the table. We reported the closest gene, if there was not an intronic hit or any gene within 50 kb TSS distance listed in one of the databases.
6	Found in database with x tumours associated to the gene listed in the retroviral tagged cancer gene database (RTCGD).
	Not found in any of the databases.

Supplementary Table 3 - CIS analysis

Gene Symbol	Chr.	Raw Distance	TSS Distance	CIS	RTCGD	NCG	Bushman	Deichmann
Mid1	X	intronic	318851	21	8			
Lrp1b	2	intronic	1656003	11				
Gm10664	8	-245218	-245275	8				
Clasp2	9	intronic	30094	8				
Rab2a	4	56175	-56215	7				
Gm22	8	104488	-104604	7				
Gm44	X	5129	6122	7				
Ckap2l	2	intronic	33	6				
Gm6531	8	24919	-24953	6				
Fkbp6	5	intronic	58173	6				
1700054O13Rik	X	51371	-51409	6				
Otoa	7	intronic	16953	6				
Nub1	5	intronic	16737	6	1			
Lamb1	12	intronic	43839	6				
St5	7	intronic	15469	5	1			
Samd9l	6	-228431	255746	5				
Ch25h	19	-85477	86827	5	2			
C230021G24Rik	10	-170787	-170830	5				
Gm9966	7	487811	-487858	5				
Ldlrad2	4	-9076	12844	5				
Xrcc6bp1	10	-290161	323103	5	1			
Acss3	10	intronic	85789	5				
Sfi1	11	-1456	63070	5				
Prim2	1	intronic	54524	5				
Pgap3	11	intronic	3018	5				
Pggt1b	18	intronic	33180	5	1			
Lingo1	9	intronic	62255	5				
Klhl38	15	intronic	3710	5				
Tmprss9	10	intronic	5592	5				

Legend:
Chr. = Chromosome; TSS = Transcriptional Start Site; CIS = common integration site; RTCGD = Retroviral Tagged Cancer Gene Database (Akagi et al., Nucleic Acids Res. 2004); NCG = Network of Cancer Genes (http://ncg.kcl.ac.uk/); Bushman = Bushman Cancer Gene List (http://www.bushmanlab.org/links/genelists); Deichmann = referring to Deichmann et al., Mol Ther. 2011.

8	Found in database with x tumours associated to the IS.
	Not found in database