Successful gene therapy of Diamond-Blackfan anemia in a mouse model and human CD34+ cord blood hematopoietic stem cells using a clinically applicable lentiviral vector

Yang Liu,[1] Maria Dahl,[1] Shubhranshu Debnath,[1] Michael Rothe,[2] Emma M. Smith,[1] Tan Hooi Min Grahn,[1] Sarah Warsi,[1] Jun Chen,[1] Johan Flygare,[1] Axel Schambach[2,3] and Stefan Karlsson[1]

[1]Molecular Medicine and Gene Therapy, Lund Stem Cell Center, Lund University, Lund, Sweden; [2]Institute of Experimental Hematology, Hannover Medical School, Hannover, Germany; and [3]Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

 doi:10.3324/haematol.2020.269142

Received: August 27, 2020.

Accepted: December 23, 2020.

Pre-published: January 14, 2021.

Correspondence: *STEFAN KARLSSON* - stefan.karlsson@med.lu.se

YANG LIU - yang.liu@med.lu.se

Supplementary Appendix

Supplementary Methods

Mice

The homozygous doxycycline-inducible *Rps19*-deficient mouse model used in the study was established as previously described[15]. *Rps19* deficiency was induced by 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 (PB) was collected from the tail vein into the microvette tube (Sarstedt) and was analyzed using Sysmex XE-5000. Red blood cells were lysed using ammonium chloride for 10 min at room temperature. To evaluate the contribution toward various blood lineages by flow cytrometry 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). Experiments were performed using a FACS LSR Fortessa cytometer (BD Biosciences) and were analyzed by FlowJo software (version 10.5.3; Tree Star). FACS analysis of the myeloerythroid compartments in BM was performed as previously described[22]. Briefly, BM cells were harvested by crushing the femur and tibia in phosphate-buffered saline (PBS) containing 2% fetal bovine serum (FBS) (Gibco). Fresh cells were stained with the following antibodies: CD45.1 (740889; BD Biosciences), CD45.2 (109806; Biolegend), CD41 (12-0411-83; eBioscience), GR1 (108410; Biolegend), CD11b (101210; Biolegend), B220 (103210; Biolegend), CD3 (100310; Biolegend), c-kit (47-1171-82; eBioscience), CD105 (120404; Biolegend), Ter119 (25-5921-82; eBioscience), 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 Fortessa cytometer (Becton Dickinson) and were analyzed by FlowJo software (version 10.5.3; Tree Star).

Transplantation of hematopoietic progenitor cells

For the transplantation of uninduced gene-corrected BM cells, after incubation for 1 day, $0.5x10^6$ bulk transduced c-kit$^+$ cells were resuspended in 300μL PBS and transplanted into the tail vein of lethally irradiated (900 cGy) wild-type recipients (CD45.1/45.2). For the transplantation of gene-corrected *Rps19*-deficient BM cells, after incubation for 1 day with doxycycline (1 μg/ml), $0.5x10^6$ bulk

transduced Lin- cells and $1x10^6$ untransduced Lin+ cells were resuspended in 300µL PBS and transplanted into the tail vein of lethally irradiated (900 cGy) wild-type recipients (CD45.1/45.2).

Transduction of human primary cord blood cells and erythroid differentiation

The shRNAs along with GFP marker were designed as previously described[23]. For transduction, cells were transduced with shRNAs at an MOI of 5. The sequences for shRNAs are shown in the Supplementary Table S1. After 48 hours of transduction, GFP^+ cells were sorted and transduced with or without EFS-RPS19 at an MOI of 5. To check the successful integration of the vector into human cells, we used genomic DNA isolated from cells with or without EFS-RPS19 transduction as templates, and primers specifically targeting at the lentiviral vector WPRE region were used for the PCR array (the *ALB* gene was used as an internal reference gene). The primer sequences are shown in the Supplementary Table S1. For erythroid differentiation, an equal number of $CD34^+$ cells (20,000) were cultured in erythroid differentiation medium supplemented with different cytokines for three stages of differentiation. The base medium for all three differentiation phases comprises: IMDM, 15% FBS, 1% BSA, 500mg/ml human holo-transferrin, 1% Insulin-Transferrin-Selenium and 1% penicillin–streptomycin and ß-mercaptoethanol. At stage I (day 1 to day 6), cells were cultured in base medium plus 50 ng/ml hSCF, 10 ng/ml human IL-3 (hIL-3) and 6U/ml Erythropoietin. At stage II (day 7 to day 10), hIL-3 was removed from the medium. At stage III (day 10 to day 16), both hIL-3 and hSCF were removed from the medium.

Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR)

RNA was isolated using RNeasy Micro Kit (Qiagen, USA), followed with reverse transcription using SuperScript™ III (Thermo Fiesher, USA). qRT-PCR was performed with SYBR Green PCR Master Mix (Applied Biosystems, USA) without specific indications. All reactions were performed in triplicates. The primer sequences are shown in the Supplementary Table S1.

Determination of transduction efficiency

The c-kit^+ cells were enriched and transduced (MOI=5-10) as mentioned above. At 48 hours after transduction, $40x10^3$ c-kit^+ transduced cells were seeded in 1.5 mL of M3434 methylcellulose (Stem Cell Technologies) and single colonies were picked after 14 days of culture. DNA was isolated from each colony, followed by an PCR assay to detect the inserted vector. We determined the vector-transduced colonies based on the positive detection of WPRE region of the vector. Transduction efficiency was determined by the total number of colonies carrying the targeted WPRE region of the

vector backbone relative to the total colony number. The experiments were performed in triplicates (100 colonies were picked each time).

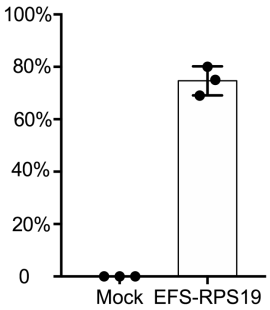
Determination of vector copy number (VCN)

Whole BM or PB cells were isolated at 16 weeks after transplantation. Genomic DNA was isolated with the DNA Blood & Tissue kit (Qiagen). The mean VCN was determined by qRT-PCR using TaqMan Gene Expression Assay (Thermo Fisher Scientific). The number of viral sequences was normalized to the genomic reference sequence. We used WPRE element to detect the viral sequences (as shown in the Supplementary Table S1), and *Tfrc* gene (4458367; Thermo Fisher Scientific) was used for genomic DNA normalization.

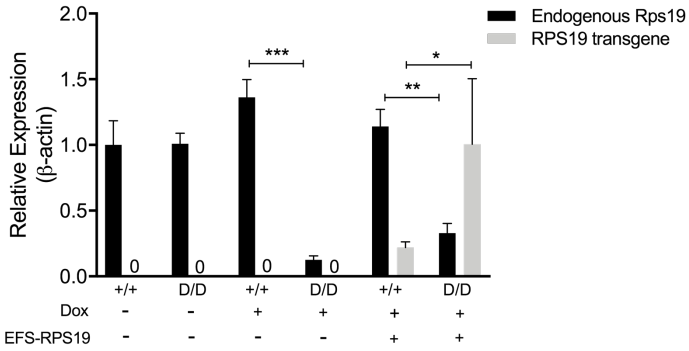
Insertion site analysis

Vector-genome junction was amplified using the INtegration Site PIpeline for paIRed-End reaDs (INSPIIRED) workflow as described by Sherman and colleagues[24]. In brief, genomic DNA was purified using AMPure beads and fragmented with a Covaris 220 sonicator. After an additional round of AMPure purification, DNA was end-repaired and subjected to linker ligation. For each sample, a unique linker was used to prevent cross-contamination between samples during batch processing. Primers specific for the SIN long-terminal repeat (LTR) region of the lentivirus in combination with linker-specific primers were used to amplify vector genome junctions. The amplification from the LTR into the vector (instead of the genome) was controlled and largely prohibited using blocking oligos. In a second nested PCR, LTR-specific index primers together with linker-specific primers were used to attach next-generation sequencing adapters for Illumina paired-end sequencing. After Bioanalyzer quality assessment, the libraries were loaded on MiSeq Nano flow cells. The sequencing reads were deconvoluted according to the primer index used during the second nested PCR step. The individual sequences were quality filtered, aligned to the target genome and quantified by sonic abundance as described in the INSPIIRED bioinformatics pipeline[25]. The iPSC-clones C14 or HD2 were used as monoclonal control samples as indicated. The pool size estimation and sequence diversity were analyzed as previously described[26-28]. Bioinformatic steps were generally performed as described[25]. Particularly, we allowed a total of 5 bp below the quality threshold per amplicon. A quality sliding window of 10 bp was chosen. A maximum fragment length of 2500 bp, a minimal overlap with the reference genome of 20 bp with 95% homology and a start of the alignment no later than 5 bp after the LTR were set in the processing parameters of the INSPIIRED pipeline. The individual files were aligned and annotated to the mouse (mm9, for all mouse samples) and human genome (hg38, for C14 or HD2).

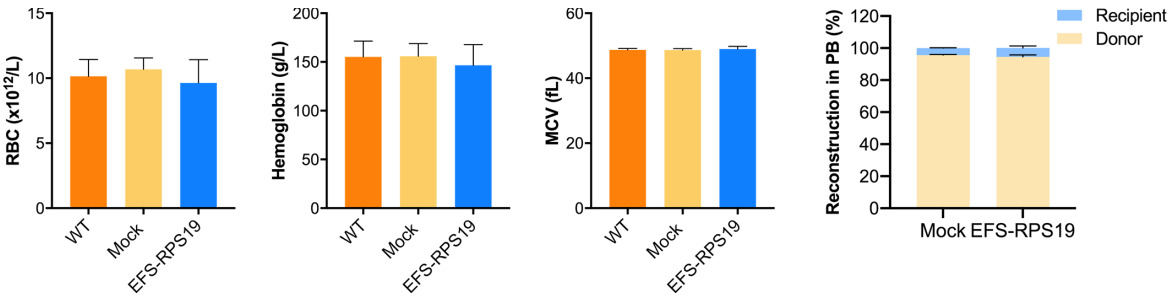
Supplementary Figures



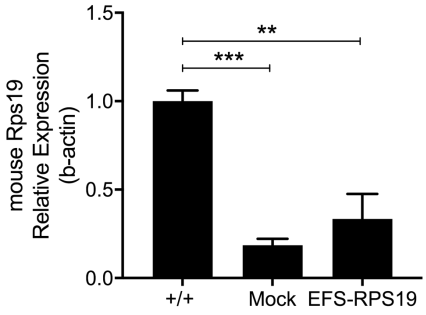
Supplementary Figure S1. Transduction efficiency of the EFS-RPS19 vector in uninduced c-kit^{+} BM cells isolated from D/D mice



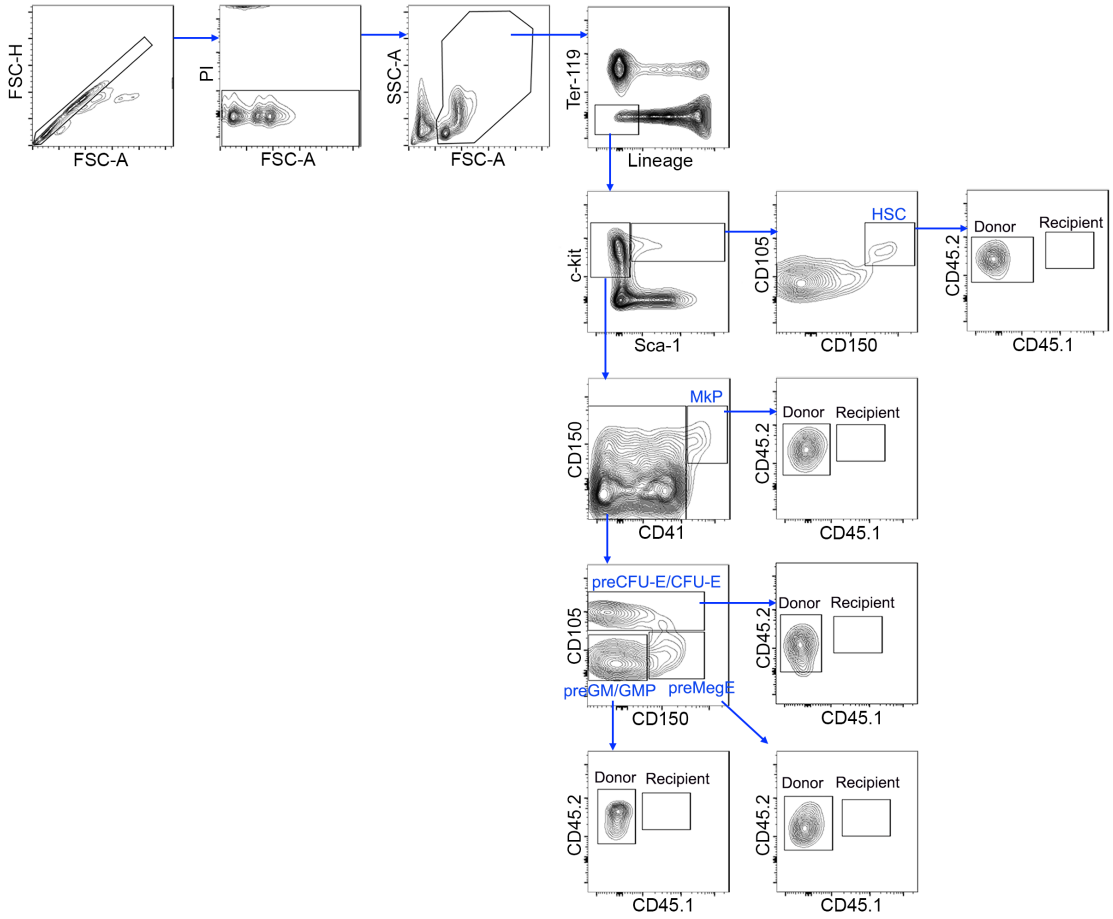
Supplementary Figure S2. Comparison of endogenous *Rps19* and transgene *RPS19* expressions in c-kit^{+} BM cells isolated from D/D and +/+ mice on day 4 after transduction (average of 3 independent experiments with 3 technical replicates in each group, error bars represent the SD, *p<0.05, **p<0.01, ***p<0.001 by student's t-test)



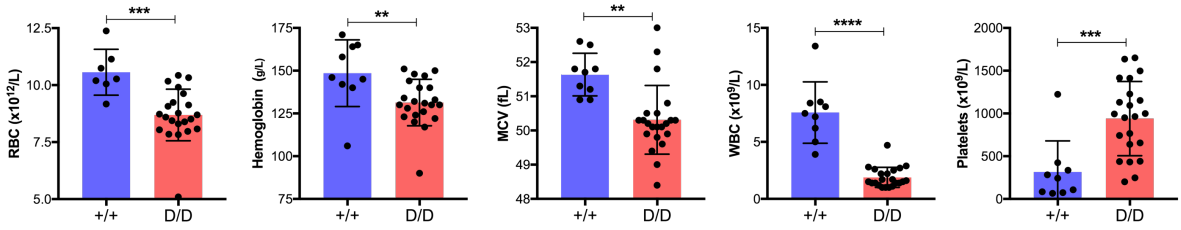
Supplementary Figure S3. Normal blood cellularity and absence of recipient-derived hematopoiesis in recipients receiving uninduced gene-corrected cells or mock transduced cells at 2 months after transplantation (n=13-16, MCV: mean corpuscular volume)



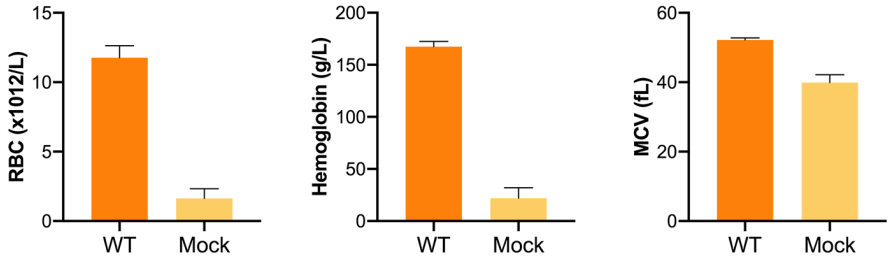
Supplementary Figure S4. Inhibition of mouse *Rps19* expression in the mock and gene-corrected cells induced with doxycycline for 16 weeks (n=3-6 in each group, error bars represent the SD, ***p*<0.01, ****p*<0.001 by one-way ANOVA)



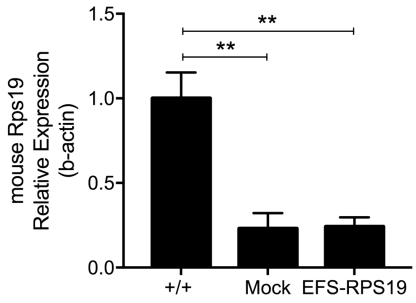
Supplementary Figure S5. The applied FACS strategy allowing the fraction of myeloerythroid progenitors and erythroid precursors



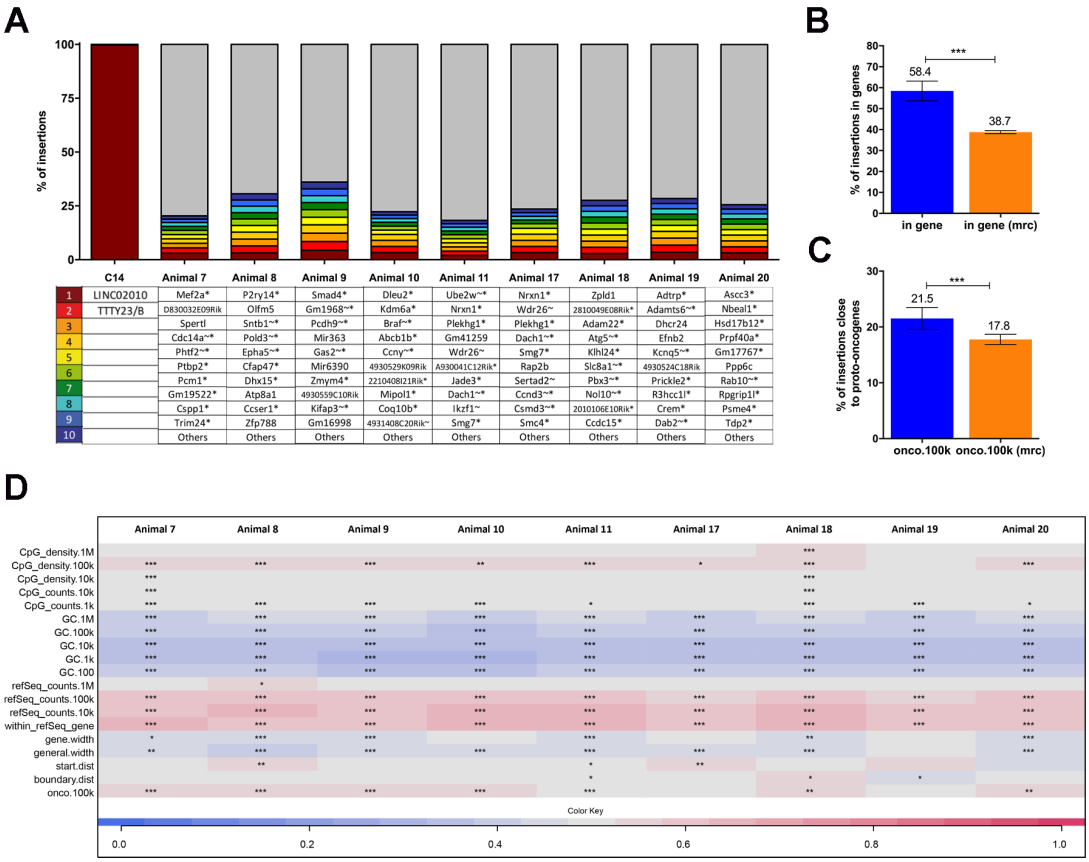
Supplementary Figure S6. Anemia phenotype in *Rps19*-deficient mice receiving doxycycline administration for 1 week (n=9 for +/+ group, n=22 for D/D group, ***p*<0.01, ****p*<0.005, *****p*<0.001 by student's t-test. RBC: red blood cell; MCV: mean corpuscular volume; WBC: white blood cell)



Supplementary Figure S7. Mice in the mock group had to be sacrificed due to severe anemia within 2-3 weeks after transplantation in the *Rps19*-deficient cell transplantation model (RBC: red blood cell; MCV: mean corpuscular volume)



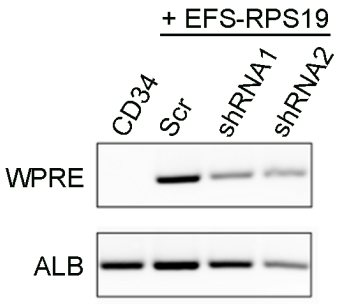
Supplementary Figure S8. Inhibition of mouse *Rps19* expression in the mock and vector-treated *Rps19*-deficient cells induced with doxycycline for 16 weeks (n=4-6 in each group, error bars represent the SD, ***p*<0.01 by one-way ANOVA)



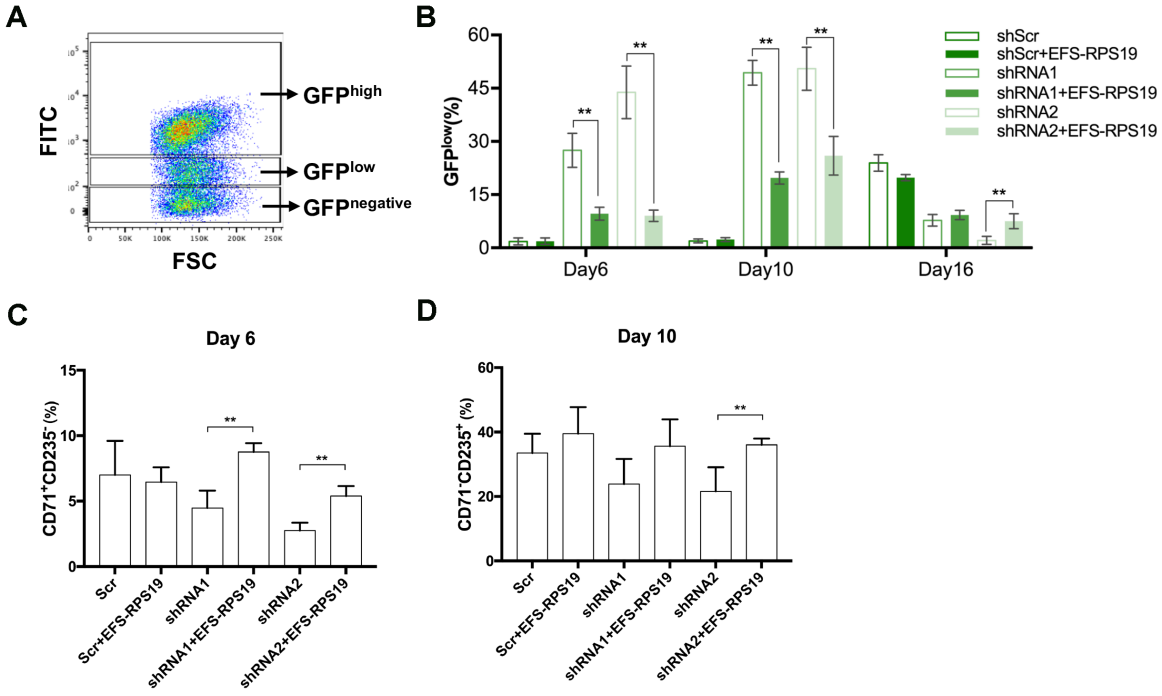
Supplementary Figure S9. Gene-corrected *Rps19*-deficient BM cells show a vector integration pattern that indicates low risk of mutagenesis and a highly polyclonal insertion site pattern

(A) Top 10 integration sites in each sample (*indicates that the integration was within a transcription unit, ~ indicates that the insertion was within 50 kb of a cancer-related gene). (B-C) Percent of all integrations inside of transcriptional units (B) and percent of integrations within 100 kb of proto-oncogenes compared to matched random control sites (C). (D) Genomic heatmap analysis of insertion site profile. (mrc: matched random control, ***$p < 0.001$ by unpaired t-test).

Supplementary Figure S10. Venn diagram displaying the shared common insertion sites in cohorts 1+2 and cohorts 3+4. In case the integration overlapped with a proto-oncogene listed in the AllOnco database, the gene symbol is marked in red, in case it was described as a previous common insertion site it is marked in orange, if both parameters apply, the gene symbol is marked in purple.



Supplementary Figure S11. Successful integration of the vector into human *RPS19*-deficient CD34^{+} cord blood cells by PCR array (untreated CD34^{+} cord blood cells were regarded as control; Human *ALB* was used as an internal reference gene)



Supplementary Figure S12. Impaired erythroid differentiation of *RPS19*-deficient CD34$^+$ cord blood cells

(A) The applied FACS strategy according to GFP intensity. (B) Percentage of GFP^{low} populations in *RPS19*-deficient CD34$^+$ cells treated with or without EFS-RPS19 during erythroid differentiation from stage I to stage III. (C) Percentage of indicated cell outputs of GFP^{high} populations on day 6. (D) Percentage of indicated cell outputs of GFP^{high} populations on day 10 (data are shown as mean±SD, $^{**}p<0.01$ by student's t-test, three independent experiments).

Supplementary Tables

For qRT-PCR		**5' to 3'**
human codon-optimized *RPS19*	F	AAGAAAAGCGGCAAACTCAAG
	R	CCGTAGATCTTGGTCATGCT
human *RPS19*	F	GCCTGGAGTTACTGTAAAAGACG
	R	CCCATAGATCTTGGTCATGGAGC
human *ACTIN*	F	AGAAAATCTGGCACCACACC
	R	GGGGTGTTGAAGGTCTCAAA
mouse *Rps19*	F	GCAGAGGCTCTAAGAGTGTGG
	R	CCAGGTCTCTCTGTCCCTGA
mouse *Actin*	F	ATGGTGGGAATGGGTCAGAA
	R	CCATGTCGTCCCAGTTGGTAA
WPRE	F	TTCTGGGACTTTCGCTTTCC
WPRE	R	CCGACAACACCACGGAATTA
Taqman PROBE		ATCGCCACGGCAGAACTCATCG
For RT-PCR		**5' to 3'**
WPRE	F	GAGGAGTTGTGGCCCGTTGT
WPRE	R	TGACAGGTGGTGGCAATGCC
mouse *Actin*	F	GCTAATGAGGCTGGTGATAAGTGG
	R	CACGCTCGGTCAGGATCTTCAT
human *ALB*	F	TGAAACATACGTTCCCAAAGAGTTT
	R	CTCTCCTTCTCAGAAAGTGTGCAT
shRNA sequences		**5' to 3'**
human *RPS19* shRNA-1		GCACAAAGAGCTTGCTCCC
human *RPS19* shRNA-2		GAGATCTGGACAGAATCGC
Scramble		GACACGCGACTTGTACCAC

Supplementary Table S1. Sequences for qRT-PCR, RT-PCR and shRNAs (F: Forward; R: Reverse; ALB: ALBUMIN)

SID#	Sample Name	Unique Sites	Chapman	Coverage	S.chao1	Gini	Shannon	UC50
1	HD2 Clone	7	8	58%	5	0.705	0.897	1
2	Animal 5	882	887	58%	1411	0.840	5.025	32
3	Animal 6	878	856	61%	1963	0.874	4.732	17
4	Animal 7	987	976	64%	1506	0.884	4.523	14
5	Animal 8	1304	1411	48%	3108	0.889	4.990	31
6	Animal 13	933	922	79%	2164	0.873	4.727	18
7	Animal 14	957	934	68%	1359	0.872	4.883	23
8	Animal 15	1115	1083	67%	1670	0.839	5.287	44
9	Animal 16	985	979	62%	1559	0.884	4.757	22

Supplementary Table S2. Analysis of the pool size (Chapman, S.chao1), clonality (Gini) and diversity (Shannon) of gene-corrected BM cells from cohorts 1+2

SID#	Sample Name	Unique Sites	Chapman	Coverage	S.chao1	Gini	Shannon	UC50
1	C14	2	2	67%	3	0.497	0.023	1
2	Animal 7	248	234	89%	610	0.604	4.880	35
3	Animal 8	317	294	83%	999	0.758	4.603	22
4	Animal 9	275	253	83%	1187	0.758	4.412	15
5	Animal 10	301	282	86%	680	0.715	4.745	30
6	Animal 11	532	501	84%	868	0.734	5.219	41
7	Animal 17	338	313	83%	934	0.699	4.903	34
8	Animal 18	346	326	88%	734	0.726	4.786	23
9	Animal 19	296	274	79%	634	0.771	4.473	21
10	Animal 20	309	290	83%	662	0.727	4.708	24

Supplementary Table S3. Analysis of the pool size (Chapman, S.chao1), clonality (Gini) and diversity (Shannon) of gene-corrected *Rps19*-deficient BM cells from from cohorts 3+4

Gene Symbol	**Animal 5**	**Animal 6**	**Animal 7**	**Animal 8**	**Animal 13**	**Animal 14**	**Animal 15**	**Animal 16**	**Overlaps**
Lrrc4c	●	●	●	●	●	●	●	●	8
Mir101c	●	●	●	●	●	●	●	●	8

Ncam2	●	●	●	●	●	●	●	●	8
Vegfc	●	●	●	●	●	●	●	●	8
Cylc2	●	●		●	●	●	●	●	7
Dach1	●	●	●	●	●	●	●		7
Dlg2	●		●	●	●	●	●	●	7
Pcmtd1	●	●	●	●		●	●	●	7
Phtf2	●	●	●	●		●	●	●	7
Rgs18		●	●	●	●	●	●	●	7
Tfec	●	●	●	●		●	●	●	7
Trps1	●	●	●	●	●	●		●	7
1700019E08Rik	●		●	●	●	●	●		6
1700019M22Rik	2.1%	●	●	●		●	●		6
4930529K09Rik	●	●	●	●		●	●		6
AA545190	●		●	●	●	●		●	6
Alcam	●	●	●		●		●	●	6
Angpt1			●	●	●	●	●	●	6
Arhgap18	●	●	●	●			●	●	6
Ascc3	●			●	●	●	●	●	6
Caap1	●	1.1%	●	●		●	●		6
Cntnap2	●		●	●		●	●	●	6
Csmd3	●	●	●	●			●	●	6
Ddi1	●		●	●	●		●	●	6
Diaph2	●		●		●	●	●	●	6
Epha7	2.0%	●		●	●	●	●		6
Gm11917		●	●		●	●	●	●	6
Gm20125	●		●	●	●	●	●		6
Gm20756	●	●		●	●		●	●	6
Gm35496	●	●	●	●			●	●	6
Gm6578			●	●	●	●	●	●	6
Hgf	●	●		●	●	●		●	6
Ikzf2	1.0%		●	●	●	●	●		6
Inpp4b		●	●	●	●	●	●		6
Gene Symbol	**Animal 5**	**Animal 6**	**Animal 7**	**Animal 8**	**Animal 13**	**Animal 14**	**Animal 15**	**Animal 16**	**Overlaps**
Kdm6a	●			●	●	●	●	●	6
Lnpep		●	●	●		●	●	●	6
Lrrn3		●	●	●	●	●	●		6
Mir6368		●	●	●	●	●	●		6
Nrxn1		●	●	●		●	●	●	6
Pla2g4a	●		●	●	●	●	●		6
Rab38	●	●	●	●			●	●	6
Stau2	●	●	●	●	●	●			6
Tbc1d5	●	●	●		●	●	●		6

Gene Symbol	Animal 5	Animal 6	Animal 7	Animal 8	Animal 13	Animal 14	Animal 15	Animal 16	Overlaps
Wac	●	●	●	●	●			●	6
Wapl	●	●		●	●		●	●	6
1700128A07Rik	●	●	●			●	●		5
4930559C10Rik		●	●	●		●		●	5
4933422A05Rik	●	●	●			●		●	5
A930001A20Rik	●	●		●			●	●	5
Adamts6	●	●		●		●		●	5
Adgrl3	●	●	●	●			●		5
Akap13		●	●	●		●		●	5
Arap2	●	●	●	●				●	5
Arhgap6		●	●		●	●	●		5
Asxl2	●	●		●	●			●	5
B3galt2			●	●		●	●	●	5
Ccser2	●		●	●		●		●	5
Cdk14	●	●		●	●		●		5
Cdk8			●	●	●		●	●	5
Celf2		●	●	●		●	●		5
Cfap47	●		●	●			●	●	5
Dcun1d5			●	●	●		●	●	5
Ddx10	●	●	2.8%	●			●		5
Fam107b	●		●	●		●	●		5
Fam174a	●	●	●		●	●			5
Gm19782	●		●	●	●	●			5
Gm6634			●	●		●	●	●	5
Gria3	●		●	●	●			●	5
Hs3st1		●	●	●			●	●	5
Mctp2	●	1.1%		●	●			●	5
Mef2c	●		●	●	●		●		5
Micu3			●	●	●		●	●	5
Mir1931	●		●		●	●	●		5
Mir3961	●		●	●		●		●	5
Mir6350	●	●		●	●			●	5
Mir6411	2.4%		●		●	●		●	5
Mllt10	●	●	●	●	●				5
Mmp16	●		●	●			●	●	5
Myo10		●	●	●		●		●	5
Nipbl		●	●	●			●	●	5
Nudt12		●	●	●	●		●		5
Pcdh7	●		●	●			●	●	5
Pisd-ps3	●			●	●	●	●		5
Plxdc2	●			●	●		●	●	5

Prex2		●	●	●	●		●		5
Rab10os	●		●			●	●	●	5
Rasa1	●	●		●		●		●	5
Rb1cc1		●	●		●		●	●	5
Rnpc3	●	●			●	●	●		5
Sema3a	●	●	●			●	●		5
Sfi1	●		●	●		●	●		5
Shank3	●				●	1.9%	●	●	5
Shprh		2.9%	●	●	●	●			5
Slc4a7	●	●	●			●		●	5
Smurf2		●		●	●		●	●	5
Spata6	●		●	●			●	●	5
Tab2	●	●	●	●		●			5
Tbl1xr1	●			●	●		●	●	5
Tlr4	●	●	1.1%	●			●		5
Ttc37		●		●	●	●	●		5
Ube3a	●			●	●	●		●	5
Utrn	●	●	●		●	●			5
Vav3	●		●	●			●	●	5
Vwc2		●	●	●		●	●		5
Zmym4			●	●		●	●	●	5

Supplementary Table S4. Common insertion sites in or near the same genes in cohort 1 (animal 5-8) and cohort 2 (animal 13-16) of mice receiving gene-corrected BM cells from uninduced donors. In case the integration overlapped with a proto-oncogene listed in the AllOnco database, the gene symbol is marked in red, in case it was described as a previous common insertion site it is marked in orange, if both parameters apply, the gene symbol is marked in purple.

Gene_Symbol	**Animal 7**	**Animal 8**	**Animal 9**	**Animal 10**	**Animal 11**	**Animal 17**	**Animal 18**	**Animal 19**	**Animal 20**	**Overlaps**
Pisd-ps3	●	●	●	●	●		●	●	●	8
Diaph3	●	●	●			●	●	●	●	7
Lrrc4c	●	●			●	●	●	●	●	7
Nrxn1		●	●		●	●	●	1.9%	●	7
1700019M22Rik	●			1.4%	●		●	●	●	6
Trps1	●		●	●	●		●	●		6
Agbl1	●			●	●		●	●		5
Epha5		3.2%		1.5%		●	●	●		5
Hgf	●	●		●	1.5%			●		5
Kdm6a	●	●		2.9%	●			●		5

Lnpep		•		•	•		•	•		5
Lrrtm4	•				•		•	•	•	5
Mef2c			•	•	•			1.2%	•	5
Mir101c	•	•					•	•	•	5
Ncam2				•	•		•	•	1.8%	5
Tmem168		•			•	•		•	•	5

Supplementary Table S5. Common insertion sites in or near the same genes in cohort 3 (animal 7-11) and cohort 4 (animal 17-20) of mice receiving gene-corrected BM cells from doxycycline-induced donors. In case the integration overlapped with a proto-oncogene listed in the AllOnco database, the gene symbol is marked in red, in case it was described as a previous common insertion site it is marked in orange, if both parameters apply, the gene symbol is marked in purple.