OMTM, Volume *9*

Supplemental Information

Preclinical Development of a Lentiviral Vector for Gene Therapy of X-Linked Severe Combined Immunodeficiency

Valentina Poletti, Sabine Charrier, Guillaume Corre, Bernard Gjata, Alban Vignaud, Fang Zhang, Michael Rothe, Axel Schambach, H. Bobby Gaspar, Adrian J. Thrasher, and Fulvio Mavilio

Supplemental Figures and Legends, Tables, Methods, and References

Supplemental Figures

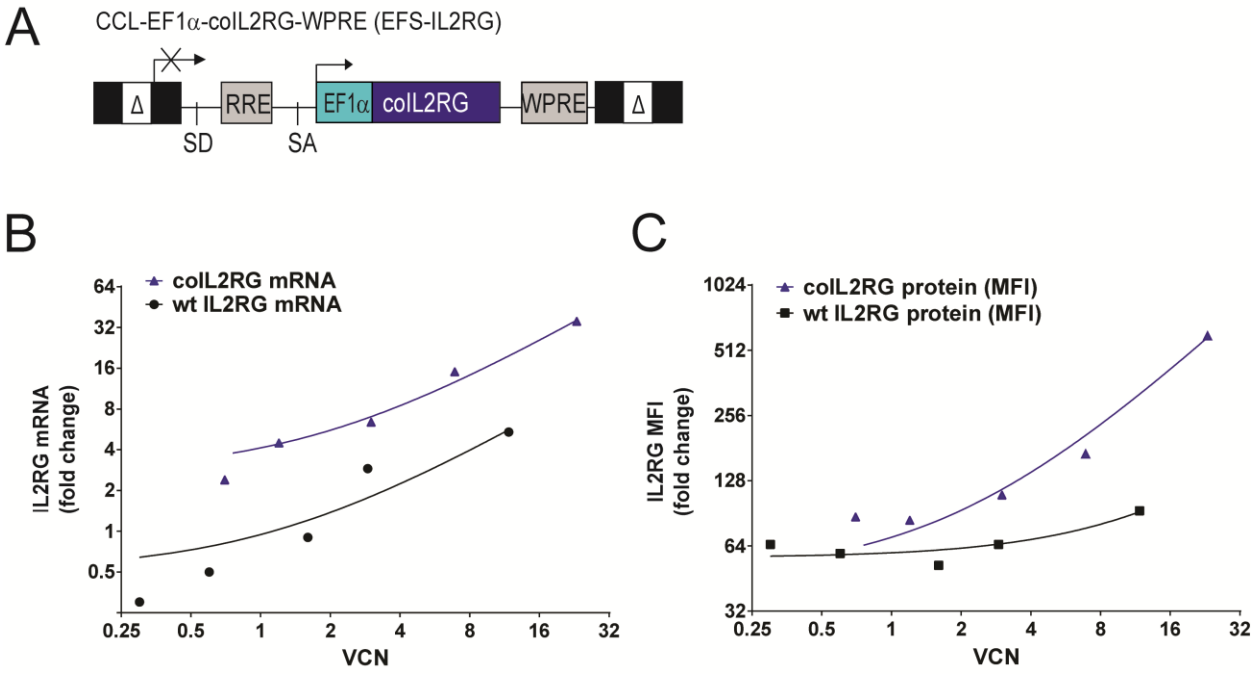

Figure S1. **A.** Scheme of the SIN lentiviral vector containing deleted LTRs (Δ), the short EF1α promoter (EF1α), the codon-optimized version of the IL2RG (coIL2RG), the HIV-1 Rev Responsive Element (RRE), the HIV-1 splicing donor (SD) and acceptor (SA) signals, and a Woodchuck Post-transcriptional Response Element (WPRE). **B.** Relative quantification of the IL2RG mRNA by RT-qPCR (on the left) and protein by flow cytometry (on the right) driven by the codon-optimized (coIL2RG) or wild type (wt) sequence in transduced ED7R cells. Relative mRNA and protein quantification is expressed as fold change with respect to the endogenous IL2RG in Jurkat cells. VCN, Vector Copy Number; MFI, Mean Fluorescence Intensity.

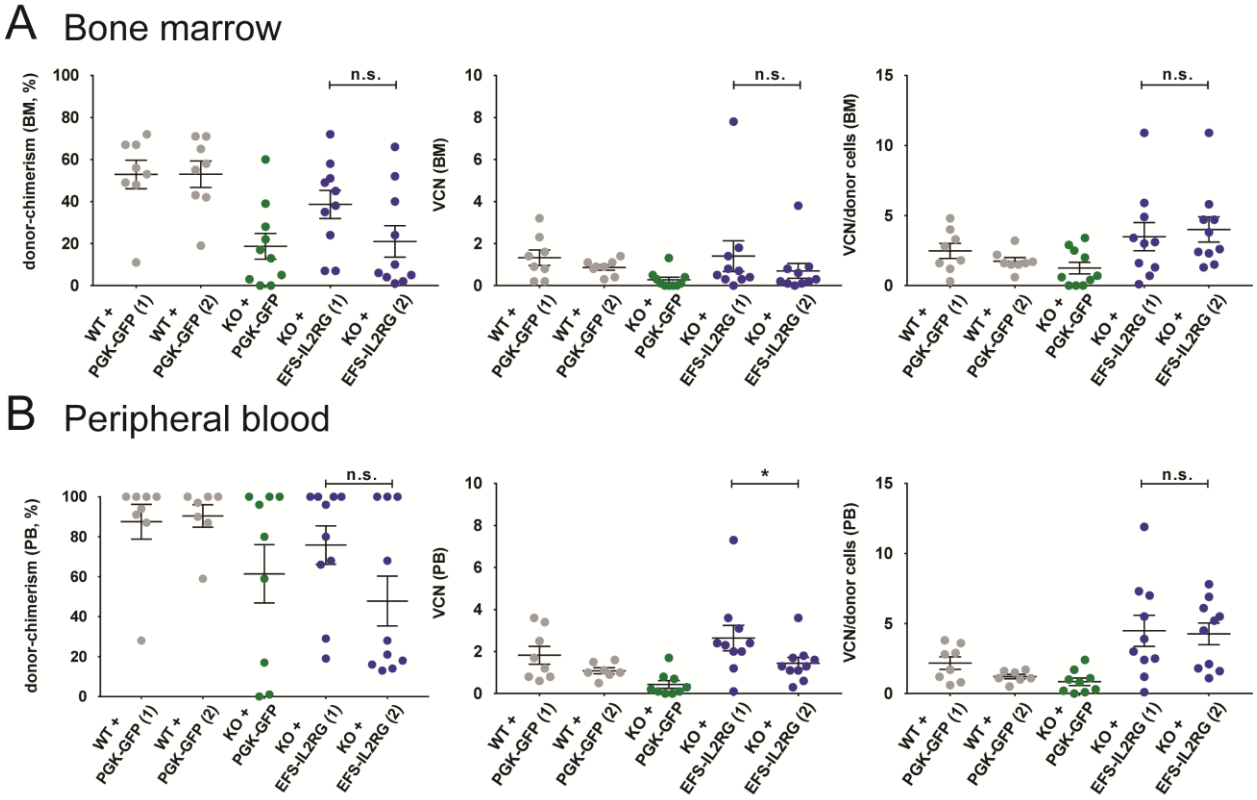

Figure S2. **Transplantation of HSPCs transduced once or twice with the EFS-IL2RG in Il2rg-deficient mice.** Chimerism, VCN and VCN/donor cell in the bone marrow (up) and peripheral blood (down) of *Rag2*$^{-/-}$/*Il2rg*$^{-/-}$ mice transplanted with wild-type C56Bl6 (WT) or *Il2rg*$^{-/-}$ (KO) Lin^{-} cells transduced once (1) or twice (2) with the EFS-IL2RG vector (KO+EFS-IL2RG) or a PGK-GFP vector (WT+PGK-GFP, KO+PGK-GFP), six months after transplantation. Data are presented as individual animals, and as means ± SEM. Statistical differences are expressed as n.s., non-significant, or * $p<0.05$.

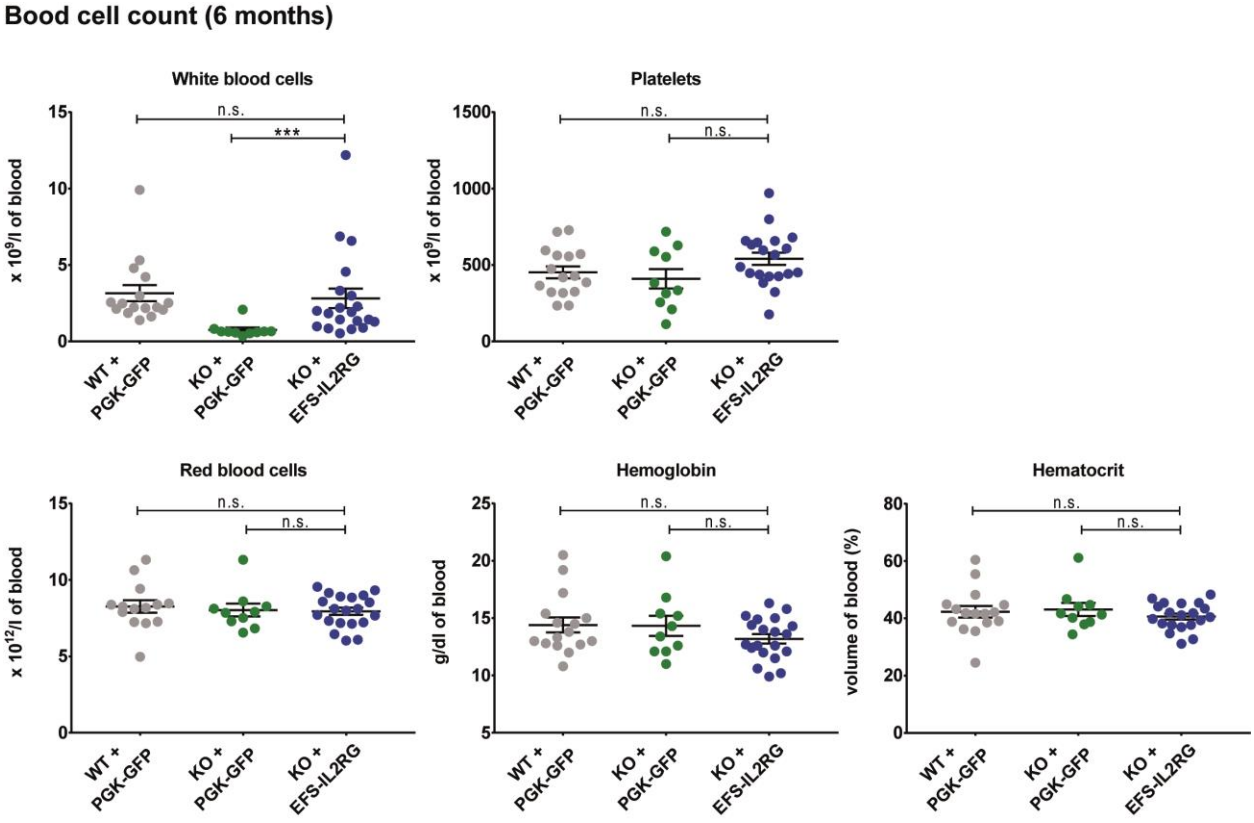

Figure S3. **Blood cell count in Il2rg-deficient mice after HSPC-transplantation.** Count of white blood cells, red blood cells and platelets, and quantification of hemoglobin and hematocrit in the blood of *Rag2*$^{-/-}$/*Il2rg*$^{-/-}$ mice transplanted with wild-type C56Bl6 Lin^{-} cells transduced with a PGK-GFP vector (WT+PGK-GFP), *Il2rg*$^{-/-}$ Lin^{-} cells transduced with the PGK-GFP vector (KO+PGK-GFP), and *Il2rg*$^{-/-}$ Lin^{-} cells transduced with the EFS-IL2RG vector (KO + EFS-IL2RG), six months after transplantation. Data are presented as individual animals, and as means ± SEM. Statistical differences are expressed as n.s., non-significant, or *** $p<0.001$.

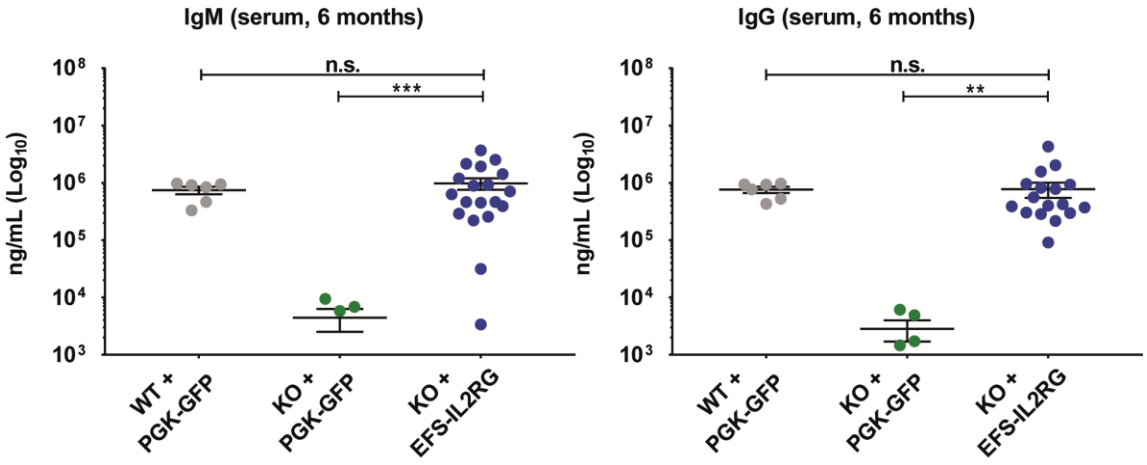

Figure S4. **Serum immunoglobulin in Il2rg-deficient mice after HSPC-transplantation.** Level of IgM and IgG in the serum of *Rag2*$^{-/-}$*/Il2rg*$^{-/-}$ mice transplanted with wild-type C56Bl6 Lin^{-} cells transduced with a PGK-GFP vector (WT+PGK-GFP), *Il2rg*$^{-/-}$ Lin^{-} cells transduced with the PGK-GFP vector (KO+PGK-GFP), and *Il2rg*$^{-/-}$ Lin^{-} cells transduced with the EFS-IL2RG vector (KO + EFS-IL2RG), six months after transplantation. Data are presented as individual animals, and as means ± SEM. Statistical differences are expressed as n.s., non-significant, ** $p<0.01$, *** $p<0.001$.

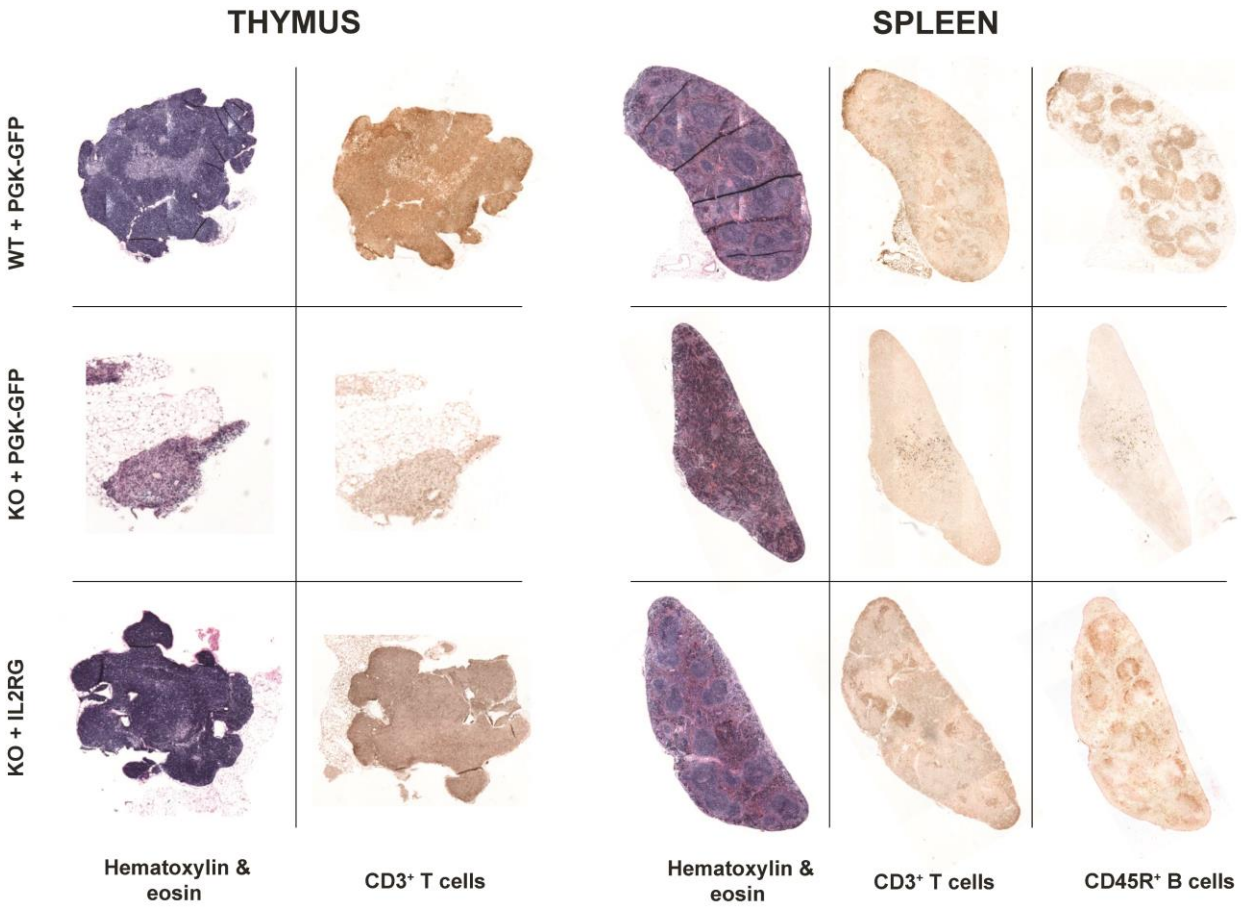

Figure S5. **Histopathological analysis of thymus and spleen of Il2rg-deficient mice after HSPC-transplantation.** Staining with Hematoxylin and Eosin, and antibodies against CD3 and CD45R antigens in representative thymus (left) and spleen (right) sections of $Rag2^{-/-}/Il2rg^{-/-}$ mice transplanted with wild-type C56Bl6 Lin^- cells transduced with a PGK-GFP vector (WT+PGK-GFP), *Il2rg*$^{-/-}$ Lin^- cells transduced with the PGK-GFP vector (KO+PGK-GFP), and *Il2rg*$^{-/-}$ Lin^- cells transduced with the EFS-IL2RG vector (KO + EFS-IL2RG), six months after transplantation.

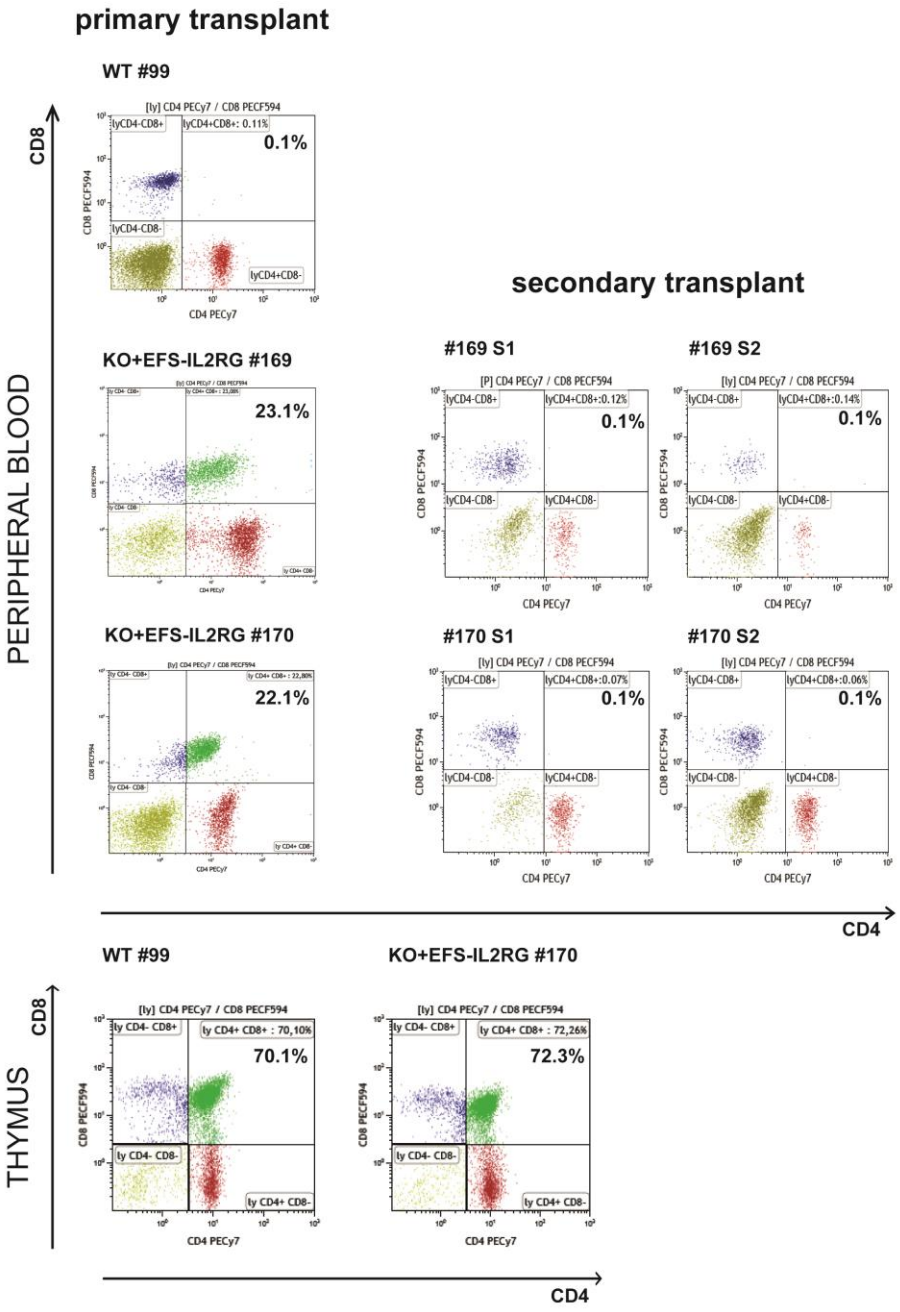

Figure S6. **T-cell phenotype in peripheral blood and thymus of mice #169 and #170.** Flow cytometry analysis of the expression of CD4 (x axis) and CD8 (y axis) in peripheral blood T cells from the *Rag2*$^{-/-}$*Il2rg*$^{-/-}$ mice #169 and #170 (upper panel, left), transplanted with *Il2rg*$^{-/-}$ Lin- cells transduced with the EFS-IL2RG vector (KO+EFS-IL2RG), and of a control mouse transplanted with wt cells (#99). The upper right panels show the same analysis in two mice (S1 and S2) secondarily transplanted with the BM from mice #169 and #170. The bottom panel shows the flow cytometry analysis of the thymus from mouse #170 and from the control mouse #99.

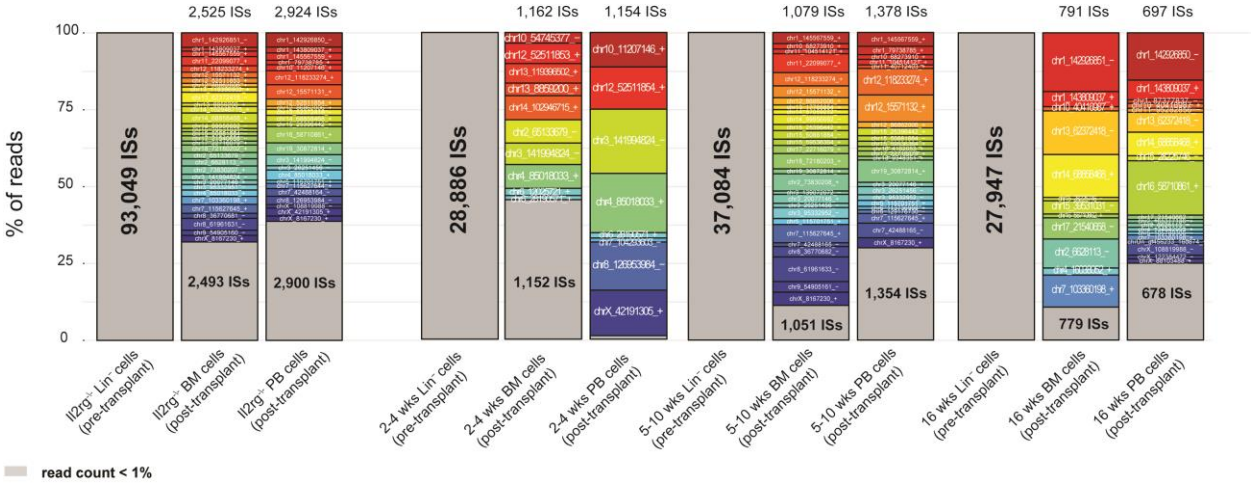

Figure S7. **Integration profile of the EFS-IL2RG vector in individual transduction groups.** ISs were retrieved by LM-PCR and Illumina high-throughput sequencing in pre-transplant Lin$^-$ cells and post-transplant BM and PB of *Rag2*$^{-/-}$/*Il2rg*$^{-/-}$ mice transplanted with and *Il2rg*$^{-/-}$ Lin$^-$ cells transduced with the EFS-IL2RG vector (see Figure 1). The histograms show the percentage of sequence reads associated to each IS in the pre- and post-transplant BM and PB of mice belonging to each transduction group as pools (2-4 wks, 5-10 wks and 16 wks; right) or the merge of them (left). In each bar, the total number of ISs retrieved is indicated inside (pre-transplant samples) or at the top (post-transplant samples), the fraction of ISs with read count <1% is shown in gray and the corresponding number of ISs is indicated inside the box, while ISs with read count >1% are shown in different colors and identified by their genomic coordinates.

Supplemental Tables

Table S1. RT-qPCR analysis of *Mecom* gene expression in transduced cell clones. The table shows the gene expression ratios of Mecom/β-actin for the different expanded clones as calculated by the ΔΔ-Ct-method. The 141203-c1 (highlighted in grey) served as a reference for comparison of fold up- or downregulation. (na, not analyzed).

Sample ID	**Vector**	**CT β-Actin**	**CT Mecom**	**ΔCt**	**ΔΔCt**
141203-c1	Mock	22.88	27.52	4.00E-02	1.00
141203-c2	Mock	22.83	26.79	6.44E-02	1.61
141203-c3	Mock	na	na	na	na
141203-c4	RSF91	29.27	30.71	3.68E-01	9.21
141203-c5	RSF91	na	na	na	na
141203-c12	EFS-IL2RG	18.48	24.89	1.17E-02	0.29
141203-c13	EFS-IL2RG	18.20	26.64	2.88E-03	0.07
141203-c14	EFS-IL2RG	18.91	27.22	3.15E-03	0.08
141203-c15	EFS-IL2RG	21.37	31.12	1.17E-03	0.03
141203-c16	EFS-IL2RG	na	na	na	na
141203-c17	EFS-IL2RG	20.78	29.98	1.70E-03	0.04

Table S2. Gene Ontology categories enriched in the pre-transplant *Il2rg*$^{-/-}$ Lin^{-} cells (all samples). Only categories with a p-value ≤0.05 after Bonferroni correction for false discovery rate are considered significantly enriched.

GO identifier	GO term	p-value	Bonferroni-corrected p-value
GO:0044267	cellular protein metabolic process	2.2E-40	7.5E-37
GO:0006464	protein modification process	5.2E-23	1.8E-19
GO:0016070	RNA metabolic process	1.6E-21	5.4E-18
GO:0006259	DNA metabolic process	5.9E-20	2.0E-16
GO:0030163	protein catabolic process	6.3E-19	2.1E-15
GO:0043632	modification-dependent macromolecule catabolic process	1.9E-18	6.4E-15
GO:0044257	cellular protein catabolic process	5.0E-18	1.7E-14
GO:0006396	RNA processing	1.2E-15	3.8E-12
GO:0015031	protein transport	1.5E-15	4.9E-12
GO:0006281	DNA repair	5.1E-14	1.8E-10
GO:0016071	mRNA metabolic process	8.6E-13	2.9E-09
GO:0000279	M phase	2.9E-12	9.9E-09
GO:0016310	phosphorylation	3.3E-12	1.1E-08
GO:0016568	chromatin modification	4.7E-12	1.6E-08
GO:0006397	mRNA processing	2.5E-10	8.6E-07
GO:0007242	intracellular signaling cascade	4.0E-10	1.4E-06
GO:0008380	RNA splicing	4.3E-10	1.5E-06
GO:0000087	M phase of mitotic cell cycle	6.9E-10	2.3E-06
GO:0007067	mitosis	1.8E-09	6.1E-06
GO:0006886	intracellular protein transport	3.1E-07	1.1E-03
GO:0034660	ncRNA metabolic process	5.4E-07	1.8E-03
GO:0051056	regulation of small GTPase mediated signal transduction	6.6E-07	2.2E-03
GO:0009966	regulation of signal transduction	8.4E-07	2.9E-03
GO:0006310	DNA recombination	8.5E-07	2.9E-03
GO:0006260	DNA replication	1.0E-06	3.4E-03
GO:0051246	regulation of protein metabolic process	6.7E-06	2.3E-02
GO:0043087	regulation of GTPase activity	1.1E-05	3.6E-02
GO:0050657	nucleic acid transport	1.1E-05	3.7E-02
GO:0050658	RNA transport	1.1E-05	3.7E-02
GO:0016311	dephosphorylation	1.2E-05	3.9E-02
GO:0030097	hemopoiesis	1.4E-05	4.8E-02

Table S3. KEGG pathways enriched in the pre-transplant *Il2rg*$^{-/-}$ Lin^{-} cells (all samples). Only categories with a p-value ≤0.05 after Bonferroni correction for false discovery rate are considered significantly enriched.

GO identifier	GO term	p-value	Bonferroni-corrected p-value
mmu04120	Ubiquitin mediated proteolysis	7.9E-09	1.5E-06
mmu03040	Spliceosome	2.0E-07	3.9E-05
mmu04144	Endocytosis	7.8E-07	1.5E-04
mmu04070	Phosphatidylinositol signaling system	1.9E-06	3.6E-04
mmu04520	Adherens junction	6.5E-06	1.3E-03
mmu04660	T cell receptor signaling pathway	6.8E-06	1.3E-03
mmu04662	B cell receptor signaling pathway	2.0E-05	3.9E-03
mmu05211	Renal cell carcinoma	2.6E-05	5.1E-03
mmu04510	Focal adhesion	3.1E-05	6.1E-03
mmu00230	Purine metabolism	4.7E-05	9.0E-03
mmu05221	Acute myeloid leukemia	5.6E-05	1.1E-02
mmu04722	Neurotrophin signaling pathway	6.3E-05	1.2E-02
mmu04666	Fc gamma R-mediated phagocytosis	7.3E-05	1.4E-02
mmu04110	Cell cycle	1.2E-04	2.3E-02
mmu04114	Oocyte meiosis	1.2E-04	2.3E-02
mmu04910	Insulin signaling pathway	1.3E-04	2.6E-02
mmu04012	ErbB signaling pathway	1.5E-04	2.8E-02
mmu04720	Long-term potentiation	1.7E-04	3.2E-02
mmu04810	Regulation of actin cytoskeleton	1.7E-04	3.2E-02
mmu05213	Endometrial cancer	2.3E-04	4.4E-02

Table S4. List of human and mouse orthologue gene (NCBI HomoloGene definition) over-targeted at statistically significant frequency with respect to random ($p < 0.001$) by the EFS-IL2RG vector in human mobilized peripheral blood (mPB) $CD34^+$ and murine Lin^- hematopoietic stem/progenitor cells. Genes are ranked by the number of integrations sites (ISs) retrieved from the two cell types. Genes found frequently targeted in pre- and/or post-transplant hematopoietic cells in LV-based gene therapy clinical trials (see main text) are indicated in bold. The only gene found in common between the two lists (ARAP2) is indicated in red.

HomoloGene Symbol (human)	# ISs in mPB $CD34^+$ cells	HomoloGene Symbol (mouse)	# ISs in wt Lin^- cells
KDM2A	61	*Dach1*	40
PACS1	47	*Ascc3*	33
GPATCH8	47	*Trps1*	31
FCHSD2	45	*Tbc1d5*	31
ASH1L	41	*Stag1*	31
EIF4G3	38	*Diaph3*	31
EHMT1	35	*Epha7*	27
NSD1	32	*Zfp407*	26
NFAT5	30	*Mycbp2*	24
WNK1	30	*Kansl1l*	24
RBM6	30	*Cdc73*	23
ZZEF1	29	*Dmxl1*	23
TNRC6B	29	*Rb1cc1*	22
IP6K1	28	*Smarcad1*	21
NPLOC4	27	*Wrn*	21
PBX3	27	*Vps13a*	21
PPP6R2	27	*Phip*	21
SBF2	27	*Spata5*	21
CCDC57	27	*Phf14*	20
GRB2	26	*Adgrl4*	20
DLG1	26	*Cdk13*	20
STAT5B	26	*Lyst*	19
SMG1	26	*Ube3a*	19
ARAP2	25	*Stxbp6*	19
RPTOR	25	*Lnpep*	19
NF1	24	*Kdm4c*	19
SPATS2	24	*Hgf*	18
VMP1	24	*Pot1a*	18
SETD2	24	*Rev3l*	18
FNBP1	24	*Pcm1*	17
UNK	22	*Hjurp*	17
SMG6	22	*Sfi1*	17
TNRC6C	22	*Pla2g4a*	17
RTN3	22	*Chd6*	17
NFATC3	22	*Rbm26*	17
RFX2	22	*Rttn*	17
ANKRD11	22	*Arap2*	16
FOXJ3	21	*Pde3b*	16
SIK3	21	*Rasa1*	16
CAPN1	20	*Tsc22d2*	16
BLM	20	*Wwp1*	16
SMARCC1	20	*Trip12*	16
UBE2G1	19	*Cnot6l*	16
VAV1	19	*Slc4a7*	15
MUM1	19	*Uba6*	15

RABL6	19	*2310035C23Rik*	15
TAOK1	19	*Wls*	15
METTL16	19	*Cnot4*	15
MROH1	19	*Arhgap5*	14
CKAP5	18	*Uri1*	14

Supplemental Materials and Methods

List of primers and probes used for qPCR (VCN, chimerism and Il2RG mRNA evaluation). The suffix h or m indicate human or mouse genome.

Oligo. name	Type	Sequence
HIV-Psi	forward	5'-CAGGACTCGGCTTGCTGAAG-3'
HIV-Psi	reverse	5'-TCCCCCGCTTAATACTGACG-3'
HIV-Psi	probe	5'-CGCACGGCAAGAGGCGAGG-3'
WPRE	forward	5'-AGGAGTTGTGGCCCGTTGT-3'
WPRE	reverse	5'-TGACAGGTGGTGGCAATGC-3'
WPRE	probe	5'-TGTTTGCTGACGCAACCCCCACT-3'
hIL2RG	forward	5'-CAGGAGACAGGCCACACAGA-3'
hIL2RG	reverse	5'-CACTCAGTTTGTGAAGTGTTAGGTTCT-3'
hIL2RG	probe	5'-CTAAAACTGCAGAATCTGGTGATCCCCTGG-3'
hTFIID	probe	5'-TGTGCACAGGAGCCAAGAGTGAAGA-3'
hTFIID	forward	5'-GAGAGCCACGAACCACGG-3'
hTFIID	reverse	5'-ACATCACAGCTCCCCACCAT-3'
hALB	forward	5-GCTGTCATCTCTTGTGGGCTGT-3'
hALB	reverse	5'-ACTCATGGGAGCTGCTGGTTC-3'
hALB	probe	5'-CCTGTCATGCCCACACAAATCTCTCC-3'
mY	probe	5'-TTCTCCAGGACCAGTGACTGGAGATCA-3'
mY	forward	5'-GTGCTAAGGAGTAGAGCGGAGAA-3'
mY	reverse	5'-CATGGTAACTGCTCAAGCGGT-3'
mTtn	probe	5'-TGCACGGAAGCGTCTCGTCTCAGTC-3'
mTtn	forward	5'-AAAACGAGCAGTGACGTGAGC-3'
mTtn	reverse	5'-TTCAGTCATGCTGCTAGCGC-3'
mIL2RG	forward	5'-TCGAAGCTGGACGGAACTAATAG-3'
mIL2RG	reverse	5'-CTCCGAACCCGAAATGTGTAC-3'
mIL2RG	probe	5'-TGAACCTAGATTCTCCCTGCCTAGTGTGGA-3'
mTFIID	probe	5'-TGTGCACAGGAGCCAAGAGTGAAGA-3'
mTFIID	forward	5'-ACGGACAACTGCGTTGATTTT-3'
mTFIID	reverse	5'-ACTTAGCTGGGAAGCCCAAC-3'

List of antibodies used for the immunophenotype analysis in the animal study

Antibody	Clone	Fluorochrome	Supplier	Reference	Isotype
CD3	145-2C11	APC	BD	533066	Ar Ham IgG1, κ
CD4	RM4-5	PECy7	BD	552775	Rat IgG2a, κ
CD8	53-6-7	ECD	BC	A88606	IgG2a kappa, Rat (LOU/Ws1/M)
IgD	11-26c	PE	southern-B	1120-09	Rat IgG2a, κ
IgM	1B4B1	APC	southern-B	1140-11	Rat IgG1κ
CD19	1D3	PECy7	BD	552854	Rat IgG2a, κ
NK1.1	PK136	PE	BD	557391	Ms IgG2a, κ
CD11b	M1/70	APC	BD	553312	Rat IgG2b, κ
TCRgd	GL3	PE	BD	553178	ArHam IgG2, κ
TRCab	H57-597	PECy7	BD	560729	Ar Ham IgG2, λ1
CD44	IM7	PE	BD	553134	Rat IgG2b, κ
CD25	3C7	APC	BD	558643	Rat IgG2b, κ

List of primers and probes used for LM-PCR. **NNNN** represent the 4-bp sample specific tags

Oligonucleotide	Sequence
linker-	(Phos)TAGTCCCTTAAGCGGAG(AmC7)
linker+	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC
HIV 3' LTR primer	AGTGCTTCAAGTAGTGTGTGCC
Linker primer	GTAATACGACTCACTATAGGGC
ILLU linker nested tag primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGGGCTCCGCTTAAGGGAC
ILLU H3LTRnest tag primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**NNNN**GTCTGTTGTGTGACTCTGGTAAC

Bioinformatics pipeline

The bioinformatics pipeline developed for IS identification was based on the SnakeMake workflow manager to call the different programs needed for the trimming, cleaning, filtering, mapping, annotation and reporting steps. R1 reads (forward) from 300bp Paired-end Illumina MiSeq sequencing were firstly demultiplexed by "cutadapt" (DOI:10.14806/ej.17.1.200) and a reference database containing anchored 5' tags (4 bp). We allowed one-mismatch tolerance over the 4 bp, considering that the minimal editing distance between tags is two nucleotides. R1 reads containing the correct tag were counter-selected if containing the provirus sequence or the plasmid sequence, otherwise they were trimmed by skewer for the sequences corresponding to the primer on the viral Long Terminal Repeat (LTR), to the remaining LTR (up to the last 3 nucleotides that are processed later) and the adaptor sequence. For each trimming step, we allowed a mismatch rate of 0.12, a minimal length of 20 bp after trimming, and a minimal match length equal to the sequence length to trim, except for the adaptor, for which the minimal matching length was set to 10 bp and the mismatch rate to 0.3, considering the low quality of the read at the 3' end. Output reads are then trimmed without tolerance for the very end of the LTR (the last three nucleotides GCA) in order to increase the accuracy of the mapping of the adjacent host genome. Then, a filter selected reads $\geq$ 20 bp and an average PHRED quality of 30. Trimmed reads are mapped by bowtie2 (configuration: -N 1 -L 25 -i S,25,0 --score-min L,0,-0.15 --gbar 10) on the GRCh37.75/hg19 (human) or GRCm38/mm10 (mouse) reference genome assembly. Unmapped reads are searched for Tru91restriction (TTAA) site and if present, the upstream sequence before that position is remapped. This step removes concatemeres of gDNA fragments that may have been ligated during the library preparation. Again, unmapped read are search for TTA motif, cut and remapped. All mapped outputs from these three steps are collapsed in bam/sam file and processed to remove low MAPQUAL alignments (<10), multi-hits alignments (XS bam tag), alignments with mismatches in the first 3 nucleotides (NM bam tag) and IS position is determined from chromosome, mapping position corrected for strand and CIGAR INDELS. Alignments mapping to the same position are collapsed together keeping track of the read count and IS falling in repeated regions are collapsed together to a single position. All IS are then annotated using a RefSeq modified table with consensus transcripts for each gene. Finally, a script performs a reporting file including IS position, genomic annotation, read count and percentage reads from parental step, averaged read length, quality composition, and other additional features.

Over-targeting gene definition

In order to estimate the over-targeting of a gene in pre-transplant Lin- cells, we compared the targeting frequency experimentally observed to a expected frequency given by a random dataset of IS. Because of the technical procedure used to recover the ISs, proviruses can only be detected if they are in close proximity to a Tru91 restriction site. Therefore, the less Tru91sites are in a genomic region, the less chances we have to retrieve ISs located there. For this reason, the targeting-frequency of a gene is also related to the number and distribution of the intragenic Tru91sites. So, instead of generating random virtual genomic positions in a arbitrary proximity to an Tru91site, we decided to use a corrected random dataset based on Tru91 restriction site coordinates, since every IS can be associated to its closest Tru91 site. For each sample, a number of random Tru91 sites equal to the number of experimental ISs was sampled with replacement 200 times from the whole set of ~19 million restriction site

coordinates (corresponding to 5,330,124 and 5,812,911 intragenic positions in the human and murine genome respectively). For each gene, we computed the average frequency of Tru91 sites detected in all replicates. We use a Poisson distribution with a mean equal to the average computed frequency and extract the probability to observe a higher frequency. Probabilities were corrected for multiple comparisons using the Bonferroni correction. Over-targeted genes were defined as genes associated to a Bonferroni-corrected p value ≤ 0.001 and at least 3 experimental ISs.

In post-transplant samples, we determined the differential enrichment in the targeting frequency of any gene, or in any GO or KEGG category, with respect to the pre-transplant datasets. Briefly, for each post-transplant sample, we calculated an expected targeting frequency for each gene by multiple (1,000) random sampling of a comparable number of integrations from the corresponding pre-transplant IS collection. The difference between observed and expected targeting frequency was associated to a p-value for each gene. For the GO or KEGG categories, the DAVID analysis on the list of targeted genes in the post-transplant samples was run using the corresponding pre-transplant lists as background.

In vitro immortalization assay (IVIM)

Mecom gene expression was analyzed in one cell clone transduced with IL2RG which was above the Q1-threshold in the IVIM assay. Cells were expanded from 96-well to 24-well plates for Mock (n=2), RSF91 (n=2) and IL2RG (n=6) assays. After 7 days, total mRNA was isolated with the Direct-zol™ RNA MiniPrep kit (Zymogene). 1 µg of total mRNA was reverse transcribed and analyzed by q-PCR for Mecom gene expression with the forward primer 5'-TGATCAGTGTCCCAAGGCATTT-3' and the reverse primer 5'-ATGTGTCGCTGAAGGTTGCT-3'. The Mecom concentration relative to β-Actin (Mm_Actb_1_SG QuantiTect Primer Assay) was quantified with a StepOnePlus q-PCR device (Applied Biosystems).