Multiple Integrated Non-clinical Studies Predict the Safety of Lentivirus-Mediated Gene Therapy for β-Thalassemia

Maria Rosa Lidonnici,[1] Ylenia Paleari,[1] Francesca Tiboni,[1] Giacomo Mandelli,[1] Claudia Rossi,[1] Michela Vezzoli,[1] Annamaria Aprile,[1] Carsten Werner Lederer,[1,2] Alessandro Ambrosi,[3] Franck Chanut,[4] Francesca Sanvito,[5] Andrea Calabria,[1] Valentina Poletti,[6] Fulvio Mavilio,[7] Eugenio Montini,[1] Luigi Naldini,[1,3] Patrizia Cristofori,[1,4] and Giuliana Ferrari[1,3]

[1]San Raffaele Telethon Institute for Gene Therapy (SR-TIGET), IRCCS San Raffaele Scientific Institute, Milan, Italy; [2]Department of Molecular Genetics Thalassaemia, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; [3]Vita Salute San Raffaele University, Milan, Italy; [4]GlaxoSmithKline Ware, Hertfordshire, UK; [5]Department of Pathology, IRCCS San Raffaele Scientific Institute, Milan, Italy; [6]Genethon, Evry, France; [7]Department of Life Sciences, University of Modena and Reggio Emilia, Modeno, Italy

Gene therapy clinical trials require rigorous non-clinical studies in the most relevant models to assess the benefit-to-risk ratio. To support the clinical development of gene therapy for β-thalassemia, we performed *in vitro* and *in vivo* studies for prediction of safety. First we developed newly GLOBE-derived vectors that were tested for their transcriptional activity and potential interference with the expression of surrounding genes. Because these vectors did not show significant advantages, GLOBE lentiviral vector (LV) was elected for further safety characterization. To support the use of hematopoietic stem cells (HSCs) transduced by GLOBE LV for the treatment of β-thalassemia, we conducted toxicology, tumorigenicity, and biodistribution studies in compliance with the OECD Principles of Good Laboratory Practice. We demonstrated a lack of toxicity and tumorigenic potential associated with GLOBE LV-transduced cells. Vector integration site (IS) studies demonstrated that both murine and human transduced HSCs retain self-renewal capacity and generate new blood cell progeny in the absence of clonal dominance. Moreover, IS analysis showed an absence of enrichment in cancer-related genes, and the genes targeted by GLOBE LV in human HSCs are well known sites of integration, as seen in other lentiviral gene therapy trials, and have not been associated with clonal expansion. Taken together, these integrated studies provide safety data supporting the clinical application of GLOBE-mediated gene therapy for β-thalassemia.

INTRODUCTION

The development of retroviral vectors and optimized transduction protocols has permitted gene transfer into hematopoietic stem and progenitor cells (HSPCs), leading to the application and translation of gene therapy strategies for a number of hematopoietic diseases and leukodystrophies (reviewed in Dunbar et al.,[1] Eichler et al.,[2] Aiuti et al.,[3] Biffi et al.,[4] and Ribeil et al.[5]).

β-Thalassemia, an inherited anemia characterized by reduced or absent production of hemoglobin β chains, was one of the first candidate diseases for gene therapy.[6] Current management strategies for transfusion-dependent thalassemia patients include blood transfusion, iron chelation, and splenectomy. To date, allogeneic hematopoietic stem cell transplantation (HSCT) is a cure only available for less than 30% of patients.[7] The benefits and limitations of current therapies for β-thalassemia are well discussed by Cappellini et al.[8] *Ex vivo* engineering of autologous HSPCs and administration of genetically modified cells potentially represents a cure applicable to all patients regardless of donor availability and free from transplant-related immunological complications such as graft rejection and graft versus host disease. The evidence gained from allogeneic HSCT indicates that 20%–30% of donor chimerism can be curative in β-thalassemia patients and provides the rationale for a gene therapy approach.[9] A recent study of 4 patients treated with HSCT demonstrated that, despite low myeloid chimerism, the majority of circulating erythrocytes and progenitors were of donor origin, suggesting that they have a selective advantage in *vivo*.[10] These data indicate that the proportion of genetically modified nucleated progenitor cells necessary to achieve a therapeutic level of circulating red blood cells (RBCs) is within the frequency of gene transfer by β-globin lentiviral vector (LV).[10] Furthermore, non-clinical studies conducted in human HSPCs *in vitro* and murine models *in vivo*, as well as the first clinical trial, have demonstrated the feasibility of the gene therapy approach for the treatment of β-thalassemia.[11,12]

Despite the therapeutic success in clinical trials, lymphoproliferative disorders and pre-malignant expansion of myeloid progenitors

have been reported in some patients treated with HSPCs transduced with Moloney murine leukemia virus (MoMLV)-derived vectors, where the retroviral vector was inserted into and activated a proto-oncogene.[13–15] Proviral elements that are involved in the transcriptional regulation of the provirus can alter gene transcription; in particular, elements present in the long terminal repeat (LTR) and the splice donor sites normally used for splicing of viral transcripts. Certain genotoxic events could lead to activation of the transforming potential of a cellular proto-oncogene after integration of a retroviral vector within or near the proto-oncogene, as reviewed by Naldini.[16] Non-clinical studies showed that LVs are less likely to cause insertional gene activation than retroviral vectors,[17] although they can still cause interference with normal gene regulation by inducing alterations in transcriptional activity,[18–21] aberrant splicing,[22,23] and premature transcript termination, as observed in the first clinical trial of gene therapy for β-thalassemia.[11] In this study, a clone harboring an insertion site located within the proto-oncogene HMGA2 accounted for 30% of all integration sites (ISs) in whole blood cells 3 years after gene therapy. In the erythroid cells, the integration event was associated with the expression of an HMGA2 transcript whose 3′ UTR was replaced by vector sequences. This chimeric transcript lacked the target site of the repressive miRNA let-7, contributing to elevated HMGA2 expression together with an increase in the rate of transcriptional initiation. Although the treated patient still remains healthy 11 years after therapy, HMGA2 has been found to be de-regulated in some cancers.

Indeed, understanding the interactions between viral vectors and the genome is instructive to assess the characteristics of transduced cells, although the contribution of different clones to hematopoietic reconstitution is not predictable from non-clinical studies. In gene therapy for β-thalassemia, genotoxicity by *trans*-activation is expected to be very limited because of the erythroid-restricted activity of promoter and/ or enhancer elements included in the self-inactivating vector and because of the nature of the transferred gene (human β-globin chain). No report of general toxicity or development of tumors associated with the transfer of the human β-globin gene by LV has been reported so far in a large number of non-clinical studies in mouse models.[24–29] Therefore, in view of future clinical application, safety studies of genetically modified cells are preferably conducted following guidelines issued by the regulatory agencies such as the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA). Crucial parameters to be assessed, when appropriate and feasible, include the toxicity of the vector, the profile of vector integration, the *in vivo* biodistribution of transduced cells, and germline transmission. Although *in vitro* tests for transduced HSPCs are defined, surrogate *in vivo* assays may rely on transplantation of transduced cells in murine strains and/or humanized immunodeficient mouse models.[30–35]

We have previously reported the development of GLOBE LV and demonstrated expression of therapeutic levels of β-globin and long-term correction with *in vivo* selection of genetically corrected erythroid cells in a severe β-thalassemia intermedia mouse model[26] as well as restoration of normal erythroid differentiation from transduced CD34+ cells of β-thalassemia patients.[36] In both mouse and human cells, we obtained proof of efficacy with the achievement of therapeutic levels of β-globin expression with a low vector copy number and in the presence of a limited proportion of transduced HSPCs.

To move forward with clinical development, we identified the best-performing β-globin LV by comparing GLOBE and newly derived vectors for their transcriptional activity and potential interference with the expression of surrounding genes. Furthermore, to support the use of HSCs transduced by GLOBE LV for the treatment of β-thalassemia, we followed the Guideline on the Non-clinical Studies Required before First Clinical Use of Gene Therapy Medicinal Products (EMEA/ CHMP/GTWP/125459/2006), which defines scientific principles and provides guidance to applicants developing gene therapy medicinal products to facilitate a harmonized approach in the European Union (EU). To guarantee quality, robustness, and traceability, we designed and performed toxicology, tumorigenicity, and biodistribution studies following the Guidelines for GLP (Good Laboratory Practice) in compliance with the Italian GLP Regulations (DL 50, March 2, 2007; G.U. 8, April 13, 2007) and the Organisation for Economic Co-operation and Development (OECD) Principles of GLP (as revised in 1997, ENV/MC/CHEM(98)17). We evaluated the toxic and oncogenic potential associated with the administration of HSPCs transduced with a high dose of GLOBE LV to C57BL6/Hbb^th3 mutant (*th3/+*) mice, which represents the homologous disease model of severe thalassemia intermedia. Moreover, to assess the effect of vector transduction on the ability of human CD34+ cells to engraft and differentiate correctly in the hematopoietic districts, we transplanted GLOBE LV-transduced CD34+ cells into an immunodeficient mouse strain (non-obese diabetic [NOD]/ NOD.Cg-Prkdcsid Il2rgtm1Wjl/SzJ [NSG]) and followed their distribution and differentiation in hematopoietic and non-hematopoietic organs. Both studies were supplemented by vector IS analysis to corroborate safety predictions and HSC-related information.

Overall, our integrated approach, including *in vitro* and *in vivo* experiments, provides results predictive of safety for β-thalassemia gene therapy, complying with the applicable regulatory requirements for advanced medicinal products.

RESULTS

Analysis of Perturbation of Expression in Genes Near or Containing GLOBE LV

Changes in the expression of host genes flanking lentiviral ISs are a safety concern when associated with changes in cell biology. To investigate the transcriptional perturbation of genes targeted by GLOBE LV integrations, we utilized human erythroleukemia (HEL) cells, which are permissive but negative for β-globin endogenous expression. HEL cells were transduced with three different LVs, specifically GLOBE, GATA-GLOBE, and cHS4-GLOBE, a derived vector containing the chicken HS4 insulator element in the 3′ LTR, and grown as single clones. Specifically, GATA-GLOBE is a derived vector containing, in the 3′ LTR, the erythroid enhancer GATA1 HS2 element, potentially acting as an insulator, as we reported previously.[37] The strategy of including insulator elements in the vector LTR aimed to provide protection from the effect of surrounding chromatin to
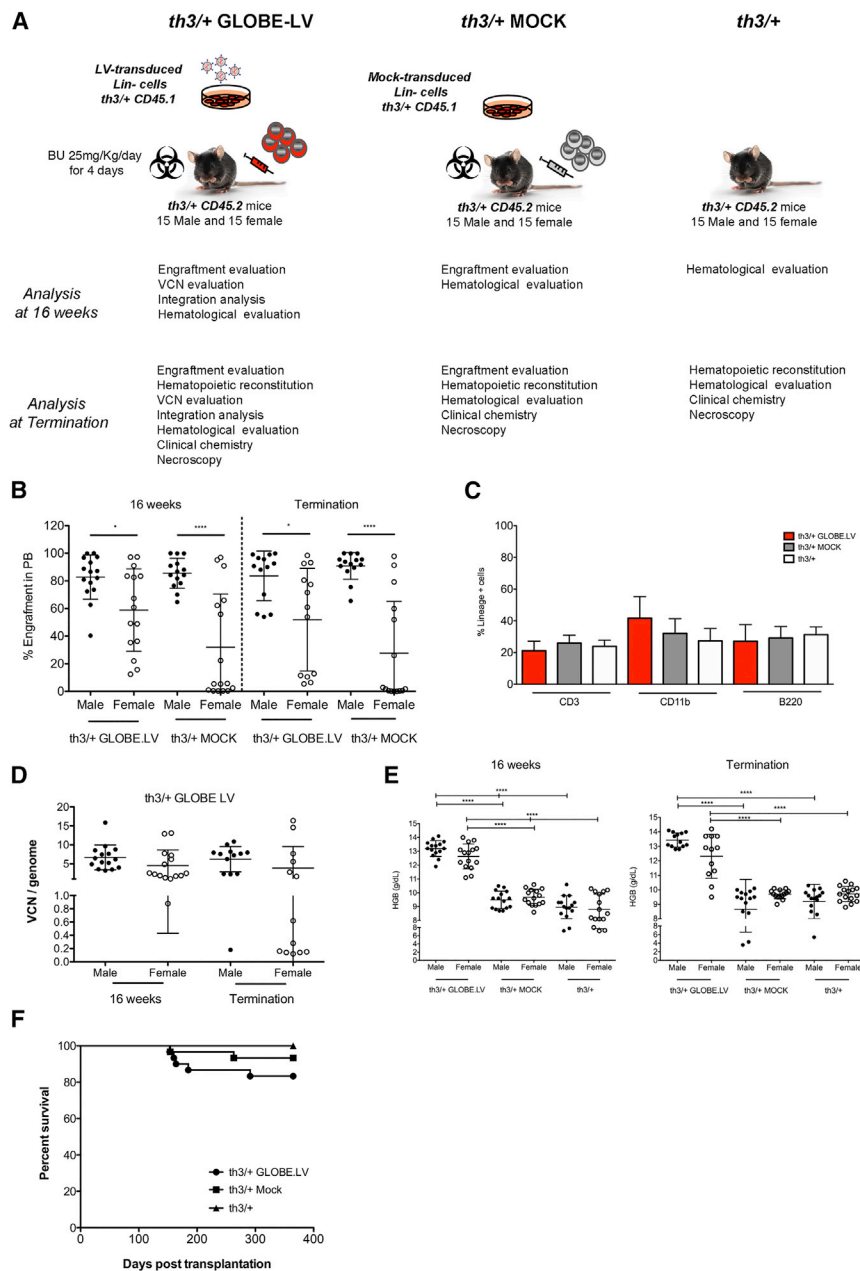
Figure 1. Toxicology and Tumorigenicity Study in a Thalassemia Murine Model

(A) Experimental plan of the toxicology and tumorigenicity study. Lin⁻ cells from donor C57BL6/Hbb^th3 mutant heterozygous (*th3/+*) mice were MOCK-transduced or transduced with GLOBE LV at MOI 100 and incubated overnight. The recipient mice were pre-treated with busulfan (25 mg/kg/day for 4 days). Mice were intravenously injected with a dose of 4×10^5 Lin⁻ cells/mouse. Test group consisted of mice transplanted with GLOBE LV-transduced Lin⁻ cells (*th3/+* GLOBE LV). Control groups consisted of mice pre-treated with busulfan and transplanted with MOCK-transduced Lin⁻ cells (*th3/+* MOCK) and age-matched untreated mice (*th3/+*). Chimerism and cell origin were monitored using antibodies specific for CD45.1 or CD45.2 surface antigen. (B) Donor cell engraftment in male and female recipients was determined by cytofluorimetric analysis for the expression of CD45.1 in peripheral blood mononuclear cells (PBMCs) at 16 weeks and 12 months after transplantation (termination). Values are shown as mean ± SD. Student's t test was applied (*p < 0.05, ****p < 0.0001). (C) Analysis of hematological reconstitution on peripheral blood was assessed by cytofluorimetric analysis. Percentages of myeloid and lymphoid (B and T) cells were evaluated by using antibodies against CD11b, B220, and CD3, respectively. Values are shown as mean ± SD. The GLOBE LV th3/+ group is shown in red, the MOCK th3/+ group in gray, and the th3/+ group in white. (D) VCN per genome was evaluated in GLOBE LV th3/+ group males and females. Values are shown as mean ± SD. (E) HGB concentration was evaluated in all groups of mice. Values are shown as mean ± SD. One-way ANOVA and *post hoc* Bonferroni's multiple comparisons were applied (****p < 0.0001). (F) All animals were observed for signs of ill health or overt toxicity from day1 until the end of the study. Mortality was observed in the GLOBE LV th3/+ and MOCK th3/+ groups from day 153 onward. 5 deaths occurred in the GLOBE LV th3/+ group (circles). In the MOCK th3/+ group (squares), two animals died. No mortality was observed in the th3/+ group (triangles). Survival fractions were calculated using the product limit (Kaplan-Meier) method. Survival curves were compared using the log rank test (also called the Mantel-Cox test). No statistical difference was observed between the groups.

integrated proviruses to obtain reduction of variability and improvement in the β-globin expression level in genetically modified cells. Southern blot analysis of DNA extracted from clones showed that the addition of GATA1-HS2 and cHS4 elements did not affect the stability of the GLOBE LV (data not shown). The vector copy number (VCN) per cell from single clones was measured, and no difference was detected among the three vectors (Figure S1A), indicating a similar transduction ability.

The expression of the β-globin transgene was analyzed by qPCR and western blot in HEL cells induced to erythroid differentiation by hemin exposure and normalized for the VCN (Figures 1B and 1C). No difference in transcriptional level was observed between GLOBE and cHS4-GLOBE (Figure S1B), whereas GATA-GLOBE LV produced 4-fold more β-globin transcript than GLOBE LV (p < 0.001). However, the transcriptional difference between GLOBE and GATA-GLOBE LV was not confirmed at the protein level (Figure S1C).

To identify vector ISs, HEL cell clones were subjected to ligation-mediated (LM)-PCR and DNA sequencing. Change in expression level of target genes because of transcriptional activity at viral ISs

was analyzed by low-density array (LDA) real-time PCR technology (Applied Biosystems). Genes sited within a window of 200 kb around the IS were included in this analysis. The transcriptional activity of genes was assessed in HEL clones before (uninduced) and after induction to erythroid differentiation (induced) (Table S1). HEL cells are capable of both spontaneous and induced globin expression and, thus, are often used as an erythroid model system.[38] β-Globin chains are absent, and this pattern of globin synthesis is observed after 4–5 days of induction by hemin. To determine individual deregulated genes in association with viral ISs, the Z score test with an adj. p < 0.001 was employed. There were no differences in the frequency of the deregulated target genes among all LVs observed. In uninduced clones, there were 4, 6, and 1 deregulated genes associated with viral integration of GLOBE, GATA-GLOBE, and cHS4 LVs, respectively (Figure S1D; Table S1). In induced clones, there were 6, 1, and 6 deregulated genes associated with viral integration of GLOBE, GATA-GLOBE, and cHS4 LVs, respectively (Figure S1E; Table S1). Considering that there are no significant advantages of the newly derived vectors and given the long development and deep characterization of the GLOBE LV,[22] we elected to continue the safety characterization with GLOBE LV.

GLOBE LV-Transduced Cells Correct Thalassemia in the Murine Model

To evaluate the effect of GLOBE LV transduction on hematopoietic cells *in vivo*, we designed and performed a bone marrow (BM) transplantation study in a mouse model of β-thalassemia (th3/+ strain, in compliance with OECD principles and GLP (Figure 1A). The GLOBE LV batch used in this study was produced at medium scale and purified by MolMed with a titer of 9.69×10^7 transduction units (TU)/mL. Lin$^-$ cells from th3/+ mice were MOCK-transduced (th3/+ MOCK) or transduced with GLOBE LV (th3/+ GLOBE LV) at an MOI of 100 and transplanted into female and male recipient th3/+ mice (n = 30 per group), which were monitored for up to 12 months after gene therapy. An age-matched untreated control group was included in the study (th3/+, n = 30).

The transduction protocol was designed to achieve a high VCN superior to that expected in human cells (VCN $\geq$ 2). The transduction efficiency was 100%, and the mean VCN in liquid culture and colony-forming units (CFUs) ranged from 6.03 to 9.69 (number of experiments = 3) (Table S2A).

Recipient mice were pre-treated with a myeloablative dose of busulfan, an alkylating agent causing myeloablation to favor the engraftment of donor cells in recipient HSCs, and transplanted with 0.4×10^5 th3/+ Lin$^-$ cells/mouse, equivalent to approximately 25×10^6 human CD34$^+$ cells/kg.

Donor and recipient hematopoietic cells were derived from two coisogenic strains and could be distinguished by alternative expression of the allelic form of the pan-leukocyte surface marker CD45, being CD45.1$^+$ and CD45.2$^+$, respectively, and detected by cytofluorimetric analysis. Engraftment of th3/+ GLOBE LV− cells and th3/+ MOCK

cells was evaluated 16 weeks after transplantation and at termination in peripheral blood (PB). Overall, engraftment percentages (mean ± SD) did not significantly change in th3/+ GLOBE LV mice (70.78 ± 26.47 at 16 weeks) versus th3/+ MOCK mice (56.93 ± 39.56 at 16 weeks) (Tables S2B and S2C) and were maintained over time (67.78 ± 32.85 to 58.14 ± 42.20, respectively) (Figures 1B and 1C; Tables S2B and S2C). The difference in engraftment percentages between female and male, observed in the th3/+ GLOBE LV group (82.71 ± 16.01 male versus 58.87 ± 29.83 female at 16 weeks; 83.64 ± 17.92 male versus 51.93 ± 37.16 female at termination) was not vector-related because it was also present in th3/+ MOCK mice (Figure 1B). The immunophenotype analysis of blood nucleated cells for the percentage of cells expressing B, T lymphoid, and myeloid-specific surface markers (B220, CD3, and CD11b, respectively) showed a normal pattern of lineage development and no evidence of abnormal expansion (Figure 1C). High VCN (mean ± SD) was detected in PB at 16 weeks (6.686 ± 0.8421 male; 4.535 ± 1.060 female) and in BM cells at termination (6.224 ± 0.9259 male; 3.868 ± 1.576 female), demonstrating sustained long-term engraftment of gene-modified cells (Figure 1C; Table S2B). No statistical difference in VCN was observed between the genders. Overall, these data indicate that transduced cells were able to engraft at a high level with no alteration in the hematopoietic reconstitution.

Transplantation of GLOBE-transduced HSPCs in the mouse model of β-thalassemia showed treatment-related efficacy in terms of rescue of anemia, with increased hemoglobin levels, RBCs, hematocrit (HCT), and amelioration of RBC-related parameters (mean corpuscular volume [MCV], mean corpuscular hemoglobin [MCH], and MCH concentration [MCHC]) (Tables S3A–S3C). A low RBC mass level (RBC count, hemoglobin [HGB], and HCT), consistent with clinical anemia present in th3/+ mice, was observed in control groups (th3/+ MOCK and th3/+ mice) (Tables S3B and S3C; Figure 1E). In contrast, th3/+ mice, receiving GLOBE LV gene therapy, showed an increase in RBC mass levels toward normal circulating levels at both sampling times, indicating a treatment-related response (Table S3A; data not shown). The mean RBC count was 6.4–6.8×10^6/mm^3 in control groups and 8.5–9.3×10^6/mm^3 in th3/+ GLOBE LV mice and a mean HGB from 8.7–9.7 g/dL to 12.3–13.3 g/dL and mean HCT from 25.0%–28.7% to 35.8%–40.5% (Table S3A; Figure 1E). A correlation between percentages of transduced engrafted cells and transgene expression, measured by hematological analysis (HGB), was observed in most of the cases (Figures 1A, 1D, and 1E). Because VCN analysis is performed on nucleated blood and BM cells, and CD45$^+$ expression is absent in the erythroid lineage, in a few cases, a therapeutic level of transgene expression in the presence of a low frequency of transduced cells and/or low VCN was detected as a consequence of positive selection of corrected erythroid cells over the untransduced ones. As in a previous study,[26] correction of hematological parameters was achieved even in mice with a frequency of BM-transduced cells of less than 15%, presumably because of positive selection of erythroid progenitors carrying productive LV integrations. The differences in all hematological parameters in gene therapy-treated mice versus those in control groups were statistically significant (GLOBE LV

Table 1. Incidence of Tumors Observed in Both Decedents and/or Pre-terminally Sacrificed Animals and at Termination

	Males			Females		
Group (n = 15/sex)	th3/+ GLOBE LV	th3/+ MOCK	th3/+	th3/+ GLOBE LV	th3/+ MOCK	th3/+
Lymphoma (various organs)	1 (ICD)	0	3 (TK)	2 (ICD +TK)	2 (ICD+ TK)	2 (TK)
Histiocytic sarcoma (various organs)	1	0	0	0	0	0
Sex cord stromal tumor in the ovary	0	0	0	1 (TK)	0	0
Bronchiolo-alveolar adenoma in the lungs	0	0	0	0	0	1 (TK)
Sarcoma (not otherwise specified) abdominal	0	1 (ICD)	0	1 (ICD)	0	0

ICD, intercurrent death (days after transplantation); TK, terminal kill sacrificed at 50 ± 1 week.

versus th3/+ MOCK and GLOBE LV versus th3/+, p < 0.0001), indicating correction of disease-related abnormalities provided by the gene therapy treatment. Increased hemoglobin levels compared with control groups were observed in all mice at both time points (16 weeks and 12 months after treatment), with the exception of animals 952 and 620 at termination, which also showed a low percentage of transduced cells (Figure 1E; Table S3A). Interestingly, in th3/+ MOCK and untreated mice, extramedullary hematopoiesis (EMH) is present in the spleen and liver, with occurrence of related pigmented macrophages or Kupffer cells. In comparison with both aforementioned groups, histological analysis of mice transplanted with GLOBE LV-transduced Lin⁻ HSPCs showed a reduction of these cells (Figure S2).

Evaluation of Toxicity and Tumorigenic Potential of GLOBE LV-Transduced Murine HSCs

To assess the potential toxicity and/or tumorigenicity of the transplantation of GLOBE LV-transduced cells, mortality and clinical signs were monitored during the 12-month observation period; clinical chemistry, hematologic parameters, necropsy, and full microscopic examination were performed at termination. A board-certified pathologist performed the microscopic examination of a full list of organs of all animals. A peer review of the histopathological analysis was carried out by a second board-certified pathologist.

Differences in survival among the groups were not statistically significant (Figure 1F). No mortality related to the transplantation procedure of Lin⁻ HSPCs transduced with GLOBE LV or MOCK-transduced were observed in the first 4 weeks after treatment. Mortality was observed in the th3/+ GLOBE LV and th3/+ MOCK groups from day 154 onward. There were 7 premature decedents, 5 from the th3/+ GLOBE LV group and 2 from the th3/+ MOCK group (Figure 1F). Deaths were associated with generalized lymphoma in 2 animals in the th3/+ GLOBE LV group and in 1 animal in the th3/+ MOCK group (Tables 1 and 2). Only in one instance was the cause of death unknown.

Abdominal sarcomas in 1 animal in the th3/+ GLOBE LV group and 1 animal in th3/+ MOCK group as well (Table 1) were considered to be related to the intraperitoneal injection of busulfan in an acetone and peanut oil solution and secondary to the local inflammation

induced by the irritation.[39,40] In another case, abnormal dental growth was the likely cause of decreased food intake, resulting in premature death in the th3/+ GLOBE LV group. No mortality was observed in the untreated th3/+ control group (Figure 1F).

Some tumors were recorded throughout all groups with no difference in tumor incidence or type between the groups, and they were all considered within the normal background incidence in th3/+ mice of this age (Table 1). No signs of toxicity and tumorigenicity related to transplantation of GLOBE LV transduced cells were observed.

Clinical chemistry parameters were evaluated at termination in all groups to detect potential toxic effects in the liver, kidney, pancreas, gut, and cardio-circulatory system. No signs of toxicity were found, and no difference was detected between the three groups in any parameter (Figure S3).

Hematopoietic Reconstitution in Gene Therapy-Treated Mice Is Highly Polyclonal and Dynamic

To assess the potential genotoxicity of the GLOBE LV and to analyze the clonal dynamics of hematopoietic repopulation in the transplanted mice, we performed a molecular analysis by retrieving the vector genomic integrations in PB- and BM-derived cells 16 weeks and 1 year after transplantation, respectively. We also analyzed the *in vitro* products, both liquid culture and CFUs, 10 days after transduction.

Vector-genome junctions were retrieved by means of linear amplification-mediated PCR (LAM-PCR) by using three different restriction enzymes, as described previously.[3,4] To identify the vector ISs, the sequences of the PCR products generated by MiSeq Illumina sequencing were mapped on the mouse genome by a dedicated bioinformatics pipeline.[41] Overall, we mapped 1,673 ISs from *in vitro* cultures, 3,080 from PB (at 16 weeks), and 1,738 BM cells (at 12 weeks). Tables S2A and S2B summarize the number of ISs identified for each sample. For the next set of biological analyses, we applied a threshold for the minimum number of reads for each IS (n = 3) with the goal to minimize the number of false positives. To study the clonal dynamics of the transduced cells during the hematopoietic reconstitution, we evaluated the sharing degree of ISs between the PB at 16 weeks and BM 1 year after transplantation

Table 2. Distribution of Hemolymphoproliferative Neoplasia in Both Decedents and/or Pre-terminally Sacrificed Animals and at Termination

Group (n = 15)	Animal Number and Sex	ICD or TK (days)	Lymphomas	VCN PB (BM)	VCN Tissue	% CD45.1 /(CD45.1+Cd45.2) in PB[a]
th3/+ GLOBE LV	617 F	ICD (154)	lymphoma systemic	12.92 (16.53)	8.69	67.3
	690 M	ICD (181)	lymphoma systemic	15.87 (NA)	7.82	80.9
	627 F	TK	lymphoma systemic	6.2 (0.14)	NA	12.8
	605 M	TK	histiocytic sarcoma systemic	7.44 (10.87)	NA	88.1
th3/+ MOCK	654 F	ICD (260)	lymphoma systemic	–	–	62.2
	647 F	TK	lymphoma lung	–	–	0.2
th3/+	661 M	TK	lymphoma mesenteric lymph node	–	–	–
	664 M	TK	lymphoma Peyer's patch jejunum	–	–	–
	665 M	TK	lymphoma mesenteric lymph node	–	–	–
	677 F	TK	lymphoma Peyer's patch ileum	–	–	–
	689 F	TK	lymphoma liver	–	–	–

ICD, intercurrent death (days after transplantation); TK, terminal kill sacrificed at 50 ± 1 week; NA, not available; F, female; M, male; –, not applicable.
[a]Percent chimerism in PB at the latest time point.

and found that 30% of PB ISs were shared with the ISs in BM, indicating self-renewal and multilineage potential of the transduced engrafted HSPCs (Figure 2A).

Moreover, to quantify clonal diversity over time, we calculated the Shannon diversity index. This index measures the entropy of an IS dataset, taking into account the total number of ISs and their relative contribution. The diversity of ISs in BM 12 months after transplantation was lower compared with that of PB at 4 months in almost all sample analyzed, suggesting that a diverse clonal repertoire is established after transplantation with different active gene-corrected long-term (LT)-HSC contributing to hematopoiesis (Figure 2B).

To better investigate this aspect, we measured the proportion of sequencing reads representing each IS within our datasets as a surrogate readout for the relative abundance of a cell clone harboring that integration at a given time. In PB at 16 weeks, this analysis identified 1–2 clones in 70% of the analyzed mice with a relative abundance above 10%, a threshold arbitrarily set for defining dominant clones (Figure S4A). Only one mouse (952) harbored 2 integrations above 30% (Figure S4A). Thus, PB at 16 weeks showed a highly polyclonal reconstitution pattern.

At termination, in the BM, we found a less polyclonal pattern, with all mice harboring at least one clone with relative abundance above 10% (Figure S4B). Mouse 627 was of particular interest because we retrieved only 2 ISs, and one of them (targeting the Wdr78 gene) has a relative abundance of 99% in the BM, whereas it was at a very low level in PB at 16 weeks (Figure S4). The engraftment and VCN in this mouse were very low at both time points (Table S2B), indicating that few transduced clones are participating to host recovered hematopoiesis at low levels (Figures 2C and 2D). Furthermore, almost every cell clone marked by a specific IS accounted for only a fraction

of a percentage of the total hematopoietic clones at any given time (Figures 2C and 2D).

We also analyzed the genomic distribution of ISs and the gene classes preferentially hit by LV integration. The genomic distribution of ISs, both in the *in vitro*-cultured Lin⁻ cells and *in vivo*, matched the previously reported LV preference for integration within transcriptional units (on average, 80% of ISs were within genes).[42] In all groups, ≈80% of ISs were found between 5 and 500 kb from transcription start sites (Figures S5A–S5C). From these analyses, it is evident that the hematopoietic reconstitution of the mice transplanted with transduced cells is highly polyclonal and dynamic and that GLOBE LV shows the classical integration pattern characteristic of HIV-1-derived vectors.

To assess to which groups of biological processes genes with integrated vector belong, gene ontology (GO) enrichment analysis was performed. Interestingly, in *in vivo* data, we observed an enrichment in categories belonging to metabolic and biosynthetic processes (Figures 3A and 3B).

Development of Lymphoproliferative Tumors Occurred Independently from GLOBE LV Integration

To better define the biosafety of GLOBE LV, we completed IS analysis mapping in the infiltrated spleens of the two mice dead from lymphoma (617 and 690; Table 2). No differences in genomic distribution were detected with respect to the other IS groups analyzed (*in vitro*, PB, and BM) (Figure S5D), and GO (GO) analysis did not identify enrichment in terms belonging to cancer, abnormal proliferation, or similar, but it identified a significant over-representation of the sexual development categories (Figure 4A). In these mice, VCN analysis in lymphoma-infiltrated spleens showed vector marking at a lower level compared with the other hematopoietic tissues (Table 2), with a high donor cell chimerism in the blood (67% and 81%, respectively).
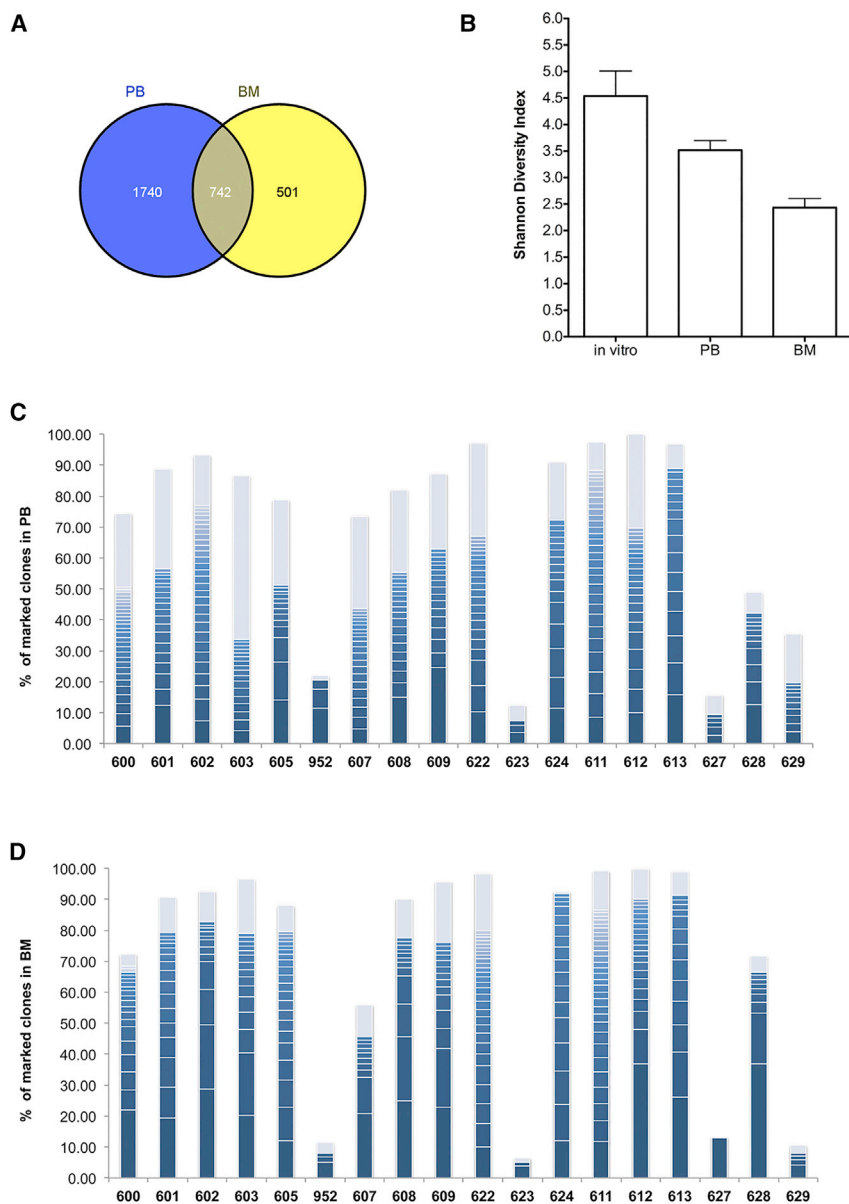
Figure 2. Hematopoietic Reconstitution of Gene Therapy-Treated th3/+ Mice Is Highly Polyclonal and Dynamic

DNA extracted from *in vitro* cultures, peripheral blood at 16 weeks, and bone marrow at termination was analyzed for vector integration. (A) Venn diagram showing the overlapping ISs between PB and BM. (B) Shannon diversity index values (y axis) in *in vitro* samples, PB, and BM. (C) Histograms of the percentage of sequence reads (y axis) for unequivocally mapped ISs from PB. The number of reads for each IS is normalized to the total number of sequence reads from the same time point and for the engraftment level. ISs below 1% were collapsed and are shown in gray. (D) ISs in BM.
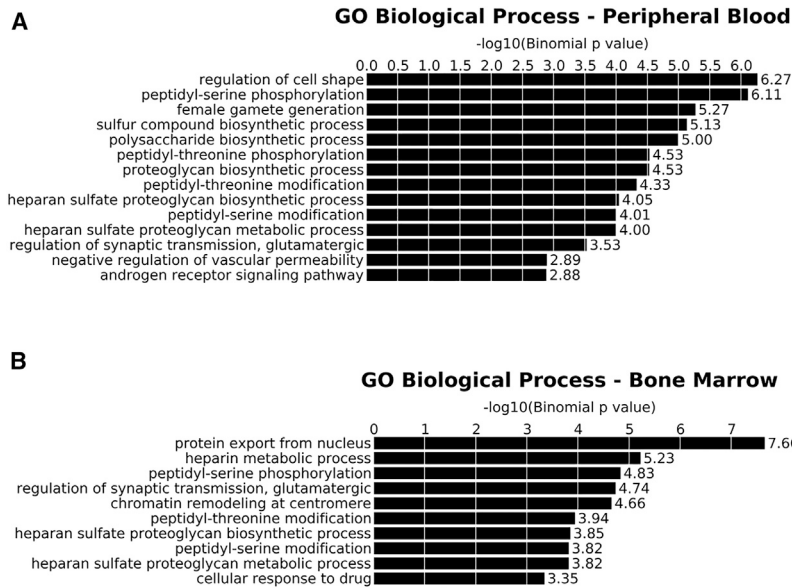
BM at a higher relative abundance level (56.4%), whereas, at an earlier time point, it was at a very low level (2.8%) (Figure S6). Considering the correction for engraftment level, this clone accounted for 32.8%–37.9% of total hematopoiesis. The integration targeting a gene (Klra21) is not cancer-related; thus, it indicates that the tumor is marked by a neutral integration. Similar results were obtained in mouse 690, where we identified two ISs with a relative abundance of 25.5% and 21.7% that, considering the engraftment level, accounted for the ≈20% of total hematopoiesis (Figure 4B). Comparing these results with those in tumor-free mice, the clonal expansion was not relevant for cancer development (Figure 4B). Moreover, mice bearing hemolymphoproliferative tumors recorded at termination (605 and 627) showed dominant clones with lower abundance (≈12% in BM), whereas tumor-free mice had clones with a similar or higher abundance (≈37% in BM) (Figure 4B).

Then, we performed an analysis of cancer-associated genes using the Retroviral Tagged Cancer Gene Database (RTCGD) database. In all four samples (*in vitro*, BM, PB, and SP samples), ≈20% of genes are present in the database (Table S4). From the analysis of tumors, we did not see any enrichment in the abundance of integrations targeting cancer genes, suggesting that the integrations are not involved in the transformation process.

Hemolymphoproliferative neoplasia was identified in two other mice (605 with histiocytic sarcoma and 627 with lymphoma) at termination.

Animal 617 was found dead and showed enlargement and discoloration of the spleen, thymus, mesenteric lymph nodes, and kidney. Samples of the spleen and thymus were frozen, and IHC was performed. A severe infiltration of atypical enlarged cells involving red and white pulp was seen. Infiltrating cells were positive for staining with anti-Ki67, CD3, and CD45.1 antibodies, suggesting a T origin of the lymphoma cells from donor-derived cells (data not shown). In the spleen of mouse 617, we identified one clone with a relative abundance of 48.8% (Figure S6). This IS was also retrieved in the

To determine whether other hallmarks of insertional mutagenesis were present in the GLOBE LV integration profile, we assessed the occurrence of common ISs (CISs), insertional hotspots that may result from integration bias at the time of transduction or *in vivo* selection of clones harboring integrations that confer a growth advantage. CISs were identified with an algorithm based on Abel et al.[43] and by the Grubbs test for outliers.[34] CIS analysis did not show any

A

GO Biological Process - Peripheral Blood

-log10(Binomial p value)

Process	Value
regulation of cell shape	6.27
peptidyl-serine phosphorylation	6.11
female gamete generation	5.27
sulfur compound biosynthetic process	5.13
polysaccharide biosynthetic process	5.00
peptidyl-threonine phosphorylation	4.53
proteoglycan biosynthetic process	4.53
peptidyl-threonine modification	4.33
heparan sulfate proteoglycan biosynthetic process	4.05
peptidyl-serine modification	4.01
heparan sulfate proteoglycan metabolic process	4.00
regulation of synaptic transmission, glutamatergic	3.53
negative regulation of vascular permeability	2.89
androgen receptor signaling pathway	2.88

B

GO Biological Process - Bone Marrow

-log10(Binomial p value)

Process	Value
protein export from nucleus	7.66
heparin metabolic process	5.23
peptidyl-serine phosphorylation	4.83
regulation of synaptic transmission, glutamatergic	4.74
chromatin remodeling at centromere	4.66
peptidyl-threonine modification	3.94
heparan sulfate proteoglycan biosynthetic process	3.85
peptidyl-serine modification	3.82
heparan sulfate proteoglycan metabolic process	3.82
cellular response to drug	3.35

Figure 3. Gene Ontology Analysis of ISs

Gene ontology (GO) analysis was performed using the GREAT software on *in vitro* cultures, peripheral blood at 16 weeks, and bone marrow at termination at 12 weeks. The integrations retrieved from each group showed overrepresentations of the gene functions indicated on the left. Only gene classes with a false discovery rate of less than 0.05 were considered. We interrogated the Biological Process database in (A) peripheral blood and (B) bone marrow.

alarming sign of genotoxicity because the only significant CIS was found in the control dataset of *in vitro*-cultured cells. A less stringent correction of the p values obtained in our CIS analysis also allows finding "weaker" CISs in the integration datasets *in vivo* (PB and BM). Interestingly, the CISs found in BM 1 year after transplant are less and are targeted by fewer integrations than the CIS identified in PB 16 weeks after transplantation. Thus, even a less stringent correction of the CIS significance does not show any abnormal enrichment of CISs over time. Moreover, CISs were not represented by high sequence counts and did not preferentially target oncogenes.

Overall, these analyses showed a safe integration profile of GLOBE LV in murine HSC and that GLOBE LV is not directly linked to tumor development.

Human CD34[+] Cells Transduced with GLOBE LV Are Able to Engraft and Reconstitute NSG Mice

As a further step, we assessed whether the distribution of GLOBE genetically modified human CD34[+] cells in the organs of the immunocompromised NSG mouse model was comparable with that of unmodified cells. We tested GLOBE LV-transduced human CD34[+] cells for their ability to engraft, differentiate into lineage-specific progeny, and localize into lympho-hematopoietic murine tissues.

Cord blood-derived CD34[+] cells (UCB-CD34[+]) were transduced with a GLOBE LV batch manufactured under good manufacturing practice (GMP) conditions. Transduced (CD34 GLOBE LV) or MOCK-transduced (CD34 MOCK) cells were injected intravenously (i.v.) into sub-lethally irradiated NSG mice (CD34 GLOBE LV, n = 12; CD34 MOCK, n = 12). A third group of age-matched mice was included in the study as an untreated control group (CD34 Ctrl, n = 6). The mice were monitored daily for survival and weekly for

clinical signs and body weight until 12 weeks after transplantation. At the end of the study, the engraftment, differentiation to specific lineages, and persistence of transduced human cells in hematopoietic organs (blood, BM, spleen, and thymus) were assessed. After extensive perfusion to reduce contaminating blood, gonads and brain were analyzed by qPCR to evaluate the distribution of human cells also in non-hematopoietic off-target organs (Figure 5A). There were no clinical signs considered to be related to the treatment and no mortality was observed during the study up to termination, except for one mouse in the CD34 MOCK group that died after bleeding at 11 weeks (day 77).

The immunophenotype analysis data showed successful and robust engraftment of transduced and MOCK-transduced UCB-CD34[+] cells in NSG mice after sub-lethal irradiation. Statistical analysis showed a lower percentage of human CD45[+] cells detected in the periphery in CD34 GLOBE LV mice (mean ± SD, 13.96 ± 3.146) compared with CD34 MOCK mice (31.24 ± 4.83) (p = 0.0066); however, no difference was found in the BM (25.23 ± 2.28 versus 33.40 ± 4.22) and in the spleen (17.86 ± 4.34 versus 12.56 ± 7.23) (Figure 5B). As expected, in the NSG model,[44] the higher proportion of human cells was represented by B cells (CD19[+] cells) in the PB and in the hematopoietic organs (Figures 5C–5E). The BM showed human HSPC engraftment (immunostaining with anti-human CD34) with multi-lineage differentiation of lymphoid and myeloid cells (immunostaining with anti-human CD19, CD3, and CD11b). No significant differences in CD34[+] levels were observed between groups in the BM, indicating a comparable capacity of CD34 GLOBE LV cells (3.25 ± 0.50) and CD34 MOCK cells (4.76 ± 0.66) to engraft and reconstitute *in vivo* (Figure 5D).

No differences in VCN were observed among the target organs PB, BM, and SP (4.04 ± 1.59, 5.09 ± 2.97, and 4.38 ± 1.03, respectively) (Figure 5F). The average VCN measured in all examined hematopoietic tissue was comparable with the VCN measured in the bulk population of the transduced cells prior to administration (4.11).

At termination, non-hematopoietic organs (ovaries and brain) were collected from NSG mice after perfusion to assess the risk of germline transmission through vector shedding and secondary transduction or inadvertently generated replication-competent lentivirus (RCL).
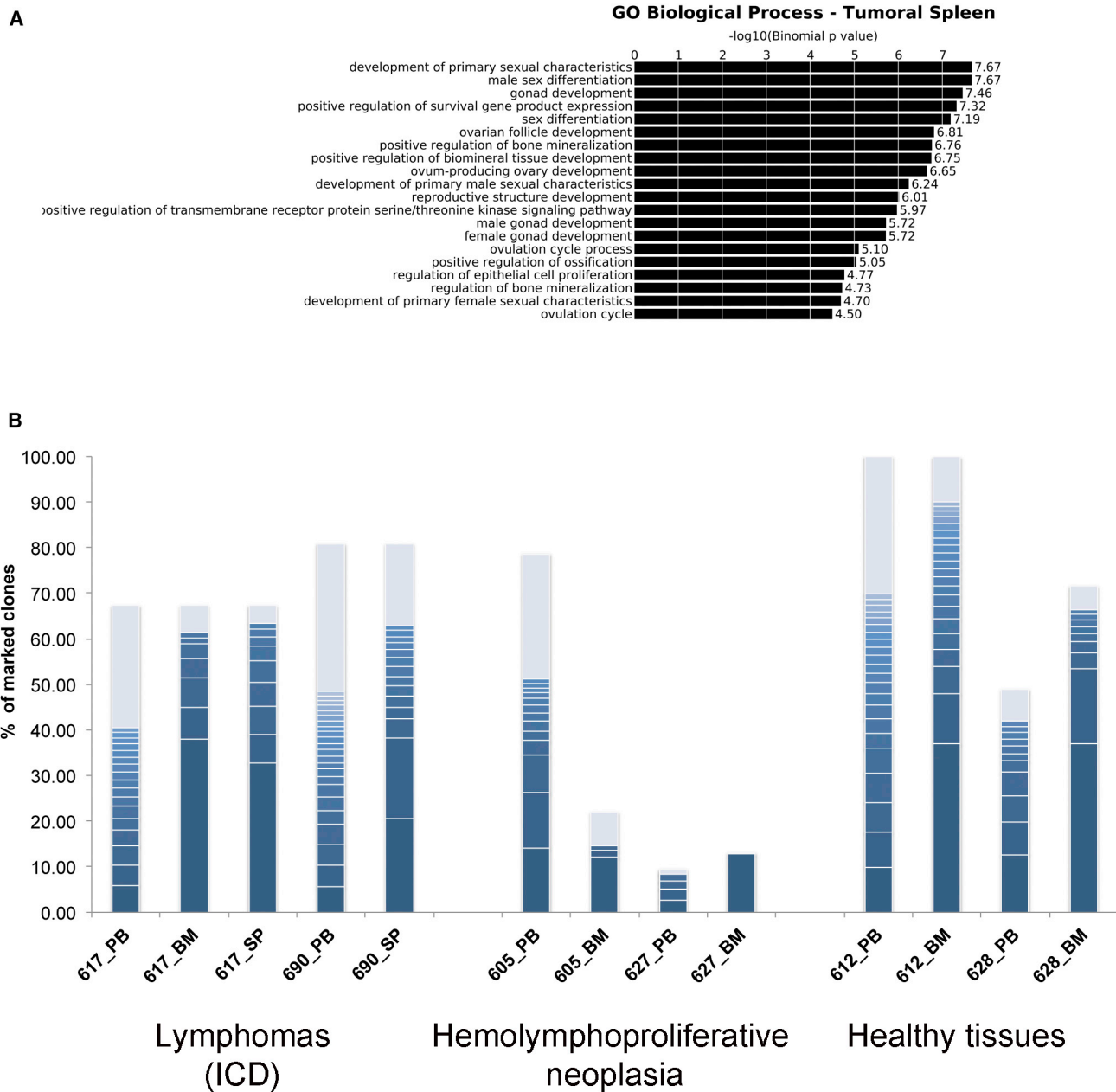
A

GO Biological Process - Tumoral Spleen

B

Lymphomas (ICD)

Hemolymphoproliferative neoplasia

Healthy tissues

Figure 4. Gene Ontology Analysis and Abundance Analysis of ISs from Infiltrated Tumors

DNA extracted from infiltrated tumor spleens was analyzed for vector integration (mice 617 and 690). (A) GO analysis was performed using the GREAT software. The integrations retrieved from tumors showed overrepresentations of the gene functions indicated on the left. Only gene classes with a false discovery rate of less than 0.05 were considered. We interrogated the Biological Process database. (B) Histograms of the percentage of sequence reads (y axis) for unequivocally mapped ISs from mice that developed lymphomas and hemolymphoproliferative neoplasia and healthy controls. The number of reads for each IS was normalized to the total number of sequence reads from the same time point and for the engraftment level. ISs below 1% were collapsed and are shown in gray.

The ovaries of 10 of 12 mice in the CD34 GLOBE LV group were negative or below the LOQ (low limit of quantification) for VCN analysis (Figure 5G). In the ovaries of the other 2 mice, less than 0.5% of human cells were detected. The VCN of 12.26 in the ovary sample from mouse 1-05 was generated by a ratio between low LV copies and an amount of human DNA very close to the LOQ. The brains of 11 of 12 mice were negative or below the LOQ for VCN analysis (Figure 5G). The VCN of 3.60 in mouse 1-11 was below the PB VCN of the same mouse (6.43). These data are likely were likely caused by false positives or related to incomplete perfusion so that

residual blood containing high VCNs can be measured in these samples, as shown in Visigalli et al.[45] Calculation of VCN in mouse tissues, as a ratio between murine genomes (β-actin target ng) and LV genomes (HIV ng), excluded the association of LV with murine cells (data not shown). Occasionally, a low signal for vector DNA was detected in non-target organs of control group mice, setting the background threshold of sample processing and the qPCR procedure.

Serum samples from CD34 GLOBE LV mice (1-01, 1-02, 1-05, 1-08, and 1-09) that showed a low level of LV DNA in ovary and/or brain in the presence of low or undetectable levels of human DNA (hTELO ng) were also analyzed by HIV p24 assay to exclude the presence of circulating aberrantly generated RCL. All analyzed samples were negative in the HIV p24 detection assay; i.e., below the analytical sensitivity of the ELISA test (<3.125 pg/mL).

Overall, these results showed that the GLOBE medicinal product successfully engrafts and differentiates into the myeloid and lymphoid progeny at a level comparable with that observed for MOCK-transduced CD34$^+$ cells. Notably, the proportion of human cells retaining the primitive CD45$^+$/CD34$^+$ phenotype was comparable in mice receiving GLOBE LV-transduced and MOCK-transduced cells. No evidence of LV transfer in non-target organs, gonads and brain, was found.

In Vitro and Ex Vivo ISs Profiles of GLOBE LV in Human CD34$^+$ Cells

The GLOBE LV integration profile was determined by LM-PCR coupled to high-throughput sequencing in vitro in the pool of transduced human CD34$^+$ cells before transplantation and ex vivo in the BM of the 10 transplanted NSG mice at the end of the study. All mice were injected with 0.5×10^6 transduced CD34$^+$ cells, with a resulting chimerism of 20%–35% human cells and a VCN ranging from 1 to 10 (Figure 6A). Overall, 14,023 unique LV ISs were retrieved in the transduced CD34$^+$ cells and 2,518 ISs in the pool of transplanted mice at the end of the study, indicating a highly polyclonal BM repopulation. The LV integration profile was also determined in individual animals, retrieving an average of 354 ISs/mouse (range, 146–709 ISs; SD = 183; Table S5).

Overall, no or few dominant clones (i.e., accounting for 10% or more of the total reads) were evidenced in post-transplant libraries by analyzing the final ISs read count (Figure 6B). In single-mouse

libraries, most of the ISs had a read count of lower than 1% (low frequency, shown in gray in Figure 6B), whereas a variable number of ISs (6–22) had a relatively higher frequency (more than 1%, colored boxes in Figure 6Bb), most likely reflecting the limited number of long-term repopulating hematopoietic stem cells in the transplanted BM. Altogether, ISs with a read count of more than 1% in any individual animal targeted 128 genes, which were not significantly enriched for any biological function by GO Database for Annotation, Visualization and Integrated Discovery (DAVID) analysis (data not shown). Interestingly, none of these ISs was found to be in common among the different animals.

The relative distribution of ISs inside and outside of genes changed in the post-transplant samples with respect to pre-transplant-transduced CD34$^+$, with a significant increase in the fraction of intergenic ISs (χ^2 test p < 0.0001). This effect was particularly evident in two of ten mice analyzed (TM024-1-01 and TM024-1-05; Table S5; Figure S8).

Overall, we retrieved 4,802 and 1,316 genes targeted by LV GLOBE in pre-transplant CD34$^+$ cells and post-transplant BM, respectively. The number of targeted genes was positively correlated with the number of ISs (Figure S7A), with most of the targeted genes hosting 1 provirus (data not shown), ranging from 1–50 ISs/gene in pre-transplant CD34$^+$ cells (50 ISs in PACS1, all low-frequency ISs), 1–12 ISs/gene in merged BM samples (12 ISs in LOC440434, all low-frequency ISs), and 1–9 ISs in individual mice. Among the top 100 genes targeted ex vivo, we found genes preferentially targeted by LVs, such as PACS1, KDM2A, and others, as observed in the pre-transplant CD34$^+$ cells (Figure S7C) and in three LV-based clinical trials.[3,46] The genes targeted in pre-transplant CD34$^+$ cells were significantly enriched in several biological functions (Figure S7D). Targeted genes in post-transplant BM were not significantly enriched in any GO biological function category or molecular pathway (Kyoto Encyclopedia of Genes and Genomes [KEGG] pathways dataset) with respect to the pre-transplant sample by DAVID analysis (data not shown), suggesting no selection for clones carrying LV integration in genes involved in particular functions. Although most of the targeted genes were in common in post- versus pre-transplant cells (Table S6), only 1 or 2 individual ISs were shared between pre- and post-transplant samples (Figure 6C; Table S7). On the contrary, several ISs were shared among individual mice (Figure 6C; Table S7), ranging from 8% to 64% of the total ISs (Figure S8).

Figure 5. Biodistribution Study in the Human Chimera Murine Model

(A) Experimental plan of the biodistribution study. Female NSG (NOD.Cg-Prkdcsid Il2rgtm1Wjl/SzJ) recipient mice were conditioned by a sub-lethal dose of irradiation to provide depletion of endogenous BM cells, allowing engraftment of human cells. After conditioning, NSG GLOBE LV mice (n = 12) were intravenously injected with transduced CD34$^+$ cells, NSG MOCK mice (n = 12) were treated with MOCK-transduced cells, and NSG control mice (n = 6) were left untreated and served as controls. At the end of the study, the engraftment, the differentiation to specific lineages, and the persistence of transduced human cells in hematopoietic organs were evaluated. After extensive perfusion to eliminate contaminating blood, gonads and brain were analyzed by qPCR to evaluate the distribution of human cells in non-hematopoietic organs. HIV p24 was assayed in mouse serum to test for the absence of replication-competent lentivirus. (B) Analysis of hematological reconstitution was assessed by cytofluorimetric analysis in PB, BM, and SP by using anti-human CD45$^+$ (**p < 0.01, two-tailed Student's t test). (C–E) PB (C), BM (D), and SP (E) showed multi-lineage differentiation of lymphoid and myeloid cells (CD19$^+$, CD3$^+$, and CD13$^+$) both in transduced and MOCK-transduced cells. (F) VCN per cell was determined in PB, BM, SP, and thymus by qPCR. (G) VCN per cell was determined in ovaries and brain. All values are shown as mean ± SD.
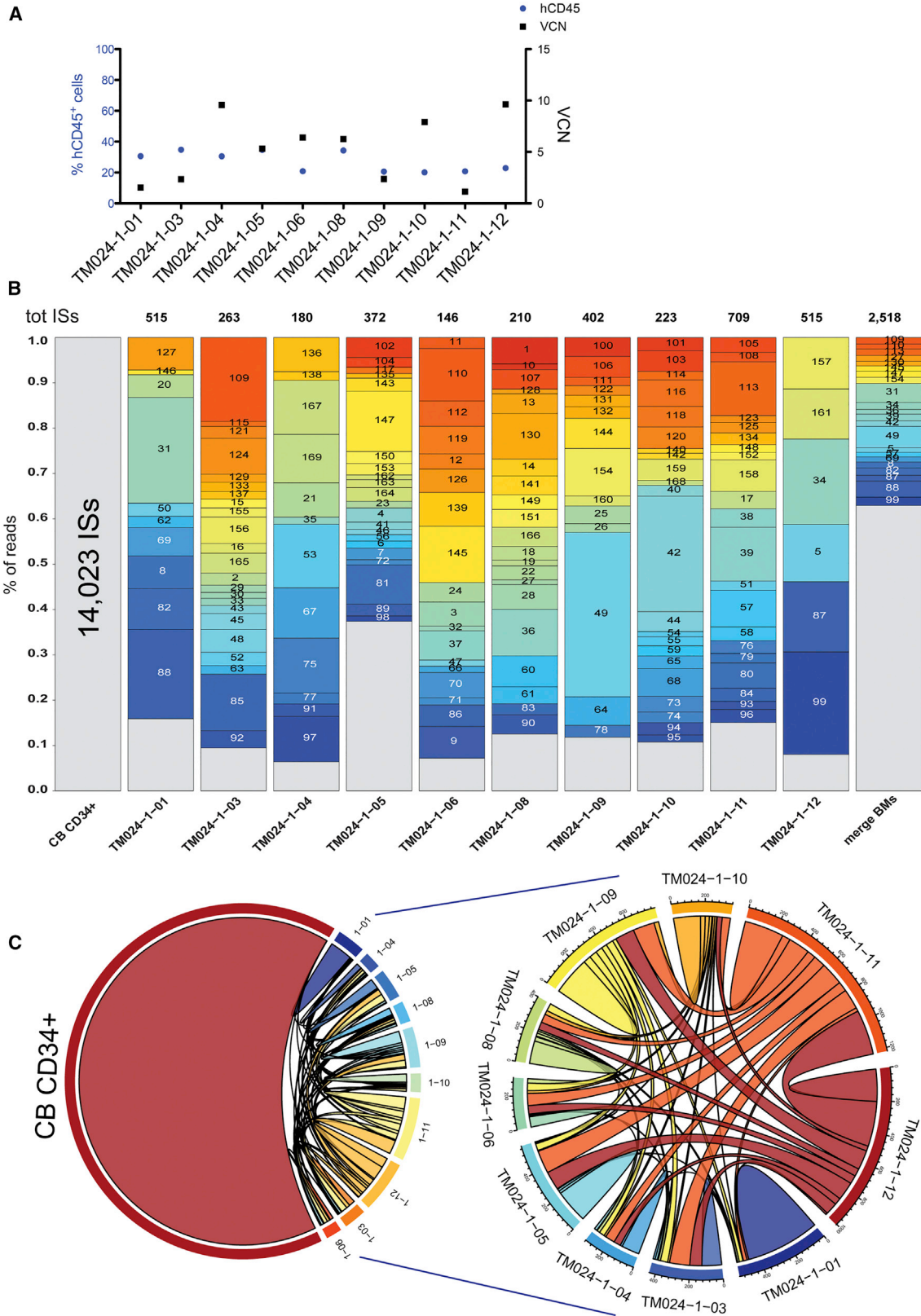
(legend on next page)

DISCUSSION

Here we analyzed transgene expression and production from three different globin LVs, specifically GLOBE, GATA-GLOBE, and cHS4-GLOBE, and their potential deregulation of cellular genes located in the vicinity of or containing the vector insertion in human cells. No significant differences were observed among LVs. On the basis of these experiments and given the long development and deep characterization of the GLOBE LV,[22] we decided to continue the safety characterization with GLOBE LV.

We observed that GATA-GLOBE LV produced more β-globin transcript than GLOBE LV. This higher transcription activity is due the presence of the GATA1-HS2 element in the LV, which binds the transcription factors CP2 and BCL11A.[47,48] Our data confirmed the potential GATA1-HS2 anti-silencing effect.[37] However, this transcriptional difference was not confirmed at the protein level. Moreover, evaluation of the perturbation of genes hit by LV integration highlights that GLOBE LV deregulated gene expression with a frequency similar to the other β-globin-expressing LVs reported in this paper and in the literature.[18–20] Hargrove et al.[20] showed that vector integration perturbed the expression of genes with a frequency of 11%, whereas GLOBE LV showed a frequency of 4% and 7% in immature and more mature erythroid cells, respectively (Table S1).

The safety and efficacy of GLOBE LV were tested *in vivo* by transplantation experiments following the guidelines for GLP in the GLP-SR-Tiget test facility. The preclinical study for gene therapy of mucopolysaccharidosis I (MPS I), recently published by Visigalli et al.,[45] was performed in this facility following the same guidelines. We designed an *in vivo* study aimed to evaluate the safety and efficacy of GLOBE LV in a mouse model of β-thalassemia intermedia (th3/+). Engraftment and transduction were very efficient, and the thalassemic phenotype was fully corrected in all mice, with two exceptions (620 and 952). Mouse 620 had very low engraftment (5.4%). Mouse 952 had an engraftment of 11.5% with a VCN in the marked cells of 1.39 (VCN 0.16 on total BM) and the most abundant clone, representing about 5% of total hematopoiesis. However, 2 mice with similar abundant clones were fully corrected with similar VCN in the marked cells (1.88–1.43). These data suggest that, in the presence of genomic low VCN per cell, the transcriptional activity of the vector might be influenced by surrounding chromatin.[49] Interestingly, GLOBE LV corrects the thalassemia phenotype with 1 copy present in about 10% of total hematopoietic cells, as shown by mouse 627, confirming previously published data.[26]

VCN per-cell values were lower and with a higher variability in the female groups, a reflection of variability in engraftment level. This finding was not vector-related because it also occurred at a higher level in the control group (th3/+ MOCK mice). Variability of engraftment might be expected because of variable bioavailability of busulfan. Pharmacokinetic studies have verified great inter-patient variations because of age, circadian rhythmicity, and underlying diseases.[50–53] As an additional factor to this variability, a gender-related different liver metabolism, influencing drug clearance in the th3/+ strain, which is characterized by hepatic iron imbalance, can be hypothesized. A higher liver iron concentration in th3/+ females[54] and a testosterone-related effect on iron homeostasis might influence metabolic liver function. Because dose adjustment, based on busulfan plasmatic dosage, is not feasible in mice, variable drug effects might have occurred, and a partial level of myeloablation because of biological variability and different sensitivity to busulfan treatment in females compared with males might have affected the level of engraftment of transplanted cells. Histopathological evaluation of tissues, clinical chemistry, and hematology did not reveal any vector-related safety concerns.

In the gene therapy-treated group, we reported 2 premature deaths because of generalized lymphoma and 2 mice with lymphoma and histiocytic sarcoma at terminal sacrifice. To understand whether these tumors could have been induced by insertional mutagenesis, the average VCN per cell was assessed in the spleens of premature decedent animals (617 and 690). These animals showed high donor cell chimerism (67% and 81%) and vector marking (12.92 and 15.87) in the blood but a lower VCN ($\approx$2-fold) in the affected tissues (8.69 and 7.82). Moreover, because these malignancies were also present in the MOCK group (1 premature death and 1 terminal kill) and in the untreated group (5 terminal kills), they are not considered a treatment-related finding. A total of 8,362,036 sequence reads resulted from the appropriate data filtering process of the original reads. The relative abundance of sequences representing each IS (a clonal mark) with respect to the total number of sequences provided an estimation of the relative abundance of a cell clone in the transduced population. The highest estimated values of the abundance of individual cell clones at early (16 weeks) or late (at termination) time points after transplant was lower than 4% of the entire dataset, which is below the threshold value generally associated with clonal dominance.[3,4] To complete the safety profile of GLOBE LV treatment, IS analyses were performed on neoplastic tissue, PB, BM, and pre-transplant samples. As expected, transplantation of a high amount number of lineage-negative mouse BM cells resulted in a polyclonal hematopoietic repopulation with preferential insertions (80% of ISs) in gene coding regions. These ISs do not show evidence of enrichment in growth- or cancer-related genes, and no CISs were retrieved in *in vitro* and *in vivo* samples.

Figure 6. *In Vitro* and *Ex Vivo* Integration Site Analysis of LV GLOBE in Human HSPCs
(A) Human versus mouse chimerism and VCN evaluation in the BM of transplanted NSG mice at the end of the study. (B) Read count distribution per single IS obtained in pre-transplant transduced CB CD34+ cells and in the different BMs of transplanted mice at the end of the study. The gray color indicates a read count lower than 1% of the total trimmed reads, whereas the other colors indicate a read count higher than 1%. ID numbers for each IS are indicated in the boxes. In the study, no ISs with a read count greater than 1% were shared among the individual transplanted mice. (C) Shown on the left and right, respectively, are chord diagrams indicating the shared ISs among the different samples, including or not including the pre-transplant CB CD34+ cells.

Moreover, GLOBE LV does not show preferred integration in either *MDS1-EVI1* or *LMO2*, which were previously linked to malignant transformation in the gene therapy setting. In fact, integration of gamma retroviral vectors into these loci has been associated with tumor development in patients treated for chronic granulomatous disease[55] or X chromosome-linked severe combined immunodeficiency and Wiskott Aldrich syndrome.[56–58] Together, these data indicate prediction of a safe integration profile of GLOBE LV.

In addition, the tumors we observed are not likely to be driven by integrations in dangerous genes, and our data are supported by a number of other studies describing spontaneous lymphomas in aging mice and undergoing conditioning.[29,32,56,59–62] In fact, lymphomas have been reported to be the most commonly observed spontaneous tumors[62] in C57BL6 mice,[61] the background of our disease model. Moreover, chemicals, retroviruses, or irradiation induces lymphoma and leukemia.[32,63,64] These studies suggest that there are other risk factors for oncogenesis in hematopoietic gene therapy protocols besides integration of LVs in the genome. Moreover, our data are in agreement with the findings of Visigalli et al.[45] In conclusion, overall phenotype correction of thalassemic mice was observed with GLOBE LV, with no alteration of BM homeostasis in the transplanted animals.

Because the experimental medicinal product is represented by human autologous transduced CD34$^+$ cells, and there is limited clinical experience with the *ex vivo* use of LV, we designed assays to assess the biodistribution and the potential risk of vector shedding, mobilization, and germline transmission using the relevant human target cells in *in vivo* models. In particular, we assessed whether the distribution of the GLOBE medicinal product (GLOBE LV genetically modified human CD34$^+$ cells) in the organs of the humanized mouse model was comparable with that of unmodified cells. Our results indicate that the clinical-grade GLOBE LV can transduce, with high efficiency, human CD34$^+$ progenitors. The two-hits protocol allowed us to reach 4 VCN/cell with no detectable toxic effect on *in vivo* differentiation of CD34$^+$ cells. It should be noted that the selected protocol is already adopted in LV-based trials for genetic diseases where no side effects have been reported so far.[3,4]

Our results show that GLOBE LV-transduced HSCs migrate to the hemolymphopoietic organs and undergo *in vivo* short-term differentiation, similar to unmodified HSCs. Our data, combined with previous observations,[45] indicate that the use of the immunodeficient strain NSG, which allows efficient engraftment of human cells, represents a suitable model for assessing the biodistribution of human CD34$^+$ cells.[33] Our data show that LV distributes along with human hematopoietic cells, remaining integrated in the human genome, and is not mobilized or shed to murine cells. In addition, HSPC with the integrated vector can be detected in several tissues, including germlines. The detection of HIV sequences in the ovaries and/or brain of 5 mice together with the absence of HIV p24 gag antigen in the serum excluded the possibility that an RCL may have been generated. The safety features of the third-generation packaging system combined with a SIN transfer vector used in this study make it extremely unlikely that a productive viral recombinant is generated during vector production or cell transduction. Thus, the most likely interpretation of the findings is an incomplete perfusion procedure, so that residual blood containing high VCNs and human cells also in the MOCK group can be measured in these mice. As in the murine study, we performed IS analysis of the infused product to determine the safety profile of the GLOBE LV medicinal product. The transduced CD34$^+$ cells showed a high polyclonal pattern of integration with no expansion or selection of specific ISs. Moreover, the genes preferentially targeted by GLOBE LV were PACS1, KDM2A, and others, which are well known hotspots in preclinical studies of LVs[34,65] and in lentiviral gene therapy trials and have not been associated with clonal expansion.[3,4,46] We were able to detect several ISs shared among individual mice, suggesting that our transduction protocol maintains engraftment potential and that hematopoietic primitive cells remain multipotent and are able to generate new blood cells.

In summary, our data demonstrate that the GMP-grade β-globin encoding GLOBE LV can efficiently transduce CD34$^+$ cells under validated clinical conditions with no toxicity. These studies provide important information about the quality and the safety data of transduced HSCs for thalassemic patients, complementing previous *in vitro* and *in vivo* studies performed in the murine th3/+ model, which showed efficient engraftment of vector-transduced cells and immune reconstitution without the occurrence of LV-induced toxicity and tumorigenicity tumors. Thus, this work strongly supports the translation of our gene therapy approach to the clinic as a phase I and II gene therapy clinical protocols based on infusion of GLOBE LV-transduced autologous CD34$^+$ cells. Our trial received approval from regulatory authorities and has recently started (ClinicalTrials. gov identifier NCT02453477).

MATERIALS AND METHODS
LVs
The GLOBE and GATA-GLOBE LVs have been described previously.[26,37] The cHS4-GLOBE LV was generated by cloning a BamHI 1.2-kb fragment containing the HS4 element from the chicken genome[18,66] into SpeI-SfiI sites of the LTR of the GLOBE vector backbone.[26]

Research-grade viral vector stocks were prepared by transient cotransfection of HEK293T cells using a three-plasmid system as described previously.[26]

For toxicology and biodistribution studies, we derived the pCCL.GLOBE Kana LV from the original PRRL GLOBE LV described previously. The pRRLsincpptW1.6Wasp-WPREmut6-Kana (a kind gift from L. Naldini) provided the main backbone.

Medium- and large-scale purified vector lots for the GLP studies were produced by MolMed (Milan, Italy). Vectors were produced by a medium- or large-scale process based on transient quadri-transfection in 293T cells, followed by purification through endonuclease

treatment, anion exchange chromatography with gradient elution, gel filtration, resuspension in serum-free medium, 0.22-μm filtration, and aseptic filling. Each lot was characterized in terms of titer, potency, purity, and safety aspects (titer of the vector in the toxicology study, 9.69×10^7 TU/mL; titer of the vector in the biodistribution study, 6.6×10^8 TU/mL).

Transduction and Culture of Cell Lines

The HEL cell line was purchased from the ECACC (European Collection of Cell Cultures). HEL cells were grown in RPMI 1640 medium (Lonza, Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS) (GE Healthcare Life Science, Little Chalfont, UK), 100 U/mL penicillin and streptomycin, and 2 mM L-glutamine (Lonza, Basel, Switzerland) and transduced with viral stocks at a multiplicity of infection (MOI) of 20 in the presence of 8 μg/mL Polybrene (Sigma-Aldrich, St. Louis, MO).

Terminal erythroid differentiation was obtained by culturing the cells in the presence of 50 μM hemin (Sigma-Aldrich, St. Louis, MO) for 5 days.

Single cell-derived clones were obtained by limiting dilution in 96-well plates (0.3 cells/well). Clones were screened by vector-specific PCR.

Southern and Northern Blot Analyses

Genomic DNA was extracted using the QIAamp DNA blood mini kit (QIAGEN, Hilden, Germany). Ten micrograms of DNA were digested overnight with AflII, EcoRV, SpeI, XbaI, or KpnI (all from New England Biolabs, UK), run on a 0.8% agarose gel, transferred to a nylon membrane (Hybond-N; GE Healthcare Life Science, Little Chalfont, UK) by Southern capillary transfer, probed with 2×10^7 dpm of a 32P-labeled HS3 probe, and exposed to X-ray film.

Total cellular RNA was extracted by TRI reagent (Sigma-Aldrich, St. Louis, MO), size-fractionated on 1% agarose-formaldehyde gel, blotted onto nylon membranes, hybridized to 10^7 dpm of 32P-labeled GATA-1 and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) probes, and exposed to X-ray film.

Western Blot Analysis

Protein extraction was performed by resuspending cells in lysis buffer (150 mM NaCl, 1% v/v NP40, 0.5% w/v deoxycholic acid, 0,1% w/v SDS, and 50 mM Trizma base [pH 7.8]) and centrifuged to remove cellular debris. Five micrograms of protein were then mixed with NuPAGE LDS sample buffer and NuPAGE sample-reducing agent (Invitrogen, Carlsbad, CA), boiled, and separated by SDS-PAGE on a 10% polyacrylamide gel. After being transferred to nitrocellulose membranes, the blots were blocked with PBS containing 0.1% Tween 20 and 5% milk for 1 hr at room temperature. The blots were probed for 1 hr at room temperature with an appropriate dilution of the primary antibodies. After washing with PBS containing 0.1% Tween 20, horseradish peroxidase-conjugated anti-mouse immunoglobulin G (IgG) was applied to the membrane for 1 hr at room temperature.

The signals were then visualized with an enhanced chemiluminescence (ECL) kit (GE Healthcare Life Science, Little Chalfont, UK), and the intensity of the bands was quantified by ImageJ software (https://imagej.nih.gov/ij/).

Quantitative Analysis of Gene Expression

Total cellular RNA was extracted by TRI reagent (Sigma-Aldrich, St. Louis, MO), purified using the Purelink RNA mini kit (Invitrogen, Carlsbad, CA), and treated with DNase I (Sigma-Aldrich, St. Louis, MO) to remove DNA contaminations. The absence of genomic DNA was verified by genome-specific PCR. Reverse transcription was performed using a high-capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA), following the manufacturer's instructions.

Gene expression levels were evaluated on a custom low-density array card ("Assay on Demand," Applied Biosystems, Foster City, CA) using an ABI PRISM 7900 HT system. Relative quantification was calculated using the ΔCt method.[67] ΔCt was calculated for every gene as $Ct_{(gene)}$ − average $Ct_{(GAPDH, β2\ microglobulin, RPLP0)}$. ΔΔCt was calculated for test clones as $Ct_{(test)}$ − average $Ct_{(controls)}$. The relative quantification (RQ) of each transcript was calculated as $2^{-ΔΔCt}$ and plotted as $\log_2$ values.

Analysis of VCN

To evaluate the number of LV copies integrated per genome, qPCR was performed using specific primers and probes for human telomerase, murine β-actin, and a LV, as described previously.[33] A reference standard was obtained from serially diluted transduced human T cell lines carrying one copy of the integrated LV or murine NIH 3T3 cell line carrying six copies of LV. The results of integrated vector copies were normalized for the number of evaluated genomes. As a negative control, samples of untransduced cells were used. All the reactions were performed according to the manufacturer's instructions and analyzed with an ABI PRISM 7900 sequence detection system or ViiA7 real-time PCR system (Applied Biosystems, Foster City, CA).

GLOBE IS Mapping and Analysis

ISs were cloned by LAM-PCR and LM-PCR.

For LAM-PCR, after linear PCR amplification of genome-vector junctions, followed by magnetic capture and second strand reconstitution, digestion with Tsp509I, HpyCH4IV, or MseI was performed. Digestion products were ligated with a restriction site-complementary linker cassette and underwent two sequential rounds of exponential PCR steps. The LAM-PCR products were tagged with megaprimers and underwent next-generation sequencing on the MiSeq instrument (Illumina). The resulting genomic sequences were automatically aligned to the mouse genome (July 2007 assembly mm9 of the University of California Santa Cruz (UCSC) genome browser; http://genome.ucsc.edu/).

The relative abundance of ISs was computed from the matrix files cleaned by contaminations (also called collisions). For each sample,

the abundance value was obtained as the percentage for each IS of the sequence reads over the total reads of the samples. When applicable, abundance was also correct for engraftment as abundance/engraftment × 100.

All GO analyses were made using the Genomic Regions Enrichment of Annotations Tool (GREAT) online software (http://bejerano.stanford.edu/great/public/html/). We uploaded the genomic coordinates of the integrations of each dataset and calculated the enrichment levels in the tested dataset by correlating positional information (based on the binomial distribution analysis for p value calculations) and the annotated function of the genes closest to the ISs (based on the hypergeometric distribution analysis for p value calculations). Biological processes of the GO database were chosen for enrichment analysis. Only gene classes with a false discovery rate of less than 0.05 were considered.

For LM-PCR, 0.150–500 ng of genomic DNA was digested with the MseI restriction enzyme (Roche). The digestion products were purified (NucleoSpin Gel and PCR Clean-up, Macherey-Nagel) and ligated to a restriction site-complementary double-stranded DNA linker. The ligation products were purified and digested with the SacI restriction enzyme (Roche) to avoid amplification of the internal provirus. Vector-genome junctions were amplified by LM-PCR, a nested PCR performed with specific primers annealing on the linker and the 3′ viral LTR, containing a 4-bp tag for sample identification. Finally, LM-PCR libraries were quantified, size-selected by gel extraction, checked by capillary electrophoresis (2100 Bioanalyzer Instrument, Agilent Technologies), and tagged with sample-specific dual indexes and Illumina sequencing adapters (Nextera XT Index Kit, Illumina) before being sequenced to saturation on the MiSeq instrument (Illumina). The resulting genomic sequences were demultiplexed by sample-specific tag and bioinformatically processed, trimmed by Skewer[68] and mapped on the human genome (February 2009, GRCh37/hg19) by Bowtie2 software.[69] During the mapping step, a quality score was assigned to each insertion site based on the total number of trimmed reads collapsed in that genomic position (read count) and the mean read quality. For each sample, the abundance value was obtained as the percentage of the read count over the total reads of the samples. DAVID software[70] was used for GO analysis LV-targeted genes (Refseq database), considered as transcriptional units with inside at least one IS.

Mice, Murine HSPC Transduction, and Transplantation

C57BL/6-Hbb[th3] CD45.2 mice were purchased by Jackson Laboratory. C57BL/6-Hbb[th3] CD45.1 mice were generated in our institute. We bred C57BL/6 CD45.1 mice to C57BL/6-Hbb[th3] CD45.2 mice and then intercrossed the C57BL/6 Hbb[th3] CD45.1/CD45.2 progeny and C57BL/6 CD45.1 to generate C57BL/6-Hbb[th3] CD45.1 mice. All procedures were performed according to protocols approved by the Committee for Animal Care and Use of the San Raffaele Scientific Institute (Committee Protocol 537).

Recipient C57BL/6-Hbb[th3] Ly-5.1 mice were treated intraperitoneally with busulfan (Sigma-Aldrich, St. Louis, MO) at 25 mg/kg/day for 4 days to obtain myeloablation prior to transplantation. Prophylactic antibacterial therapy was given by administration of gentamycin sulfate 80 mg/252 mL (0.3 mg/mL) by water intake from day −7 to day +7. Donor C57BL/6-Hbb[th3] mice were killed by CO_2 lethal inhalation, and BM from both tibia and femora was harvested into Miltenyi Biotec buffer to obtain Lin⁻ HSPCs. Lin⁻ HSPCs were purified with a Lineage Cell Depletion Kit Mouse (Miltenyi Biotec, Auburn, CA), seeded in StemSpan medium (STEMCELL Technologies, Vancouver, BC, Canada) supplemented with mIL3 (10 ng/mL), mIL6 (20 ng/mL), mFLT3L (10 ng/mL), and murine stem cell factor (mSCF; 50 ng/mL) (all from PeproTech, Rocky Hill, NY). Lin⁻ HSPCs were transduced with GLOBE LV at MOI 100 for 16–18 hr. MOCK-transduced cells were cultured under the same conditions in the absence of GLOBE LV to prepare the control item.

Transduced or MOCK-transduced cells were injected (4×10^5 cells per mouse) i.v. in 8- to 12-week-old C57BL6/Hbb[th3] mice.

In vitro liquid culture of transduced or MOCK-transduced Lin⁻ HSPC progeny was performed to conduct VCN and IS analysis. 1×10^6 Lin⁻ HPSCs were seeded in liquid culture medium supplemented with 5% FBS with mIL3 (10 ng/mL), mFLT3L (10 ng/mL), mIL6 (20 ng/mL), and mSCF (50 ng/mL) (all from PeproTech, Rocky Hill, NY) and cultured for 12 ± 2 days. 2.5×10^3 transduced or MOCK-transduced Lin⁻ HPSCs were seeded in methylcellulose medium (MethoCult M3434; STEMCELL Technologies, Vancouver, BC, Canada), colonies were scored on day 10 ± 2, and DNA was prepared for VCN analysis.

Post-transplant Monitoring

Mortality and morbidity were checked once a day each working day. Animals were observed individually once a week for the recording of clinical signs and body weight. Blood samples for hematology analyses and DNA analyses were collected 16 weeks after transplantation and at termination. Erythrocyte count, mean cell volume, packed cell volume, hemoglobin, mean cell hemoglobin concentration, mean cell hemoglobin, thrombocyte count, leucocyte count, differential white cell count with cell morphology, and reticulocyte count were determined using the Scil Vet ABC Plus⁺ analyzer (Scil Animal Care, Treviglio, Italy).

Cytofluorimetric analyses of PB were performed on a Becton Dickinson LSR II machine with BD FACSDiva software. PB from mice was analyzed for hematopoietic reconstitution using the following antibodies: phycoerythrin (PE)-conjugated anti-mouse CD11b, anti-mouse CD3, and anti-mouse B220 (BD Biosciences, San Jose, CA). Fluorescein isothiocyanate (FITC)-conjugated anti-mouse CD45.1 and allophycocyanin (APC)-conjugated anti-mouse CD45.2 antibodies (BD Biosciences, San Jose, CA) were used to evaluate donor-host chimerism at 16 weeks and termination.

Genomic DNA was extracted from the blood for determination of VCN at 16 weeks using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany).

Necropsy, Histology, and Pathology

Animals found dead during the study or sacrificed in poor condition were subjected to necropsy, and selected tissues were harvested and fixed in 10% neutral buffered formalin (Sigma-Aldrich, St. Louis, MO). At termination, animals were shipped to Accelera (Nerviano, Italy), where animals were sacrificed, and a full necropsy was executed in compliance with the Italian GLP Regulations (DL 50; March 2, 2007; G.U. 8, April 13, 2007) and the OECD Principles of GLP (as revised in 1997, ENV/MC/CHEM(98)17). Tissues were sent to GlaxoSmithKline Research and Development for histology processing, and pathological examination was performed in accordance with United Kingdom GLP (statutory instruments number 3106 and 2004 number 994) and the OECD Principles of GLP (as revised in 1997, ENV/MC/CHEM(98)17). In case of macroscopic lesions, a part of the tissue was embedded in optimal cutting temperature (O.C.T., Tissue-Tek Sakura) medium and frozen in liquid nitrogen for possible further characterization.

The following organs were collected: adrenal gland, brain including medulla and pons, cerebellar and cerebral cortex, epididymis, heart, kidneys, large intestine (cecum, colon, and rectum), liver, lungs with bronchi, lymph nodes (mesenteric and mandibular), ovaries with oviduct, pancreas, skeletal muscle, skin, small intestine (duodenum, ileum, and jejunum), spleen, sternum with BM, stomach with forestomach, testes, thymus, uterus, and vagina. Tissues were embedded in paraffin wax, sectioned, mounted on glass slides, and stained with H&E in accordance with GlaxoSmithKline (GSK) standard operation procedures (SOPs). Tissue reading was performed at GSK Research and Development in accordance with GSK SOPs. A histopathology peer review was carried out by a second pathologist in accordance with GLP SR-Tiget SOPs.

For morphological analysis, 3-μm cryosections were fixed with 4% paraformaldehyde and stained with H&E. For immunohistochemical analysis, 3-μm cryosections were fixed with 4% paraformaldehyde at room temperature and incubated with rat anti-mouse B220 (clone RA3-6B2, Serotec), anti-mouse CD3 (clone CD3-12, Serotec), anti-mouse F4/80 (clone CI, A3-1, AbD Serotec), and rabbit Ki67 (clone SP6, Lab Vision) antibodies. 3-μm cryosections were fixed with acetone at 4°C and incubated with biotin-conjugated anti-mouse CD45.1 (clone A20, BD Pharmingen) and with biotin-conjugated anti-mouse CD45.2 (clone 104, BD Pharmingen). The immunoreactions were revealed either by rat or rabbit on rodent HRP polymer (Biocare Medical) by HRP-conjugated streptavidin and using diaminobenzidine (DAB) as chromogen (Biogenex, San Ramon). Slides were counterstained with hematoxylin.

Approximately 500–800 μL of blood was collected from the aorta during necropsy for chemical analyses. Glucose, urea, creatinine, triglycerides, total protein, albumin, albumin-to-globulin ratio, globulin (calculated parameter), alkaline phosphatase, alanine aminotransferase, and aspartate aminotransferase were measured by Accelera.

Human CD34$^+$ Cell Transduction and Xenotransplantation

Cryopreserved human cord blood CD34$^+$ cells were purchased from Lonza (Basel, Switzerland). Cells were thawed and seeded on nontissue culture-treated plates coated with Retronectin (30 μg/mL; TaKaRa Bio, Shiga, Japan) () at 1×10^6 cells/mL in serum-free CellGro stem cell growth medium (SCGM; Cell Genix Technologies, Freiburg, Germany) in the presence of hSCF (300 ng/mL), hFLT3-L (300 ng/mL), thrombopoietin (100 ng/mL), and interleukin-3 (IL-3; 60 ng/mL) (all from PeproTech, Rocky Hill, NY). Following 24 hr of pre-stimulation, cells were transduced twice with GLOBE LV at MOI 100.[71] MOCK-transduced control cells were cultured under the same conditions except for the presence of LV.

For transplantation studies *in vivo*, 2×10^5 transduced or MOCK-transduced cells were administered to NSG mice (Jackson Laboratory) by injection in the lateral tail vein. Mice were maintained in a specific pathogen-free animal facility. Procedures were performed according to protocols approved by the Committee for Animal Care and Use of San Raffaele Scientific Institute (Committee Protocol 537). 9- to 10-week-old NSG female mice were conditioned by a sub-lethal dose of irradiation to provide depletion of endogenous BM cells, allowing engraftment of human cells. The recipient mice were exposed to 210 cGy of total body irradiation in an X-ray Rad Source RS-2000 irradiator (Biological System). Prophylactic antibacterial therapy was given by administration of gentamycin sulfate 80 mg/252 mL (0.3 mg/mL) by water intake from day −1 to day +28.

Ex Vivo Analysis of Human Cells in Murine Tissues

After 12 weeks, mice were anesthetized and perfused by injecting saline solution for approximately 30 min through blood vessels to eliminate hematopoietic contamination from organs before sacrifice. Mice were euthanized, and the BM, spleen, and thymus were collected and analyzed for the presence of human CD45$^+$ cells and the expression of different lineage-specific surface markers. DNA extracted from cells and tissues was analyzed for VCN by qPCR.

p24 Assay

An HIV p24 ELISA (PerkinElmer, code NEK050) was performed on serum (20–40 μL) of treated mice according to the manufacturer's instructions.

Statistical Analyses

GraphPad Prism was used to perform statistical analyses: one-way ANOVA, *post hoc* Bonferroni's multiple comparisons test or Student's t test, and two-tailed log rank Mantel-Cox test. A Z test for proportion was performed for specific analyses.

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.omtm.2018.09.001.

AUTHOR CONTRIBUTIONS

M.R.L. carried out experiments, analyzed the data, and wrote and reviewed the manuscript. Y.P. analyzed the data, helped with interpretation of the results, and wrote the manuscript. F.T., G.M., C.R., M.V., and A. Aprile performed experiments. C.L. carried out the gene expression analysis in HEL cells. A. Ambrosi performed the statistical analysis of gene expression data. F.C. and F.S. performed the histopathological analysis and reviewed the manuscript. A.C. and E.M. carried out the integration site sequencing analysis for the toxicology study. V.P. and F.M. carried out the integration analysis for the biodistribution study. L.N. contributed to discussion of the results. P.C. contributed to the histopathology report and reviewed the manuscript. G.F. provided financial support, designed experiments, interpreted results, and wrote and reviewed the manuscript.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

ACKNOWLEDGMENTS

REFERENCES

1. Dunbar, C.E., High, K.A., Joung, J.K., Kohn, D.B., Ozawa, K., and Sadelain, M. (2018). Gene therapy comes of age. Science 359, eaan4672.

2. Eichler, F., Duncan, C., Musolino, P.L., Orchard, P.J., De Oliveira, S., Thrasher, A.J., Armant, M., Dansereau, C., Lund, T.C., Miller, W.P., et al. (2017). Hematopoietic Stem-Cell Gene Therapy for Cerebral Adrenoleukodystrophy. N. Engl. J. Med. 377, 1630–1638.

3. Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C., et al. (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. Science 341, 1233151.

4. Biffi, A., Montini, E., Lorioli, L., Cesani, M., Fumagalli, F., Plati, T., Baldoli, C., Martino, S., Calabria, A., Canale, S., et al. (2013). Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. Science 341, 1233158.

5. Ribeil, J.A., Hacein-Bey-Abina, S., Payen, E., Magnani, A., Semeraro, M., Magrin, E., Caccavelli, L., Neven, B., Bourget, P., El Nemer, W., et al. (2017). Gene Therapy in a Patient with Sickle Cell Disease. N. Engl. J. Med. 376, 848–855.

6. Mercola, K.E., and Cline, M.J. (1980). Sounding boards. The potentials of inserting new genetic information. N. Engl. J. Med. 303, 1297–1300.

7. Locatelli, F., Kabbara, N., Ruggeri, A., Ghavamzadeh, A., Roberts, I., Li, C.K., Bernaudin, F., Vermylen, C., Dalle, J.H., Stein, J., et al.; Eurocord and European Blood and Marrow Transplantation (EBMT) group (2013). Outcome of patients with hemoglobinopathies given either cord blood or bone marrow transplantation from an HLA-identical sibling. Blood 122, 1072–1078.

8. Cappellini, M.D., Porter, J.B., Viprakasit, V., and Taher, A.T. (2018). A paradigm shift on beta-thalassaemia treatment: How will we manage this old disease with new therapies? Blood Rev. 32, 300–311.

9. Andreani, M., Testi, M., and Lucarelli, G. (2014). Mixed chimerism in haemoglobinopathies: from risk of graft rejection to immune tolerance. Tissue Antigens 83, 137–146.

10. Andreani, M., Testi, M., Battarra, M., and Lucarelli, G. (2011). Split chimerism between nucleated and red blood cells after bone marrow transplantation for haemoglobinopathies. Chimerism 2, 21–22.

11. Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., Down, J., Denaro, M., Brady, T., Westerman, K., et al. (2010). Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. Nature 467, 318–322.

12. Thompson, A.A., Walters, M.C., Kwiatkowski, J., Rasko, J.E.J., Ribeil, J.A., Hongeng, S., Magrin, E., Schiller, G.J., Payen, E., Semeraro, M., et al. (2018). Gene Therapy in Patients with Transfusion-Dependent β-Thalassemia. N. Engl. J. Med. 378, 1479–1493.

13. Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. 118, 3132–3142.

14. Stein, S., Ott, M.G., Schultze-Strasser, S., Jauch, A., Burwinkel, B., Kinner, A., Schmidt, M., Krämer, A., Schwäble, J., Glimm, H., et al. (2010). Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. Nat. Med. 16, 198–204.

15. Braun, C.J., Boztug, K., Paruzynski, A., Witzel, M., Schwarzer, A., Rothe, M., Modlich, U., Beier, R., Göhring, G., Steinemann, D., et al. (2014). Gene therapy for Wiskott-Aldrich syndrome–long-term efficacy and genotoxicity. Sci. Transl. Med. 6, 227ra33.

16. Naldini, L. (2011). Ex vivo gene transfer and correction for cell-based therapies. Nat. Rev. Genet. 12, 301–315.

17. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., et al. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. J. Clin. Invest. 119, 964–975.

18. Arumugam, P.I., Scholes, J., Perelman, N., Xia, P., Yee, J.K., and Malik, P. (2007). Improved human beta-globin expression from self-inactivating lentiviral vectors carrying the chicken hypersensitive site-4 (cHS4) insulator element. Mol. Ther. 15, 1863–1871.

19. Cassani, B., Montini, E., Maruggi, G., Ambrosi, A., Mirolo, M., Selleri, S., Biral, E., Frugnoli, I., Hernandez-Trujillo, V., Di Serio, C., et al. (2009). Integration of retroviral vectors induces minor changes in the transcriptional activity of T cells from ADA-SCID patients treated with gene therapy. Blood 114, 3546–3556.

20. Hargrove, P.W., Kepes, S., Hanawa, H., Obenauer, J.C., Pei, D., Cheng, C., et al. (2008). Globin lentiviral vector insertions can perturb the expression of endogenous genes in beta-thalassemic hematopoietic cells. Mol. Ther. 16, 525–533.

21. Recchia, A., Bonini, C., Magnani, Z., Urbinati, F., Sartori, D., Muraro, S., Tagliafico, E., Bondanza, A., Stanghellini, M.T., Bernardi, M., et al. (2006). Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. Proc. Natl. Acad. Sci. USA 103, 1457–1462.

22. Moiani, A., Paleari, Y., Sartori, D., Mezzadra, R., Miccio, A., Cattoglio, C., Cocchiarella, F., Lidonnici, M.R., Ferrari, G., and Mavilio, F. (2012). Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. J. Clin. Invest. 122, 1653–1666.

23. Cesana, D., Sgualdino, J., Rudilosso, L., Merella, S., Naldini, L., and Montini, E. (2012). Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. J. Clin. Invest. 122, 1667–1676.

24. Imren, S., Payen, E., Westerman, K.A., Pawliuk, R., Fabry, M.E., Eaves, C.J., Cavilla, B., Wadsworth, L.D., Beuzard, Y., Bouhassira, E.E., et al. (2002). Permanent and panerythroid correction of murine beta thalassemia by multiple lentiviral integration in hematopoietic stem cells. Proc. Natl. Acad. Sci. USA 99, 14380–14385.

25. Malik, P., Arumugam, P.I., Yee, J.K., and Puthenveetil, G. (2005). Successful correction of the human Cooley's anemia beta-thalassemia major phenotype using a lentiviral vector flanked by the chicken hypersensitive site 4 chromatin insulator. Ann. N Y Acad. Sci. 1054, 238–249.

26. Miccio, A., Cesari, R., Lotti, F., Rossi, C., Sanvito, F., Ponzoni, M., Routledge, S.J., Chow, C.M., Antoniou, M.N., and Ferrari, G. (2008). In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia. Proc. Natl. Acad. Sci. USA 105, 10547–10552.

27. Pawliuk, R., Westerman, K.A., Fabry, M.E., Payen, E., Tighe, R., Bouhassira, E.E., Acharya, S.A., Ellis, J., London, I.M., Eaves, C.J., et al. (2001). Correction of sickle cell disease in transgenic mouse models by gene therapy. Science 294, 2368–2371.

28. May, C., Rivella, S., Callegari, J., Heller, G., Gaensler, K.M., Luzzatto, L., and Sadelain, M. (2000). Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. Nature *406*, 82–86.

29. Negre, O., Bartholomae, C., Beuzard, Y., Cavazzana, M., Christiansen, L., Courne, C., Deichmann, A., Denaro, M., de Dreuzy, E., Finer, M., et al. (2015). Preclinical evaluation of efficacy and safety of an improved lentiviral vector for the treatment of β-thalassemia and sickle cell disease. Curr. Gene Ther. *15*, 64–81.

30. Benhamida, S., Pflumio, F., Dubart-Kupperschmitt, A., Zhao-Emonet, J.C., Cavazzana-Calvo, M., Rocchiccioli, F., Fichelson, S., Aubourg, P., Charneau, P., and Cartier, N. (2003). Transduced CD34+ cells from adrenoleukodystrophy patients with HIV-derived vector mediate long-term engraftment of NOD/SCID mice. Mol. Ther. *7*, 317–324.

31. Ficara, F., Superchi, D.B., Hernández, R.J., Mocchetti, C., Carballido-Perrig, N., Andolfi, G., Deola, S., Colombo, A., Bordignon, C., Carballido, J.M., et al. (2004). IL-3 or IL-7 increases ex vivo gene transfer efficiency in ADA-SCID BM CD34+ cells while maintaining in vivo lymphoid potential. Mol. Ther. *10*, 1096–1108.

32. Will, E., Bailey, J., Schuesler, T., Modlich, U., Balcik, B., Burzynski, B., Witte, D., Layh-Schmitt, G., Rudolph, C., Schlegelberger, B., et al. (2007). Importance of murine study design for testing toxicity of retroviral vectors in support of phase I trials. Mol. Ther. *15*, 782–791.

33. Scaramuzza, S., Biasco, L., Ripamonti, A., Castiello, M.C., Loperfido, M., Draghici, E., Hernandez, R.J., Benedicenti, F., Radrizzani, M., Salomoni, M., et al. (2013). Preclinical safety and efficacy of human CD34(+) cells transduced with lentiviral vector for the treatment of Wiskott-Aldrich syndrome. Mol. Ther. *21*, 175–184.

34. Biffi, A., Bartolomae, C.C., Cesana, D., Cartier, N., Aubourg, P., Ranzani, M., Cesani, M., Benedicenti, F., Plati, T., Rubagotti, E., et al. (2011). Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. Blood *117*, 5332–5339.

35. Cesani, M., Plati, T., Lorioli, L., Benedicenti, F., Redaelli, D., Dionisio, F., Biasco, L., Montini, E., Naldini, L., and Biffi, A. (2015). Shedding of clinical-grade lentiviral vectors is not detected in a gene therapy setting. Gene Ther. *22*, 496–502.

36. Roselli, E.A., Mezzadra, R., Frittoli, M.C., Maruggi, G., Biral, E., Mavilio, F., Mastropietro, F., Amato, A., Tonon, G., Refaldi, C., et al. (2010). Correction of beta-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients. EMBO Mol. Med. *2*, 315–328.

37. Miccio, A., Poletti, V., Tiboni, F., Rossi, C., Antonelli, A., Mavilio, F., and Ferrari, G. (2011). The GATA1-HS2 enhancer allows persistent and position-independent expression of a β-globin transgene. PLoS ONE *6*, e27955.

38. Martin, P., and Papayannopoulou, T. (1982). HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression. Science *216*, 1233–1235.

39. Philip, M., Rowley, D.A., and Schreiber, H. (2004). Inflammation as a tumor promoter in cancer induction. Semin. Cancer Biol. *14*, 433–439.

40. Rask-Nielsen, R., and Ebbesen, P. (1965). Reticular neoplasms induced in DBA-2 and CBA mice by intraperitoneal injections of mineral oil. J. Natl. Cancer Inst. *35*, 83–94.

41. Spinozzi, G., Calabria, A., Brasca, S., Beretta, S., Merelli, I., Milanesi, L., and Montini, E. (2017). VISPA2: a scalable pipeline for high-throughput identification and annotation of vector integration sites. BMC Bioinformatics *18*, 520.

42. Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S., and Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. Nat. Rev. Microbiol. *3*, 848–858.

43. Abel, U., Deichmann, A., Bartholomae, C., Schwarzwaelder, K., Glimm, H., Howe, S., Thrasher, A., Garrigue, A., Hacein-Bey-Abina, S., Cavazzana-Calvo, M., et al. (2007). Real-time definition of non-randomness in the distribution of genomic events. PLoS ONE *2*, e570.

44. Choi, B., Chun, E., Kim, M., Kim, S.T., Yoon, K., Lee, K.Y., and Kim, S.J. (2011). Human B cell development and antibody production in humanized NOD/SCID/IL-2Rγ(null) (NSG) mice conditioned by busulfan. J. Clin. Immunol. *31*, 253–264.

45. Visigalli, I., Delai, S., Ferro, F., Cecere, F., Vezzoli, M., Sanvito, F., Chanut, F., Benedicenti, F., Spinozzi, G., Wynn, R., et al. (2016). Preclinical testing of the safety and tolerability of LV-mediated above normal alpha-L-iduronidase expression in murine and human hematopoietic cells using toxicology and biodistribution GLP studies. Hum. Gene Ther. *27*, 813–829.

46. Cartier, N., Hacein-Bey-Abina, S., Bartholomae, C.C., Veres, G., Schmidt, M., Kutschera, I., Vidaud, M., Abel, U., Dal-Cortivo, L., Caccavelli, L., et al. (2009). Hematopoietic stem cell gene therapy with a lentivector in X-linked adrenoleukodystrophy. Science *326*, 818–823.

47. Bosè, F., Fugazza, C., Casalgrandi, M., Capelli, A., Cunningham, J.M., Zhao, Q., Jane, S.M., Ottolenghi, S., and Ronchi, A. (2006). Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. Mol. Cell. Biol. *26*, 3942–3954.

48. Jawaid, K., Wahlberg, K., Thein, S.L., and Best, S. (2010). Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus. Blood Cells Mol. Dis. *45*, 140–146.

49. Rivella, S., and Sadelain, M. (1998). Genetic treatment of severe hemoglobinopathies: the combat against transgene variegation and transgene silencing. Semin. Hematol. *35*, 112–125.

50. Hassan, N.F., Chehimi, J., Ho, W.Z., Campbell, D.E., and Douglas, S.D. (1994). Effect of hematopoietic growth factors on human blood monocytes/macrophages in in vitro culture. Clin. Diagn. Lab. Immunol. *1*, 620–625.

51. Hassan, M., Oberg, G., Bekassy, A.N., Aschan, J., Ehrsson, H., Ljungman, P., Lönnerholm, G., Smedmyr, B., Taube, A., Wallin, I., et al. (1991). Pharmacokinetics of high-dose busulphan in relation to age and chronopharmacology. Cancer Chemother. Pharmacol. *28*, 130–134.

52. Vassal, G., Challine, D., Koscielny, S., Hartmann, O., Deroussent, A., Boland, I., Valteau-Couanet, D., Lemerle, J., Lévi, F., and Gouyette, A. (1993). Chronopharmacology of high-dose busulfan in children. Cancer Res. *53*, 1534–1537.

53. Grochow, L.B., Krivit, W., Whitley, C.B., and Blazar, B. (1990). Busulfan disposition in children. Blood *75*, 1723–1727.

54. Nai, A., Pagani, A., Mandelli, G., Lidonnici, M.R., Silvestri, L., Ferrari, G., and Camaschella, C. (2012). Deletion of TMPRSS6 attenuates the phenotype in a mouse model of β-thalassemia. Blood *119*, 5021–5029.

55. Ott, M.G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U., Glimm, H., Kühlcke, K., Schilz, A., Kunkel, H., et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. Nat. Med. *12*, 401–409.

56. Siapati, E.K., Bigger, B.W., Kashofer, K., Themis, M., Thrasher, A.J., and Bonnet, D. (2007). Murine leukemia following irradiation conditioning for transplantation of lentivirally-modified hematopoietic stem cells. Eur. J. Haematol. *78*, 303–313.

57. Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415–419.

58. Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., and Fischer, A. (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N. Engl. J. Med. *348*, 255–256.

59. Ginn, S.L., Liao, S.H., Dane, A.P., Hu, M., Hyman, J., Finnie, J.W., Zheng, M., Cavazzana-Calvo, M., Alexander, S.I., Thrasher, A.J., and Alexander, I.E. (2010). Lymphomagenesis in SCID-X1 mice following lentivirus-mediated phenotype correction independent of insertional mutagenesis and gammac overexpression. Mol. Ther. *18*, 965–976.

60. Mortellaro, A., Hernandez, R.J., Guerrini, M.M., Carlucci, F., Tabucchi, A., Ponzoni, M., Sanvito, F., Doglioni, C., Di Serio, C., Biasco, L., et al. (2006). Ex vivo gene therapy with lentiviral vectors rescues adenosine deaminase (ADA)-deficient mice and corrects their immune and metabolic defects. Blood *108*, 2979–2988.

61. Zhou, S., Ma, Z., Lu, T., Janke, L., Gray, J.T., and Sorrentino, B.P. (2013). Mouse transplant models for evaluating the oncogenic risk of a self-inactivating XSCID lentiviral vector. PLoS ONE *8*, e62333.

62. Son, W.C., and Gopinath, C. (2004). Early occurrence of spontaneous tumors in CD-1 mice and Sprague-Dawley rats. Toxicol. Pathol. *32*, 371–374.

63. Gold, L.S., Manley, N.B., Slone, T.H., and Ward, J.M. (2001). Compendium of chemical carcinogens by target organ: results of chronic bioassays in rats, mice, hamsters, dogs, and monkeys. Toxicol. Pathol. *29*, 639–652.

64. Ward, J.M. (2006). Lymphomas and leukemias in mice. Exp. Toxicol. Pathol. *57*, 377–381.

65. Cattoglio, C., Facchini, G., Sartori, D., Antonelli, A., Miccio, A., Cassani, B., Schmidt, M., von Kalle, C., Howe, S., Thrasher, A.J., et al. (2007). Hot spots of retroviral integration in human CD34+ hematopoietic cells. Blood *110*, 1770–1778.

66. Puthenveetil, G., Scholes, J., Carbonell, D., Qureshi, N., Xia, P., Zeng, L., Li, S., Yu, Y., Hiti, A.L., Yee, J.K., and Malik, P. (2004). Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector. Blood *104*, 3445–3453.

67. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

68. Jiang, L., Edwards, S.M., Thomsen, B., Workman, C.T., Guldbrandtsen, B., and Sørensen, P. (2014). A random set scoring model for prioritization of disease candidate genes using protein complexes and data-mining of GeneRIF, OMIM and PubMed records. BMC Bioinformatics. *15*, 315.

69. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods. *9*, 357–359.

70. Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. *8*, R183.

71. Otsu, M., Anderson, S.M., Bodine, D.M., Puck, J.M., O'Shea, J.J., and Candotti, F. (2000). Lymphoid development and function in X-linked severe combined immuno-deficiency mice after stem cell gene therapy. Mol. Ther. *1*, 145–153.

Supplemental Information

Multiple Integrated Non-clinical Studies

Predict the Safety of Lentivirus-Mediated

Gene Therapy for β-Thalassemia

Maria Rosa Lidonnici, Ylenia Paleari, Francesca Tiboni, Giacomo Mandelli, Claudia Rossi, Michela Vezzoli, Annamaria Aprile, Carsten Werner Lederer, Alessandro Ambrosi, Franck Chanut, Francesca Sanvito, Andrea Calabria, Valentina Poletti, Fulvio Mavilio, Eugenio Montini, Luigi Naldini, Patrizia Cristofori, and Giuliana Ferrari
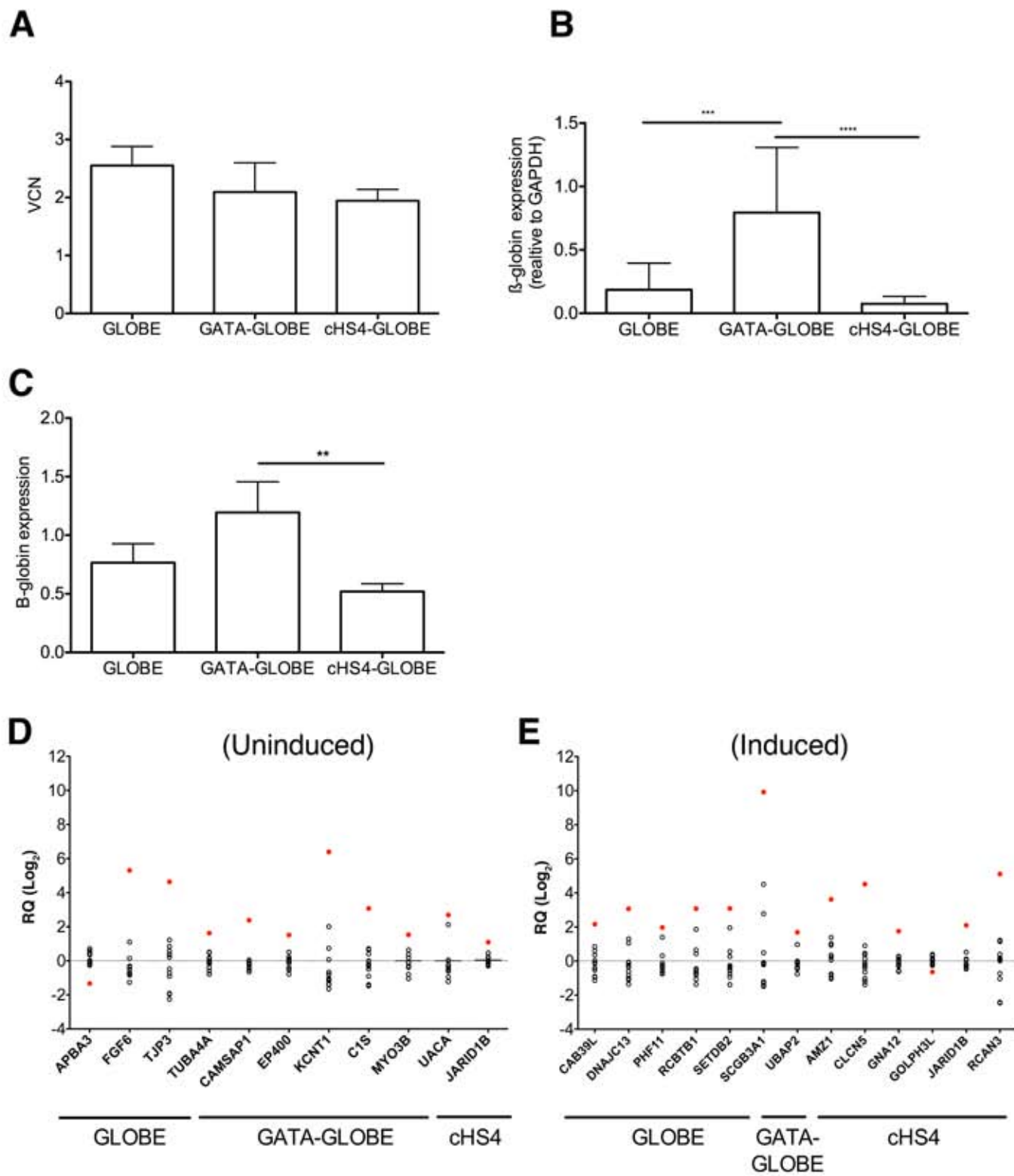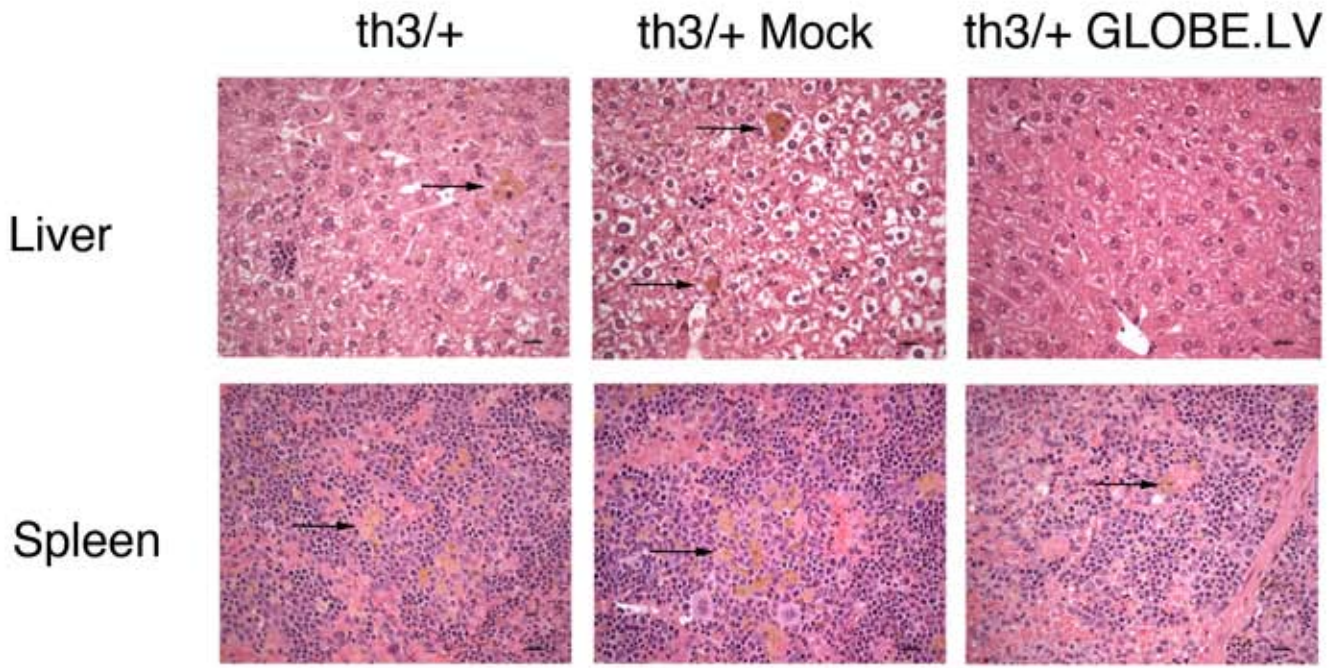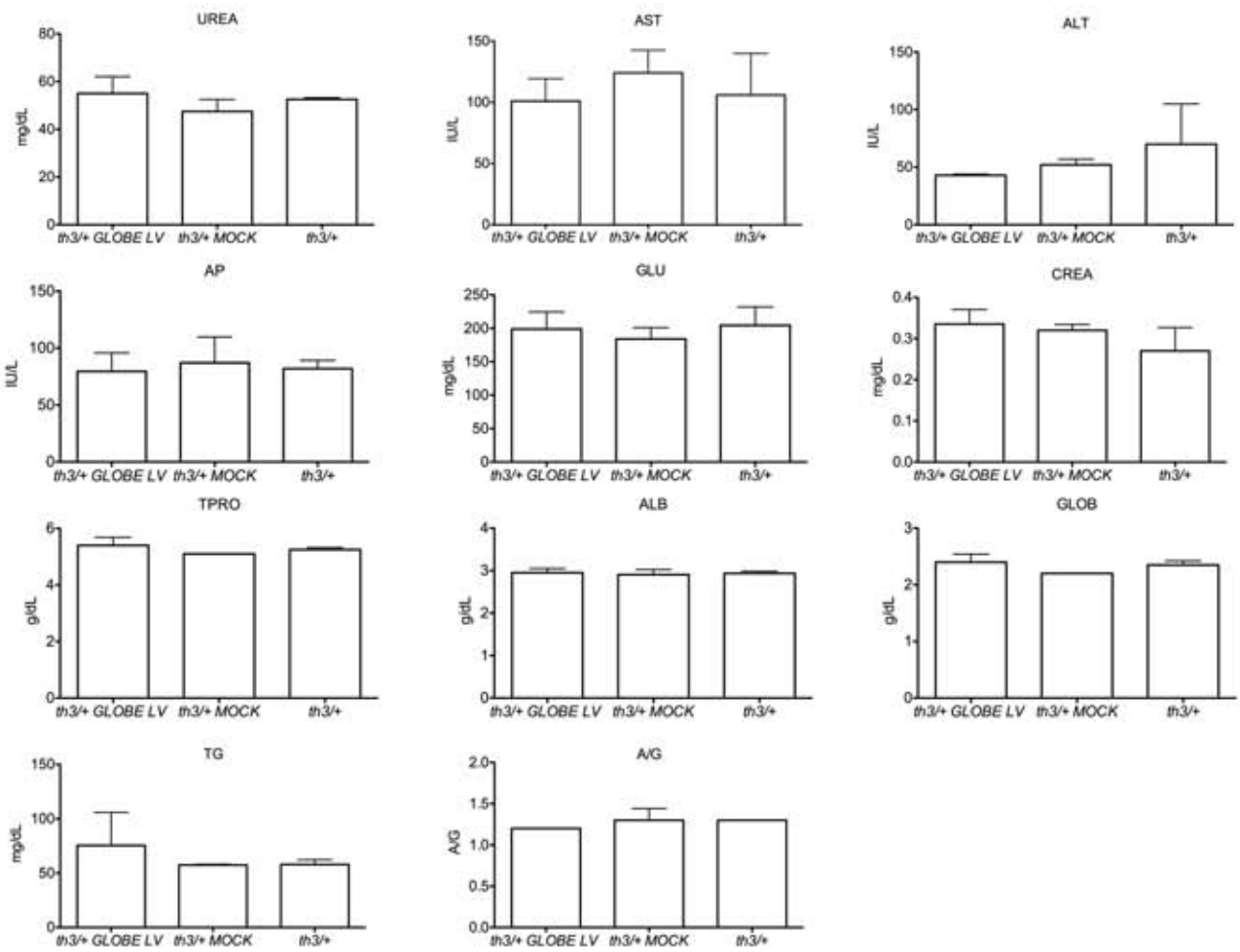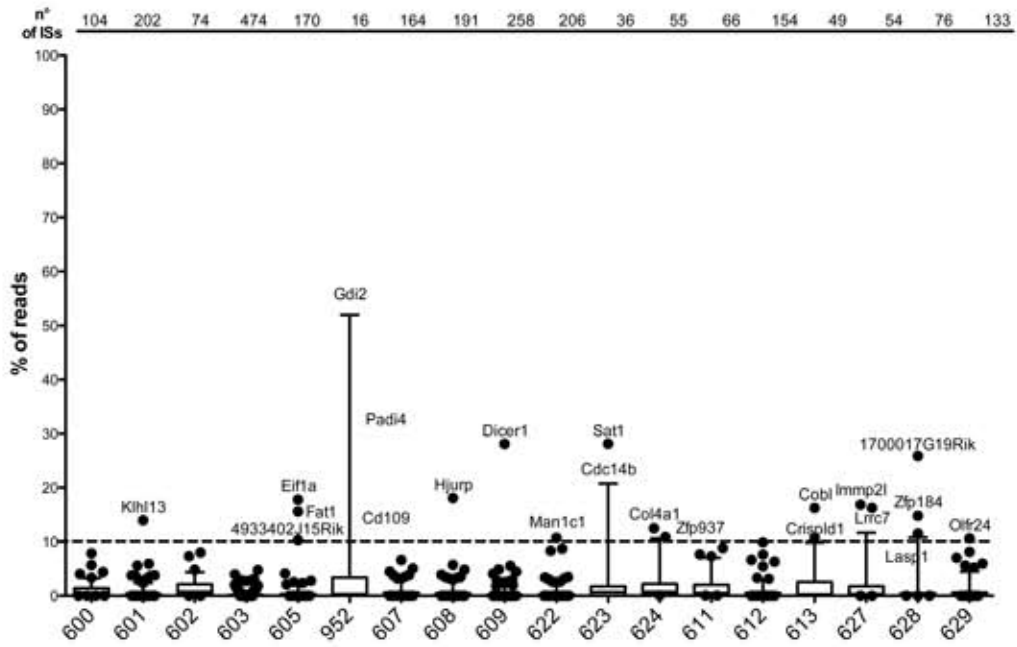
Fig. S1

Fig. S2

A

Fig. S3
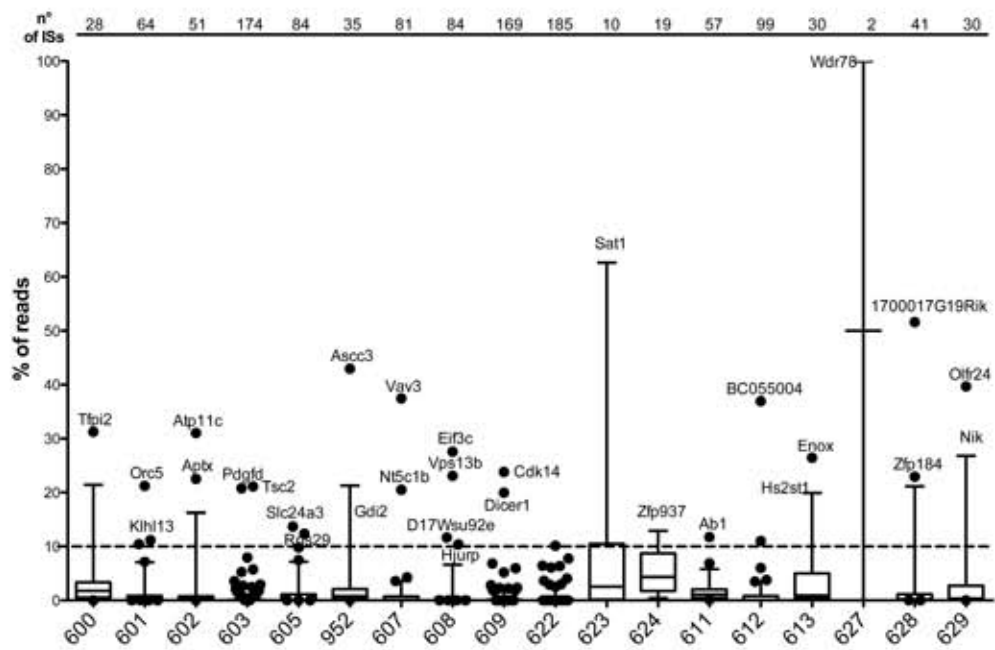
Fig. S4
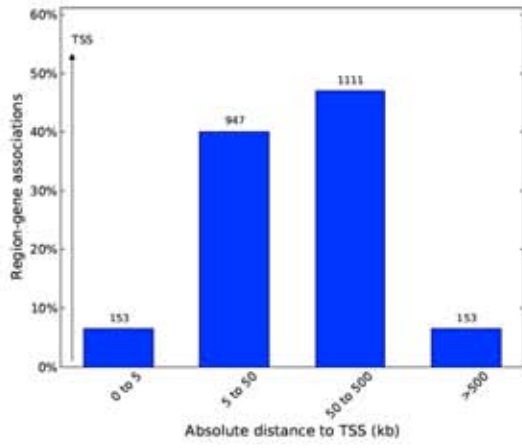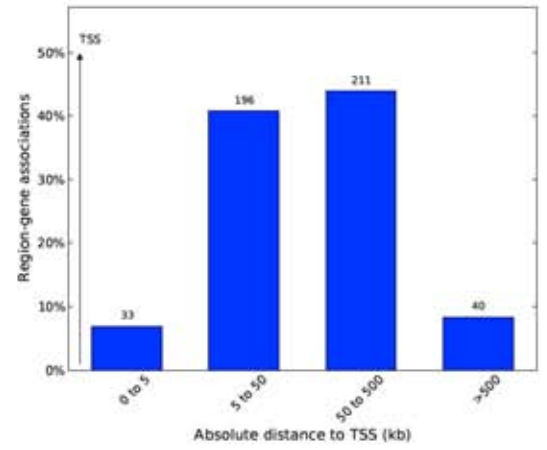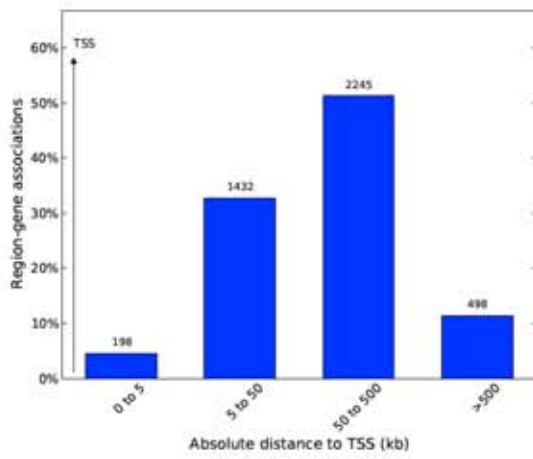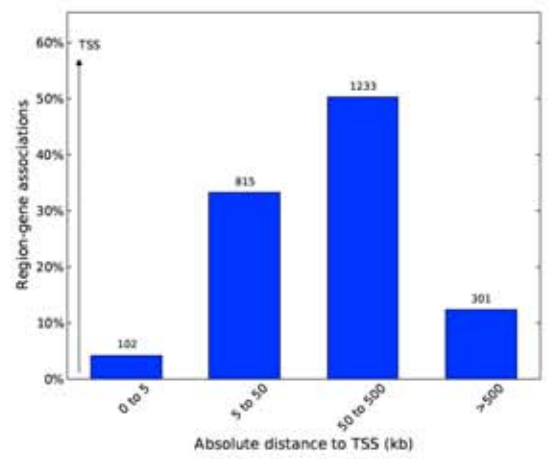
Fig. S5
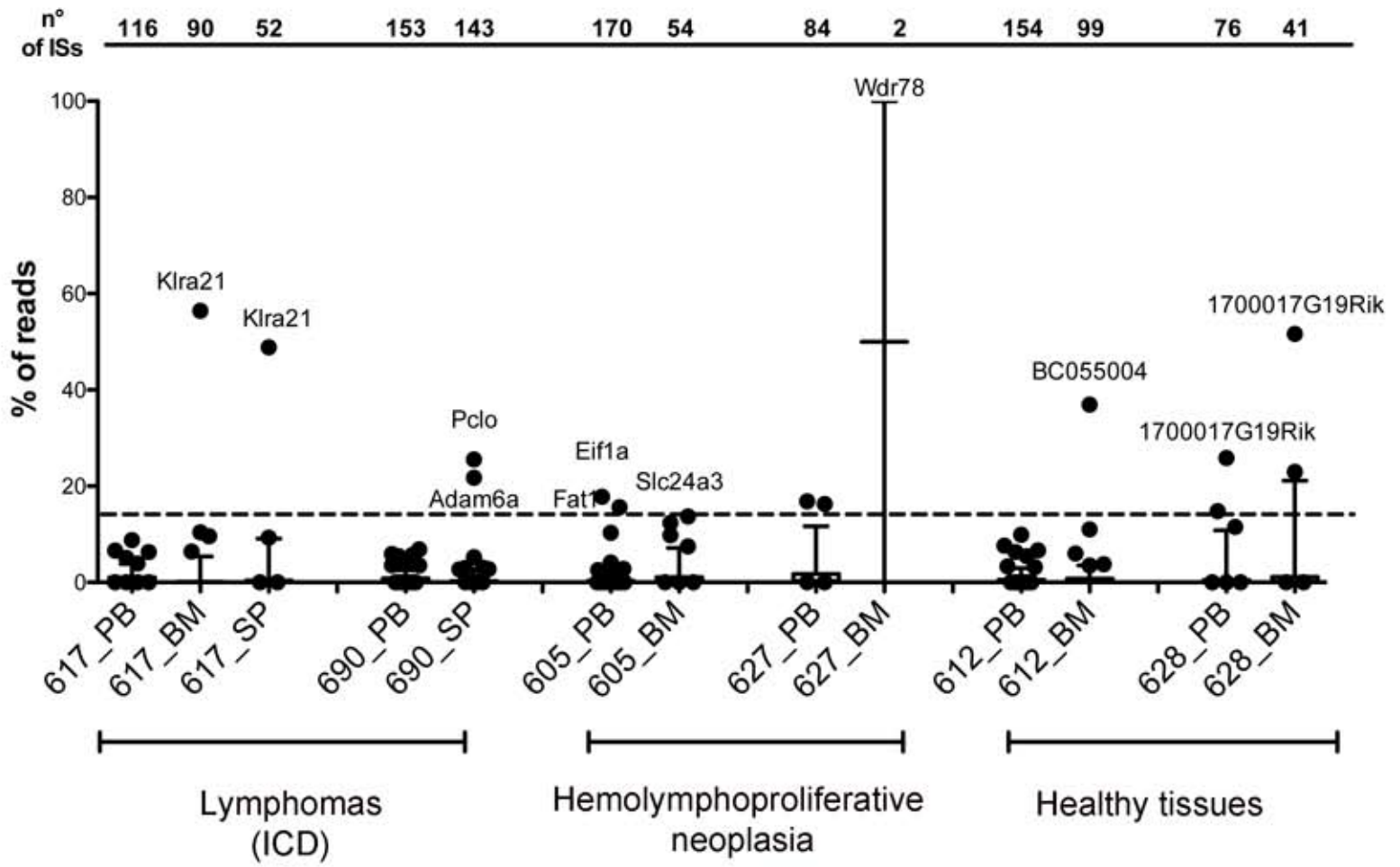
Fig. S6
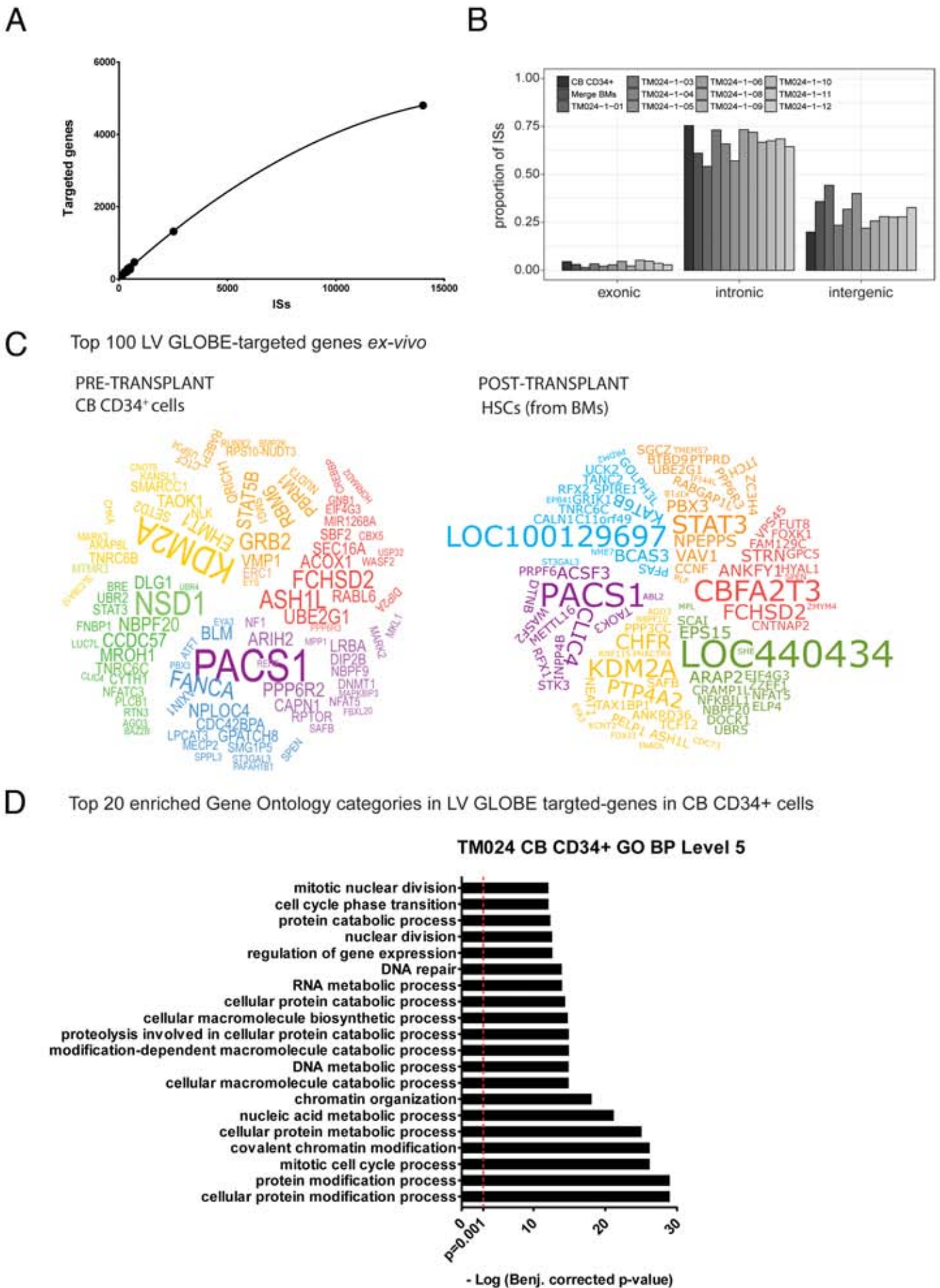
Fig. S7
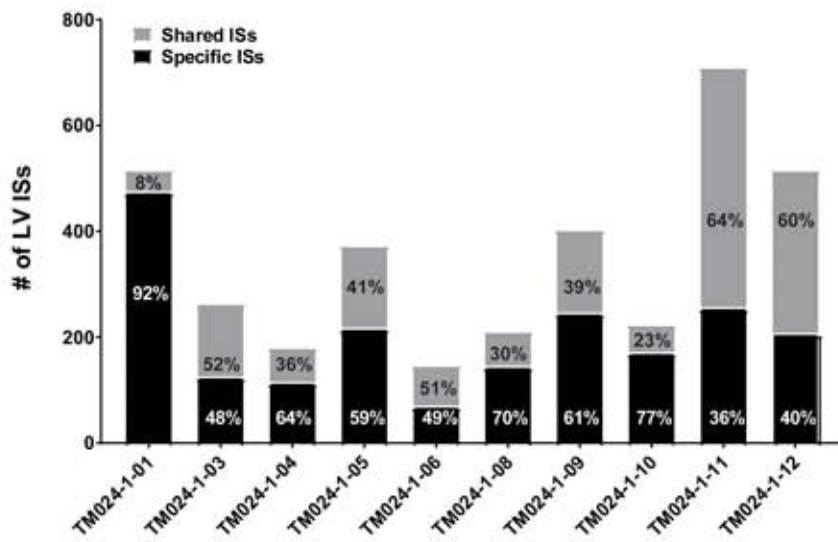
Fig. S8

Supplemental Figure Legends

Figure S1: Characterization of ß-globin expressing LVs and analysis of transcriptional perturbation

A) VCN of HEL cell clones transduced with GLOBE LV, GATA-GLOBE LV or cHS4-GLOBE LV. Values are shown as mean + SD. Student T-test analysis (two tailed) was applied. **B)** qPCR analysis of ß-globin expression in HEL cell clones transduced with GLOBE LV, GATA-GLOBE LV or cHS4-GLOBE LV induced to erythroid differentiation. Data are expressed relative to GAPDH transcript expression and normalized for VCN. Values are shown as mean + SD. Student T-test analysis (two tailed) was applied. *** = p value < 0.001, **** = p value < 0.0001. **C)** Western Blot analysis of ß-globin expression in HEL cell clones transduced with GLOBE LV, GATA-GLOBE LV or cHS4-GLOBE LV induced to erythroid differentiation. Data are expressed relative to GAPDH protein expression and normalized for VCN. Obtained values were normalized respect to clone GATA-GLOBE 12, loaded on every gel as internal control. Values are shown as mean + SD. T-test analysis (two tailed) was applied. ** = p value < 0.05. **D)** qPCR analysis of genes significantly up- or down-regulated (p value < 0.001) by proviral integration in Uninduced HEL cell clones and Induced HEL cell clones **E)**. Red dots indicate the expression value of a gene in the test clone, white dots indicate the expression values of the same gene in control clones.

Figure S2: Extramedullary hematopoiesis in liver and spleen of gene therapy treated mice

Extramedullary hematopoiesis (EMH) in spleen and liver were observed in Th3/+ mice. Related pigmented macrophages (or Kupffer cells) are considered secondary to erythrophagocytosis, Gene therapy treated animals (th3/+ GLOBE.LV) presented signs of efficacy characterized by a prominent reduction of EMH (arrowhead).

Figure S3: Toxicity evaluation of GLOBE LV transduced cells in th3/+ mouse model

A) At termination, PB was collected and used to obtain serum for clinical chemistry. Analyzed parameters were: Urea, Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Alkaline Phosphatase (AP), Glucose (GLU), Creatinine (CREA), (TPRO), Albumin (ALB), Globulin (GLOB), (TG) and the ratio between Albumin and Globulin (A/G). Values are shown as mean + SD. T-test analysis (two tailed) was applied. No difference was observed in all parameters.

Figure S4: Clonal abundance analysis of PB and BM from gene therapy treated mice

Box plot of the percentage of sequence reads (y-axis) for unequivocally mapped IS from peripheral blood at 16 weeks **A)** and bone marrow at termination **B)** is shown. The number of reads for each IS is normalized to the total number of sequence reads from the same time point. IS over the 95 percentiles of dataset are shown as dots distinct from box and whiskers, which are mostly flattened to the bottom of the plot. The total number of IS for each sample is shown on top.

Figure S5: Distribution of LV integrations along the genome of transplanted mice

Integration site (IS) frequency distribution around Transcription Start Site (TSS) of the nearest RefSeq gene, spanning 500 Kb upstream and downstream the TSS. **A)** in vitro cultures **B)** peripheral blood **C)** bone marrow **D)** tumors

Figure S6: Clonal abundance analysis of tumors and controls from gene therapy treated mice

Box plot of the percentage of sequence reads (y-axis) for unequivocally mapped IS from mice which developed lymphomas, hemolymphoproliferative neoplasia and healthy controls. The number of reads for each IS is normalized to the total number of sequence reads from the same time point. IS over the 95 percentiles of dataset are shown as dots distinct from box and whiskers, which are mostly flattened to the bottom of the plot. The total number of IS for each sample is shown on top.

Figure S7. Differential *in vitro* and *ex vivo* integration profile of LV GLOBE.
A) Overall positive correlation between number of ISs and LV-targeted genes in human HSPCs; **B)** Differential distribution of ISs among the intra- and inter-genic regions in pre-transplant CB CD34$^+$ and post-transplant samples; **C)** World cloud graphs of the top 100 LV-targeted genes in HSPCs *in vitro* (left) and *ex vivo* (right). The size of the word is proportional to the number of ISs, while the color has no particular meaning; **D)** The histogram shows the top 20 most significant enriched GO categories in the list of LV-targeted genes in pre-transplant CB CD34$^+$ cells, by DAVID analysis.

Figure S8. ISs shared among different transplanted mice. The histogram shows the portion of ISs specific for each library (black) or common to other libraries (grey).

Table S1: Perturbated genes by different LVs in Uninduced and Induced HEL cells

	Uninduced			Induced		
	GLOBE	GATA-GLOBE	cHS4	GLOBE	GATA-GLOBE	cHS4
Clones	14	11	16	12	10	12
Integrations	36	30	29	35	23	29
Total Genes	92	80	74	87	73	74
p value (of deregulated genes)	0.001	0.001	0.001	0.001	0.001	0.001
Deregulated genes	4	6	1	6	1	6
Up-regulated	3	6	1	6	1	5
Down-regulated	1	0	0	0	0	1
Frequency % (on total genes)	4.35%	7.50%	1.35%	6.90%	1.37%	8.11%

Table S2a : VCN and number of Integration sites in the liquid culture and colony forming units (CFUs)

Sample	Experiment	VCN/cell	Unique IS
CFC	1	9.96	194
LC	1	8.94	267
CFC	2	9.18	70
LC	2	6.36	647
CFC	3	6.35	na
LC	3	6.03	396

Table S2b Percentage of engraftment. VCN and number of integration sites in th3/+ GLOBE LV mice

Mouse	Sex	Engraftment (PB) 16 weeks	VCN/cell PB	Unique IS PB	Engraftment (PB) 12 months	VCN/cell BM	VCN in marked cells (normalized for engraftment)	Unique IS BM	VCN/cell tumor (SP)	Unique IS tumor (SP)
600	M	72.4	9.63	104	70.2	7.58	10.80	28		
601	M	88.8	3.23	202	90.7	8.19	9.03	64		
602	M	93.3	7.62	74	92.6	6.25	6.75	51		
603	M	86.6	5.41	474	96.5	7.25	7.51	174		
604	M	62.7	4	na	54	2.33	4.31	77		
605	M	78.7	7.44	170	88.1	10.87	12.34	84		
607	M	73.6	6.46	164	55.9	7.97	14.26	81		
608	M	82.1	6.17	191	90	7.95	8.83	84		
609	M	87.2	8.53	258	95.7	10.07	10.52	169		
610	M	40.4	5.23	na	55.6	0.18	0.32	28		
611	M	97.3	8.79	66	99.1	2.25	2.27	57		
612	M	100	4.89	154	99.9	2.35	2.35	99		
613	M	96.8	3.47	49	99	7.67	7.75	30		
614	M	99.9	3.55	na	ICD	ICD	ICD	ICD		
690	M	80.9	15.87	153	ICD	ICD	ICD	ICD	7.82	143
616	F	60.8	4.85	na	60.9	2.8	4.60	47		
617	F	67.3	12.92	116	ICD	16.53	ICD	90	8.69	52
618	F	83.5	7.75	na	80.5	16.39	20.36	na		
619	F	83.4	13.12	na	77.4	5.57	7.20	na		
620	F	49.2	2.51	na	5.4	0.28	5.19	na		
621	F	36.3	1.87	na	53.9	1.67	3.10	na		
622*	F	97.2	1.87	206	98.4	7.68	7.80	185		
623	F	12.4	0.88	36	6.4	0.12	1.88	10		
624	F	91	7.21	55	92.6	14.57	15.73	19		
625	F	97.3	2.54	na	ICD	ICD	ICD	ICD		
626	F	82.6	6.2	na	93.2	0.14	0.15	na		
627	F	15.7	1.03	54	12.8	0.14	1.09	2		
628	F	48.9	1.73	76	71.6	0.62	0.87	41		
629	F	35.4	2.22	133	10.5	0.15	1.43	30		
952	F	22	1.32	16	11.5	0.16	1.39	35		

* sacrified at day 292

Table S2c Percentage of engraftment in th3/+ Mock mice

Mouse	Sex	Engraftment (PB) 16 weeks	Engraftment (PB) 12 months
630	M	64.7	65.5
632	M	81.8	75.5
633	M	70	95
634	M	76.5	89.5
635	M	84.2	95.5
636	M	78.3	87
637	M	88.1	92.2
638	M	100	99.9
639	M	84.2	90.7
640	M	99.9	99.9
641	M	99.9	100
642	M	92.4	93.8
643*	M	86.7	91.7
644	M	90.9	95.6
645	F	96.9	97.8
646	F	95.2	79.2
647	F	0.4	0.2
648	F	0.4	0.6
649	F	0.3	0.04
650	F	0.4	0.01
651	F	5.3	1.8
652	F	1	0.4
653	F	2.4	0.6
654	F	62.2	ICD
655	F	5.4	1.7
656	F	55.9	59.9
657	F	66	52
658	F	5.7	2.8
659	F	22.3	26
692	M	90.9	91.3

* sacrified at day 158

Table S3a: Haematological evaluation th3/+ GLOBE LV group at termination

Anima No.	WBC (10³/mm³)	RBC (10⁶/mm³)	HGB (g/dl)	HCT (%)	MCV (um³)	MCH (pg)	MCHC (g/dl)	PLT (10³/mm³)
600	7.0	8.7	13.0	39.5	45.0	14.9	33.0	1013
601	4.7	8.7	12.8	38.2	44.0	14.7	33.5	1078
602	8.9	8.7	13.3	39.4	45.0	15.2	33.7	1209
603	9.8	9.2	13.1	39.1	43.0	14.3	33.6	1058
604	10.1	8.6	13.1	38.6	45.0	15.3	34.0	1181
605	6.8	9.8	14.0	43.8	45.0	14.3	32.0	858
690	na	na	na	na	na	na	na	na
607	9.3	9.6	14.1	44.3	46.0	14.7	31.8	949
608	8.6	9.3	13.7	42.3	46.0	14.8	32.4	956
609	13.0	8.4	13.0	38.8	46.0	15.5	33.5	2454
610	8.6	9.3	13.9	43.3	46.0	14.8	32.1	896
611	9.3	8.4	12.8	37.7	45.0	15.1	33.8	614
612	6.9	8.7	13.8	40.7	47.0	15.8	33.9	1650
613	9.2	9.2	14.0	41.3	45.0	15.2	34.0	881
614	na	na	na	na	na	na	na	na
952	8.9	8.4	10.2	30.8	37.0	12.1	33.1	2177
616	8.3	8.5	11.1	34.4	41.0	13.0	32.1	1451
617	na	na	na	na	na	na	na	na
618	16.8	10.2	13.7	43.1	42.0	13.4	31.9	1054
619	5.0	7.8	11.3	34.2	44.0	14.6	33.1	842
620	14.6	8.0	9.5	29.4	37.0	11.9	32.2	2004
621	6.2	8.8	11.8	36.2	41.0	13.4	32.6	1280
622	na	na	na	na	na	na	na	na
623	4.2	8.9	13.7	39.7	45.0	15.5	34.6	810
624	6.5	8.6	12.9	38.0	44.0	15.0	33.9	920
625	na	na	na	na	na	na	na	na
626	7.2	8.5	12.8	37.8	44.0	15.0	33.8	920
627	2.7	7.1	12.9	36.3	51.0	18.1	35.5	911
628	8.1	8.8	13.7	38.0	43.0	15.6	35.9	968
629	3.0	8.6	14.2	39.4	46.0	16.6	36.0	1040

Table S3b: Haematological evaluation th3/+ Mock group at termination

Anima No.	WBC (10³/mm³)	RBC (10⁶/mm³)	HGB (g/dl)	HCT (%)	MCV (um³)	MCH (pg)	MCHC (g/dl)	PLT (10³/mm³)
630	6.8	3.1	4.3	12.1	39.0	13.8	35.7	1054
692	14.2	7.0	9.1	27.3	39.0	13.0	33.3	2212
632	13.6	6.5	8.5	25.5	39.0	13.0	33.2	1862
633	15.5	7.3	9.5	28.9	40.0	13.0	32.9	1990
634	5.4	2.4	3.6	9.9	41.0	14.8	36.2	773
635	23.2	6.3	8.4	25.3	40.0	13.3	33.1	1985
636	31.0	7.3	9.9	31.3	43.0	13.6	31.7	1892
637	34.6	6.9	10.0	29.2	42.0	14.5	34.4	2256
638	35.0	6.6	9.7	29.5	45.0	14.7	33.0	2235
639	27.1	6.5	9.4	27.5	42.0	14.4	34.2	2129
640	21.5	7.0	10.1	29.8	43.0	14.4	33.9	1902
641	16.1	6.6	9.5	27.7	42.0	14.4	34.4	2090
642	17.7	6.9	9.8	29.0	42.0	14.2	33.9	2100
643	na	na	na	na	na	na	na	na
644	16.3	6.6	9.4	27.8	42.0	14.3	33.8	1800
645	14.0	7.3	9.9	28.9	39.0	13.5	34.1	2286
646	21.2	7.7	10.0	30.0	39.0	13.1	33.4	2133
647	16.3	7.0	9.5	28.4	41.0	13.6	33.4	1935
648	17.7	7.2	9.7	29.3	41.0	13.5	33.1	2006
649	37.6	7.0	10.1	30.9	44.0	14.4	32.7	2210
650	35.1	6.8	9.3	28.5	42.0	13.7	32.5	2231
651	15.0	6.8	9.6	27.3	40.0	14.2	35.3	1915
652	14.7	6.8	9.6	27.2	40.0	14.2	35.3	1717
653	15.8	6.3	9.0	26.2	42.0	14.4	34.4	1759
654	na	na	na	na	na	na	na	na
655	14.0	6.9	9.5	27.6	40.0	13.9	34.5	1977
656	28.1	6.7	9.9	27.5	41.0	14.7	35.9	2032
657	0.0	6.9	9.7	27.9	41.0	14.1	34.7	2054
658	0.0	7.0	9.8	28.5	41.0	14.1	34.5	2123
659	0.0	7.0	10.1	28.9	41.0	14.4	34.9	2073

Table S3c: Haematological evaluation th3/+ group at termination

Anima No.	WBC (10³/mm³)	RBC (10⁶/mm³)	HGB (g/dl)	HCT (%)	MCV (um³)	MCH (pg)	MCHC (g/dl)	PLT (10³/mm³)
660	4.2	6.2	5.4	24.2	39.0	8.7	22.4	1347
661	25.1	8.0	10.3	31.4	39.0	12.9	32.9	2467
662	18.2	7.2	9.4	27.9	39.0	13.1	33.6	2128
663	14.3	7.1	9.3	28.0	40.0	13.2	33.2	2078
664	18.1	5.7	8.3	24.6	43.0	14.5	33.7	1933
665	18.2	5.9	8.5	25.5	44.0	14.5	33.4	1943
666	19.6	6.2	8.9	26.7	43.0	14.4	33.2	1971
935	29.4	6.9	9.8	30.1	44.0	14.2	32.5	1655
691	38.1	6.9	10.0	30.7	44.0	14.5	32.7	2078
669	36.7	6.8	9.6	28.7	42.0	14.1	33.5	2302
670	18.9	6.7	9.5	28.1	42.0	14.3	33.8	1854
671	16.9	6.5	9.4	17.6	42.0	14.5	34.2	1821
672	15.7	6.6	9.6	27.5	42.0	14.5	34.9	2096
673	17.9	6.8	9.8	28.6	42.0	14.4	34.3	2108
674	15.0	7.3	10.1	29.2	40.0	13.9	34.6	1957
675	16.2	7.0	9.2	27.2	39.0	13.2	33.9	2220
676	16.9	7.1	9.4	27.6	39.0	13.3	34.0	2224
677	29.9	7.1	9.7	30.0	42.0	13.7	32.2	1790
678	24.7	7.5	10.3	31.1	42.0	13.9	33.3	2385
679	25.3	7.5	10.4	31.2	42.0	14.0	33.5	2335
680	29.6	7.6	10.0	30.7	41.0	13.2	32.5	2347
681	13.8	6.8	9.5	27.5	40.0	13.9	34.5	1899
682	13.2	6.5	9.0	25.9	40.0	13.9	34.6	1705
683	16.0	6.3	9.2	26.4	42.0	14.5	34.7	1726
684	21.7	7.1	10.1	30.0	42.0	14.2	33.8	1523
685	25.1	7.1	10.6	29.6	42.0	14.8	35.6	2105
686	28.5	7.1	10.0	28.8	41.0	14.1	34.6	2147
687	27.6	6.1	8.8	27.0	44.0	14.4	32.4	1609
688	28.4	6.8	9.6	28.6	42.0	14.2	33.6	2290
689	36.4	6.6	9.9	29.5	45.0	15.0	33.5	2209

Table S4: RTCGD analysis

Sample	in vitro	PB	BM	SP
n° of genes tested	1290	1765	1010	175
n° of genes in RTCGD	286	401	209	30
% of genes in RTCGD	22.2	22.7	20.7	17.1

Table S5. ISs distribution along the human genome

Sample	Raw reads	Mapped reads	# LV ISs	Inter-genic	Intra-genic (Exons)	Intra-genic (Introns)	# LV-targeted genes
CB CD34+	4.611.257	3.287.603	14.023	0.05	0.2	0.75	4.802
TM024-1-01	2.384.753	1.056.020	515	0.02	0.44	0.54	265
TM024-1-03	693.561	438.794	263	0.03	0.24	0.73	191
TM024-1-04	556.507	402.491	180	0.02	0.32	0.66	120
TM024-1-05	879.491	598.145	372	0.03	0.4	0.57	201
TM024-1-06	809.812	525.727	146	0.05	0.22	0.73	107
TM024-1-08	876.857	583.019	210	0.02	0.26	0.72	151
TM024-1-09	950.808	757.092	402	0.05	0.28	0.67	287
TM024-1-10	674.862	277.620	223	0.05	0.28	0.68	149
TM024-1-11	1.222.899	727.632	709	0.04	0.28	0.68	465
TM024-1-12	677.594	475.264	515	0.03	0.33	0.64	327
TM024 merge BMs	9.727.144	5.844.941	2.518	0.03	0.35	0.62	1.316

Table S6. ISs and targeted genes in common between pre- and post-transplant samples.

Targeted genes	TM024 1-01	TM024 1-03	TM024 1-04	TM024 1-05	TM024 1-06	TM024 1-08	TM024 1-09	TM024 1-10	TM024 1-11	TM024 1-12
Tot	265	191	120	201	107	151	287	149	465	327
Specific	119	48	35	52	27	45	90	47	124	91
Shared with CB_CD34+	146	143	85	149	80	106	197	102	341	236
Shared with CB_CD34+ (%)	55%	75%	71%	74%	75%	70%	69%	68%	73%	72%

Individual ISs	TM024 1-01	TM024 1-03	TM024 1-04	TM024 1-05	TM024 1-06	TM024 1-08	TM024 1-09	TM024 1-10	TM024 1-11	TM024 1-12
Tot	515	263	180	372	146	210	402	223	709	515
Specific	515	262	180	371	146	210	400	223	708	515
Shared with CB_CD34+	0	1	0	1	0	0	2	0	1	0
Shared with CB_CD34+ (%)	0.0%	0.4%	0.0%	0.3%	0.0%	0.0%	0.5%	0.0%	0.1%	0.0%

Table S7. Shared ISs between individual pre- and post-transplant samples

	TM024 1-01	TM024 1-03	TM024 1-04	TM024 1-05	TM024 1-06	TM024 1-08	TM024 1-09	TM024 1-10	TM024- 1-11	TM024 1-12	CB CD34+
TM024-1-01	472	0	1	4	0	11	24	3	23	12	-
TM024-1-03	0	124	13	14	6	14	47	12	129	77	1
TM024-1-04	1	13	114	17	3	19	35	13	62	44	-
TM024-1-05	4	14	17	213	8	5	44	5	147	119	1
TM024-1-06	0	6	3	8	66	28	49	17	77	54	-
TM024-1-08	11	14	19	5	28	145	48	26	55	49	-
TM024-1-09	24	47	35	44	49	48	242	30	137	124	2
TM024-1-10	3	12	13	5	17	26	30	171	52	32	-
TM024-1-11	23	129	62	147	77	55	137	52	251	287	1
TM024-1-12	12	77	44	119	54	49	124	32	287	206	-
CB CD34+	0	1	0	1	0	0	2	0	1	0	14.023