

# Ex Vivo Selection of Transduced Hematopoietic Stem Cells for Gene Therapy of $\beta$ -Hemoglobinopathies

Kanit Bhukhai,<sup>1,2,11</sup> Edouard de Dreuzy,<sup>1,2,11</sup> Marie Giorgi,<sup>1,2</sup> Charlotte Colomb,<sup>1,2</sup> Olivier Negre,<sup>1,2,3,4</sup> Maria Denaro,<sup>3</sup> Béatrix Gillet-Legrand,<sup>1,2,4</sup> Joëlle Cheuzeville,<sup>1,2,4</sup> Anaïs Paulard,<sup>1,2,4</sup> Hélène Trebeden-Negre,<sup>5</sup> Suparek Borwornpinyo,<sup>6</sup> Karine Sii-Felice,<sup>1,2</sup> Leila Maouche,<sup>1,2,10</sup> Julian D. Down,<sup>7</sup> Phillippe Leboulch,<sup>1,2,8,9</sup> and Emmanuel Payen<sup>1,2,10</sup>

<sup>1</sup>CEA, Institute of Biology François Jacob, Fontenay aux Roses 92260, France; <sup>2</sup>UMR\_007, CEA and University of Paris Saclay, Fontenay aux Roses 92260, France; <sup>3</sup>bluebird bio, Inc., Cambridge, MA 02141, USA; <sup>4</sup>bluebird bio France, Fontenay aux Roses 92260, France; <sup>5</sup>Cell Therapy Unit, Pitie-Salpetriere Hospital, Paris 75013, France; <sup>6</sup>Mahidol University, Bangkok 10400, Thailand; <sup>7</sup>Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; <sup>8</sup>Ramathibodi Hospital, Bangkok 10400, Thailand; <sup>9</sup>Harvard Medical School and Genetics Division, Department of Medicine, Brigham & Women's Hospital, Boston, MA 02115, USA; <sup>10</sup>INSERM, Paris 75013, France

**Although gene transfer to hematopoietic stem cells (HSCs) has shown therapeutic efficacy in recent trials for several individuals with inherited disorders, transduction incompleteness of the HSC population remains a hurdle to yield a cure for all patients with reasonably low integrated vector numbers. In previous attempts at HSC selection, massive loss of transduced HSCs, contamination with non-transduced cells, or lack of applicability to large cell populations has rendered the procedures out of reach for human applications. Here, we fused codon-optimized puromycin N-acetyltransferase to herpes simplex virus thymidine kinase. When expressed from a ubiquitous promoter within a complex lentiviral vector comprising the  $\beta^{\text{AT87Q}}$ -globin gene, viral titers and therapeutic gene expression were maintained at effective levels. Complete selection and preservation of transduced HSCs were achieved after brief exposure to puromycin in the presence of MDR1 blocking agents, suggesting the procedure's suitability for human clinical applications while affording the additional safety of conditional suicide.**

## INTRODUCTION

$\beta$ -Globin gene disorders are the most prevalent inherited diseases worldwide. They result from mutations that affect  $\beta$ -globin production in patients with  $\beta$ -thalassemia and hemoglobin structure in patients with sickle cell disease (SCD). Recent clinical data have provided evidence that the delivery of the  $\beta$ -globin gene by self-inactivating lentiviral vectors (LVs) is of significant clinical benefit in patients with  $\beta^{\text{E}}/\beta^0$ -thalassemia major (TM)<sup>1-4</sup> and SCD.<sup>5</sup> However, the correction of anemia and the achievement of transfusion independence are more difficult to achieve in  $\beta$ -TM subjects with no residual expression of  $\beta$ -globin chains<sup>6,7</sup> and in SCD patients receiving limited numbers of corrected hematopoietic cells.<sup>8</sup> Furthermore, a wide distribution of anti-sickling hemoglobin among the erythroid cells

of patients with SCD is required for resolving symptoms.<sup>9</sup> Thus, by genetically correcting the vast majority of cells, it should be possible to further improve the clinical course of hemoglobin disorders.

Suboptimal phenotype correction upon hematopoietic gene and cell therapy can result from low levels of expression because of site-dependent variegation and/or limited hematopoietic stem cell (HSC) transduction. Position-dependent silencing can be limited by the use of chromatin insulators,<sup>10</sup> the *ex vivo* preselection of transduced HSCs,<sup>11,12</sup> and/or the integration of multiple copies of the vector into the cell genome.<sup>13</sup> Vector copy number (VCN) in HSCs can be increased by vector optimization,<sup>14</sup> the use of a high vector titer,<sup>15</sup> and improved manufacturing practices.<sup>16</sup> The promoters of LVs encoding  $\beta$ -globin are specific to erythroid cells, minimizing the risk of oncogene activation and cell transformation upon insertional mutagenesis. Although increasing the number of vectors integrating into the genome is an attractive approach for increasing the proportion of vector-bearing HSCs and the probability of the therapeutic gene being expressed, safety concerns remain because of the potential for gene disruption and aberrant splicing events.<sup>17,18</sup> We have documented clonal expansion caused by vector integration into the *HMG2* gene and aberrant splicing in one patient with  $\beta$ -thalassemia treated with lentiviral gene therapy.<sup>1</sup> We show here that the average VCN measured in transplanted cells as a pool can be misleading

Received 10 May 2017; accepted 18 October 2017;  
<https://doi.org/10.1016/j.ymthe.2017.10.015>.

<sup>11</sup>These authors contributed equally to this work.

**Correspondence:** Emmanuel Payen, CEA de Fontenay aux Roses, 18 route du Panorama, BP6, 92265 Fontenay aux Roses Cedex, France.

**E-mail:** [emmanuel.payen@cea.fr](mailto:emmanuel.payen@cea.fr)

**Correspondence:** Phillippe Leboulch, CEA de Fontenay aux Roses, 18 route du Panorama, BP6, 92265 Fontenay aux Roses Cedex, France.

**E-mail:** [philippe.leboulch@rics.bwh.harvard](mailto:philippe.leboulch@rics.bwh.harvard)



and hide disparities between hematopoietic cells with reconstituting activity, some of which are more transducible than others. Raising the mean VCN in HSCs may thus disproportionately increase the VCN in subpopulations of cells and raise the risk of oncogenic transformation without increasing the overall probability of transduced HSCs to the expected rate. The inclusion of a system to select genetically modified cells without also increasing the number of copies of the vector per cell may therefore be an appreciable advance to increasing both the efficacy and safety of current LVs.

Post-transduction cell selection can be performed upon fluorescence-activated<sup>12</sup> or magnetic<sup>19</sup> cell sorting. Surface cell molecules present the advantage of rapid cell sorting under good manufacturing practice, but the process is costly and a proportion of gene-modified cells are lost during the procedure. Drug selection strategies generally require a long-lasting selection time, which is undesirable because increasing culture time induces loss of engraftment ability and decreased clonal diversity.<sup>20,21</sup> Increasing time may also favor survival and engraftment of clones with vector insertion near oncogenes and increase the risk of genotoxicity.<sup>22</sup> Here, we investigated the use of brief puromycin exposure in our clinical setting that enables efficient production of a  $\beta$ -globin encoding LV, the transduction of HSCs over short periods of time, transplantability with a minimal loss of *in vivo* HSC activity, the expression of the  $\beta$ -globin gene to therapeutic levels in erythroid cells, and the absence of bias toward LV integration near oncogenes. We also combined this selection strategy with a conditional suicide gene to maximize the safety of the gene therapy product.

## RESULTS

### Optimal Dose and Timing for the Selection of Transduced Hematopoietic Progenitors

Vectors expressing the puromycin *N*-acetyltransferase (*PAC*) gene are derived from LVs that have been tested in approved human trials for the gene therapy of the  $\beta$ -hemoglobinopathies (HPV569 and BB305)<sup>1,2,5</sup> (Figure 1A). Bone marrow CD34<sup>+</sup> cells were transduced with LTGCPU1, incubated for 24 or 48 hr, and transferred for 24 hr in selective medium. The absolute numbers of vector-bearing erythroid progenitors (burst-forming unit-erythroids [BFU-Es]) were similar in the presence and absence of treatment, regardless of the time allowed for *PAC* gene expression (Figure 2A). However, incubation for 2 days was required for the myeloid (colony-forming unit granulocyte macrophages [CFU-GMs]) and long-term progenitor cells (long-term culture-initiating cells [LTC-ICs]) to acquire full resistance to the antibiotic (Figure 2A). We hypothesize that optimal expression level of the *PAC* gene may be reached earlier in erythroid progenitors than in the other cells tested.

Mobilized peripheral blood (mPB) CD34<sup>+</sup> cells were also transduced with LTGCPU1 and treated with puromycin. The mean VCN per cell was determined for pooled cells and compared with the theoretical VCN obtained in optimal conditions (eradication of all non-transduced cells and survival of all vector-bearing cells). In non-selective conditions, the mean VCN was 0.17. Based on the assumption that

vector integration obeys Poisson statistics (see below), we calculated that optimal selection should give rise to 1.08 copies of the vector per cell. Mean VCN was close to the expected value if puromycin treatment began at least 2 days after transduction (Figure 2B). Selection for 1 or 2 days resulted in similar VCN values, suggesting that treatment for 24 hr was sufficient to eliminate most of the non-transduced cells. With this optimized protocol, the proportion of vector-bearing progenitors reached more than 90% following treatment with 5  $\mu$ g/mL puromycin (Figure 2C). The absolute numbers of transduced progenitors were similar in the presence and absence of puromycin (Figure 2D), confirming that selection had a minimal toxic impact on vector-bearing progenitors with this procedure.

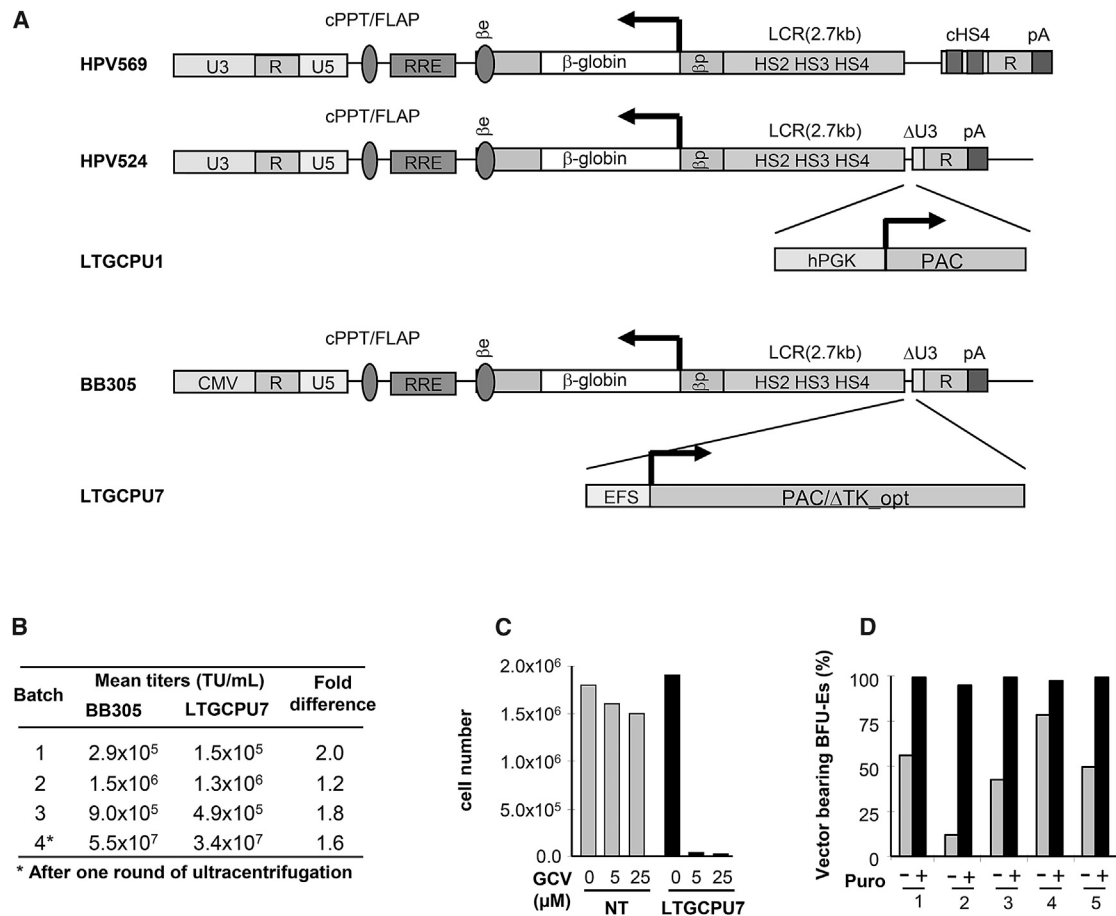
### Dual PAC and HSV1-TK Globin LV

In order to improve the safety and efficacy of the gene therapy vector, we fused a deleted version of the conditional herpes simplex virus type 1 thymidine kinase (*TK*) suicide gene to the *PAC* open reading frame (Figure 1A) and designed a sequence to optimize expression in human cells (*PAC*/ $\Delta$ TK<sub>opt</sub>; Figure S1). We replaced the human phosphoglycerate (*PGK*) promoter with the short intron-less version of the human elongation factor 1 alpha (*EF1A*) promoter (EFS), and the TAT-dependent U3 promoter/enhancer with the cytomegalovirus promoter (CMV), to decrease the size of the vector and increase its titer.<sup>14</sup> The resulting LTGCPU7 vector had a functional titer that ranged from 52% to 87% of that of the parental BB305 clinical LV (Figure 1B). Transduced NIH 3T3 cells were readily selectable with puromycin, and selected cells were sensitive to ganciclovir (Figure 1C). In addition, LTGCPU7-transduced cord blood (CB) progenitors were efficiently selected upon puromycin treatment (Figure 1D).

### Resistance of Repopulating Cells to Puromycin Selection

In order to evaluate selection at the stem cell level, we transduced CB CD34<sup>+</sup> cells with LTGCPU7, treated them with puromycin, studied them *in vitro*, and also injected them into NOD-*scid* IL2 $\gamma$ <sup>null</sup> (NSG) mice. The mean VCN (Figure 3A) and the percentage (Figure 3B) of vector-bearing cells were higher in the erythroid progenitors from puromycin-treated cells. Conversely, the mean VCN in human CD45<sup>+</sup> (hCD45<sup>+</sup>) cells isolated from immunodeficient mice receiving puromycin-treated cells was similar to that in the absence of treatment (Figure 3C), indicating an absence of selection at the SCID (severe combined immunodeficiency)-repopulating cell (SRC) level.

A study in a transfected cell line has provided evidence that puromycin is a substrate of the human multidrug resistance protein 1 (MDR1/ABCB1), but not of other multidrug resistance-related transporters such as breast cancer resistance protein (BCRP/ABCG2) or MDR-associated protein (MRP1/ABCC1).<sup>23</sup> Forced expression of *MDR1* to protect hematopoietic cells in cancer patients treated with MDR1 substrates such as anthracyclines and taxanes has led to transient selection of transduced cells,<sup>24,25</sup> suggesting that endogenous MDR1 is active in HSCs. Distinctively, colony-forming cells from *MDR1*-transduced hematopoietic human cells are protected from death.<sup>26</sup> We thus hypothesized that endogenous MDR1 activity is



**Figure 1. Lentiviral Vector Constructs Used in This Study and the Parental Vectors, Titters, and Function**

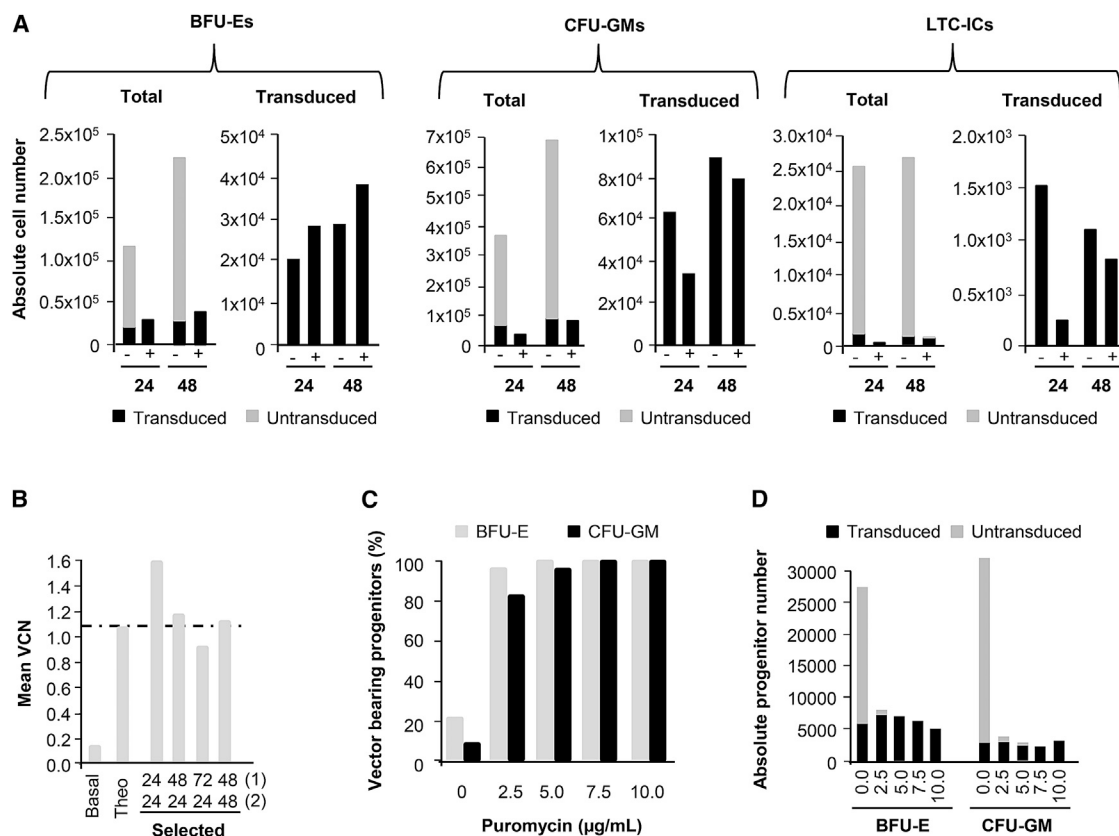
(A) All the LVs used encode the  $\beta^{\text{AT87Q}}$ -globin chain under the control of the human  $\beta$ -globin promoter ( $\beta$ p) and hypersensitive sites (HS) of the  $\beta$ -globin locus control region (LCR). The Tat-dependent HPV569, HPV524, and LTGCPU1 vectors contain a complete 5' long terminal repeat derived from HIV, whereas BB305 and LTGCPU7 contain a cytomegalovirus promoter and enhancer (CMV) instead of the HIV U3 region. The HPV524 and LTGCPU1 vector backbones are similar to the previously described  $\beta$ -globin lentiviral vector HPV569 except that they contain no chromatin insulators (cHS4). In LTGCPU1 and LTGCPU7, the human phosphoglycerate kinase 1 promoter (hPGK) or the short intron-less version of the human elongation factor 1 alpha promoter (EFS) controls expression of the puromycin *N*-acetyltransferase (PAC) gene. A deleted version of the herpes simplex virus type 1 thymidine kinase gene ( $\Delta$ TK) starting at the second ATG is fused to PAC in LTGCPU7 through a two-amino acid glycine-serine linker. We designed a sequence to optimize expression in human cells: PAC/ $\Delta$ TK\_opt (Figure S1). (B) Side-by-side comparison of BB305 and LTGCPU7 titers from four independent production experiments. (C) NIH 3T3 cells were mock-transduced (NT; gray bars) or transduced with LTGCPU7 (black bars), and LTGCPU7-transduced cells were selected with puromycin. Equivalent numbers of mock- and LTGCPU7-transduced/selected cells were treated with ganciclovir (GCV) at the indicated concentration. Living cells were counted 3 days later. (D) Five different cord blood CD34<sup>+</sup> cell samples were transduced with LTGCPU7, left untreated (–), or treated (+) 2 days post-transduction with 5  $\mu$ g/mL puromycin for 24 hr and plated on semi-solid medium. The percentage of erythroid colonies carrying the vector is indicated. 50 colonies per condition were isolated for vector detection.

higher in HSCs than in more mature cells and precludes puromycin selection at the stem cell level.

LTGCPU7-transduced CD34<sup>+</sup> cells from five CB donors were analyzed for the presence of MDR1. The percentage of MDR1<sup>+</sup> cells was low 1 day after thawing (at the time of transduction), but had increased by days 3 and 4 (Figure 3D; Figure S2), and was higher for the more primitive CD34<sup>+</sup>CD133<sup>+</sup> subset<sup>27</sup> than for CD34<sup>+</sup>CD133<sup>–</sup> cells. We concluded that day 3–4 CD34<sup>+</sup> cells were heterogeneous for MDR1 expression, and that cells with the highest

SRC activity (CD34<sup>+</sup>CD133<sup>+</sup>) had higher levels of MDR1 than other CD34<sup>+</sup> cell subsets.

Nevertheless, despite detection of MDR1 antigen in 40%–50% of CD34<sup>+</sup> hematopoietic cells (Figure S2), addition of MDR1 inhibitors was unnecessary to achieve full puromycin selection of transduced hematopoietic CD34<sup>+</sup> progenitor cells (Figures 2A–2D and 3A). We therefore hypothesized that MDR1 activity in hematopoietic progenitors was insufficient to remove puromycin at a dose of 5  $\mu$ g/mL. In order to gain insight into the possible role of MDR1 in the resistance



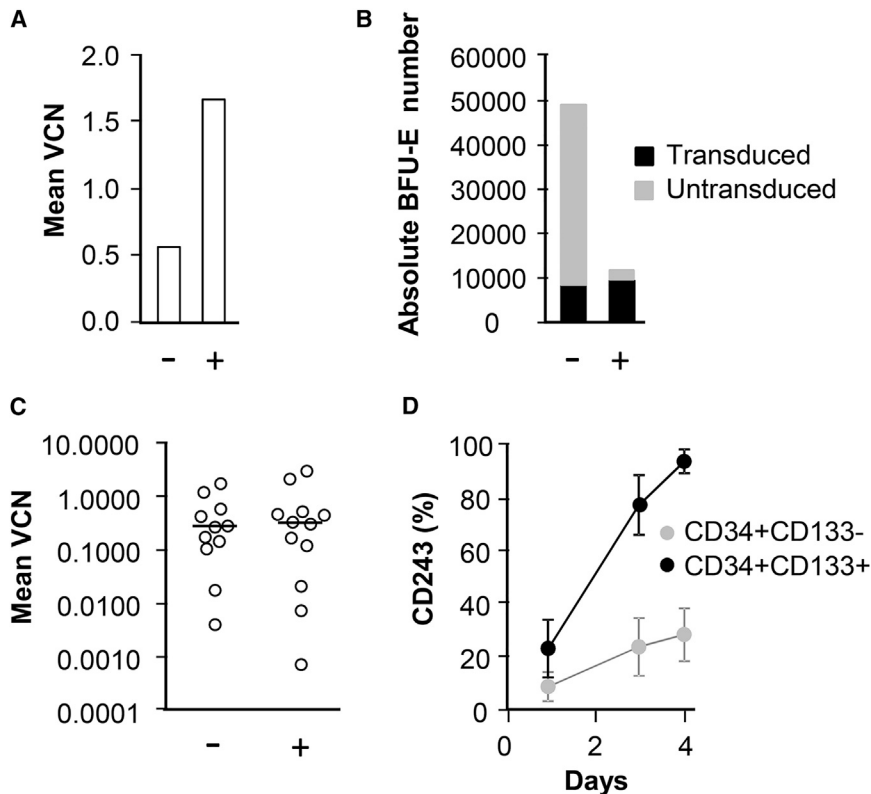
**Figure 2. Efficient Selection and Recovery of Transduced Progenitor Cells**

(A) Adult bone marrow CD34<sup>+</sup> cells were transduced with LTGCPU1. 24 or 48 hr later, cells were left untreated (–) or treated (+) with puromycin (10 µg/mL) for a further 24 hr. The cells were then immediately plated on semi-solid medium or grown in liquid culture for the evaluation of erythroid (BFU-Es) and myeloid (CFU-GM) progenitor cell numbers and to determine the number of long-term culture-initiating cells (LTC-ICs). At the end of the culture period, colony-forming cells were counted and picked (50 colonies per condition) for qPCR analyses of the presence of the vector. The absolute numbers of transduced (black) and non-transduced (gray) progenitor cells are indicated. (B) Mobilized bone marrow cells were transduced with LTGCPU1. 24, 48, or 72 hr later (1), cells were selected on puromycin (10 µg/mL) for 24 or 48 hr (2). The mean VCN for non-selected (basal) and selected cells was determined for pooled cells and compared with the theoretical expected VCN in optimal selection conditions (Theo). (C and D) Mobilized bone marrow cells were transduced and selected 48 hr later with several concentrations (0–10 µg/mL) of puromycin during 24 hr. The percentages (C) and absolute numbers (D) of transduced and non-transduced erythroid and myeloid progenitors are indicated. 25–40 colonies per condition were isolated for vector detection.

of hematopoietic cells to puromycin, we sorted CD34<sup>+</sup>CD243<sup>–</sup> and CD34<sup>+</sup>CD243<sup>+</sup> transduced cells, treated them with lower concentrations of puromycin in the absence or presence of MDR1 inhibitor, grew them *in vitro*, and compared VCNs (Figure S3). The experiment was complicated by the fact that MDR1 is distributed as a continuous rather than a discrete gradient at the surface of CD34<sup>+</sup> cells, resulting in overlapping sorted cell fractions. Nevertheless, at low puromycin concentrations (1 and 0.3 µg/mL), the effect of the MDR1 inhibitor (cyclosporine A) on the selection level was significant on MDR1<sup>+</sup> progenitors, but not on MDR1<sup>–</sup> progenitor cells. In the presence of inefficient amounts of puromycin (0.1 µg/mL), cyclosporine A had no significant impact on VCNs. At the highest puromycin concentration (2 µg/mL), selection of MDR1<sup>–</sup> and MDR1<sup>+</sup> transduced cells was independent of cyclosporine A. We therefore concluded that endogenous MDR1 activity was, at least partially, responsible for inefficient HSC selection. Expression and/or activity of MDR1 may be higher in HSCs than in more mature committed progenitor cells.

#### MDR1 Inhibitors Release the Selection Inhibition of Immature Cells

LTGCPU7 transduced CB CD34<sup>+</sup> cells were treated with puromycin, in the presence or absence of MDR1 inhibitors. Mean VCN was determined in pooled cells from methylcellulose and in human cells recovered from mouse bone marrow. In this experiment, selection level in erythroid progenitors was unexpectedly lower in the absence than in the presence of MDR1 inhibitors (Figure 4A). We hypothesize that CB progenitors with variable levels of MDR1 expression may have variable susceptibility to puromycin selection, depending on the sample considered. In the absence of MDR1 inhibitor, no selection of transduced SRCs was observed (Figure 4B). By contrast, in the presence of inhibitors, puromycin was able to select vector-bearing cells (Figure 4B). As expected, the selection of transduced SRCs and their concomitant decreased number was associated with lower proportions of human cells in mice receiving selected HSCs than in mice receiving cells treated with puromycin only (Figure 4C). The lower



**Figure 3. Inefficient Selection of SRCs Is Correlated with High-Level MDR1 in CD34<sup>+</sup>CD133<sup>+</sup> Cells**

(A–C) Cord blood CD34<sup>+</sup> cells were transduced with LTGCPU7, left untreated (–), or treated (+) 2 days post-transduction with 5 μg/mL puromycin and plated on semi-solid medium or injected into NSG mice. (A and B) The mean VCN in pooled erythroid cells (A) and the absolute counts of transduced and untransduced progenitor cells (B) were determined for erythroid colonies retrieved from methylcellulose (60 colonies per condition were isolated for vector detection). Mice received 30,000 cells (five mice in the untreated group and six mice in the treated group) or 150,000 cells (six mice in each group). (C) The mean VCN values and their median were determined for human CD45<sup>+</sup> cells sorted from the bone marrow of individual mice 3 months after transplantation. (D) Five different cord blood CD34<sup>+</sup> cell samples were transduced with LTGCPU7 and analyzed for the presence of CD243 (MDR1), CD34, and CD133 antigens on the day of transduction (day 1) and 2 and 3 days later. Results are the mean ± SD.

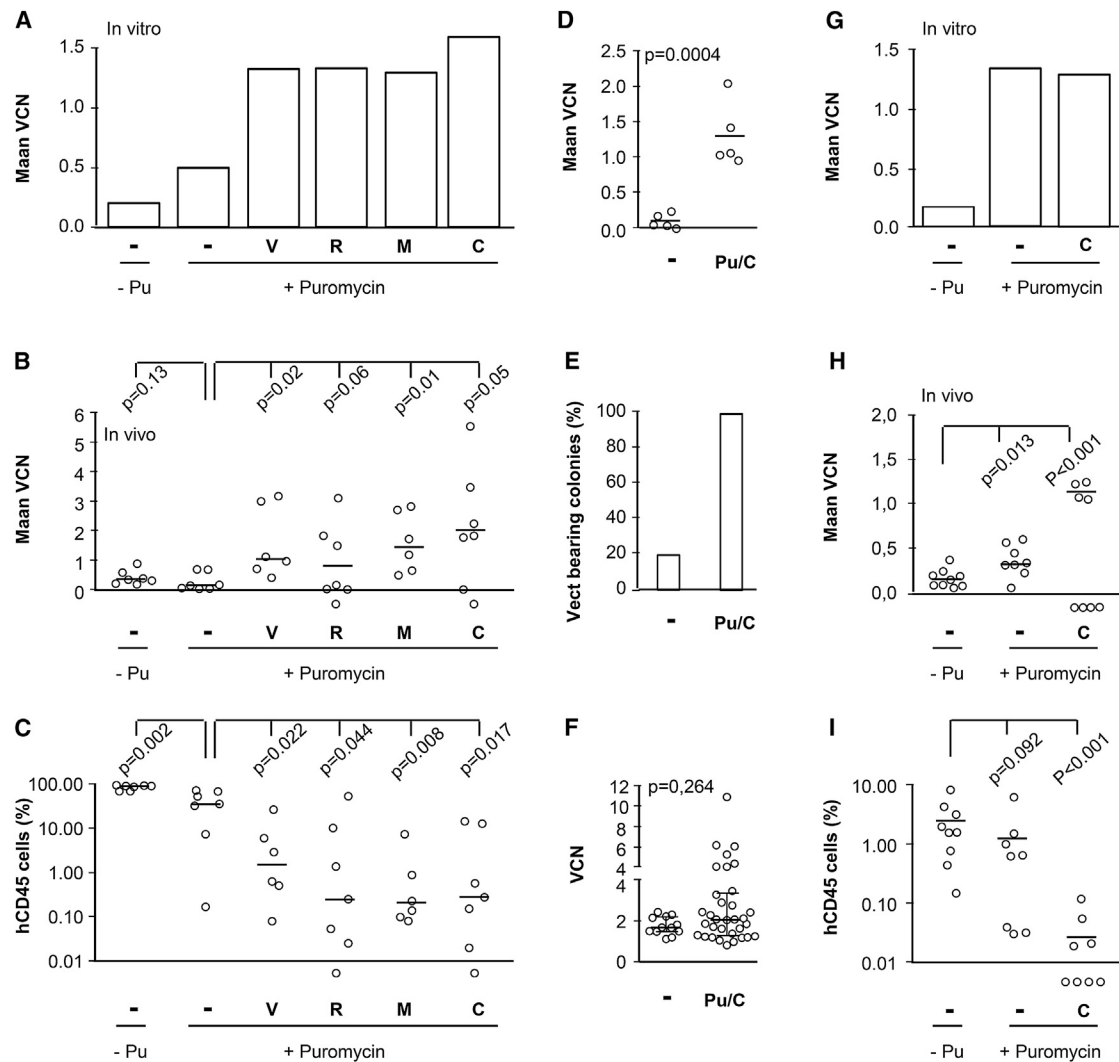
rate of human cell reconstitution in the group of mice receiving cells treated with puromycin only than in animals receiving untreated cells, in the absence of selection at the SRC level, may be explained by the influence of supporting cells on SRC homing, proliferation, and/or differentiation.

We checked that the higher VCN in human cells resulted from a higher proportion of modified SRCs, rather than the selection of a small number of highly modified cells, by determining the proportion of vector-bearing SRCs. LTGCPU7 transduced CB CD34<sup>+</sup> cells were selected and injected into NSG mice. The mean VCN in the hCD45<sup>+</sup> cells of individual mice increased upon cell selection with puromycin and cyclosporine A (Figure 4D). Erythroid progenitors retrieved from seven mice, including three from the non-selected group and four from the selected group, were recovered in the form of erythroid colonies and harvested. The vector was detected in 19% of cells from mice in the non-selected group and 100% of cells from mice in the selected group (Figure 4E), indicating that vector-bearing SRCs, from which bone marrow erythroid progenitors are derived, had been selected. VCN did not differ significantly between individual BFU-Es, suggesting that the selection of transduced cells did not particularly favor the cells with the highest VCNs (Figure 4F). However, a few colonies had high VCNs, and VCN values did not follow a normal distribution ( $p < 0.05$ , Shapiro-Wilk test), suggesting that a few cells may have been transduced more efficiently than others (see below).

mice injected with puromycin-treated SRCs than in other animals (Figure 4I).

#### Recovery of Transduced Stem Cells

By evaluating the absolute numbers of these cells after selection in a limiting dilution assay, we investigated whether the lower percentage of human CD45<sup>+</sup> cells in NSG mice resulted purely from the elimination of untransduced SRCs or whether it was also attributed to a particular toxicity to transduced SRCs (Figure 5). The frequency of SRCs was 4.4 times higher for control than for treated cells (Figure 5A), and cell selection decreased the absolute cell count by a factor of 5.4. The median transduction efficacies for the erythroid progeny of these cells were 74.9% for selected cells and 4.2% for control cells (Figure 5B). Therefore, the absolute number of vector-bearing SRCs on day 4 was only 1.34 times higher for untreated than for treated cells (Figure 5C). Reconstitution levels as a function of the numbers of SRCs injected showed a similar relationship regardless of selection, consistent with a lack of dependence of culture conditions on cell fitness (Figure 5D). Another experiment yielded similar conclusions (Figure S4). In order to ascertain that MDR1 inhibition has a limited negative impact on repopulating cells, CB hCD34<sup>+</sup> cells were treated with verapamil or cyclosporine A alone and transplanted in NSG mice. The level of engraftment, the proportion of human CD34<sup>+</sup> cells, and the relative proportions of myeloid and lymphoid cells were similar compared to those measured in the control group



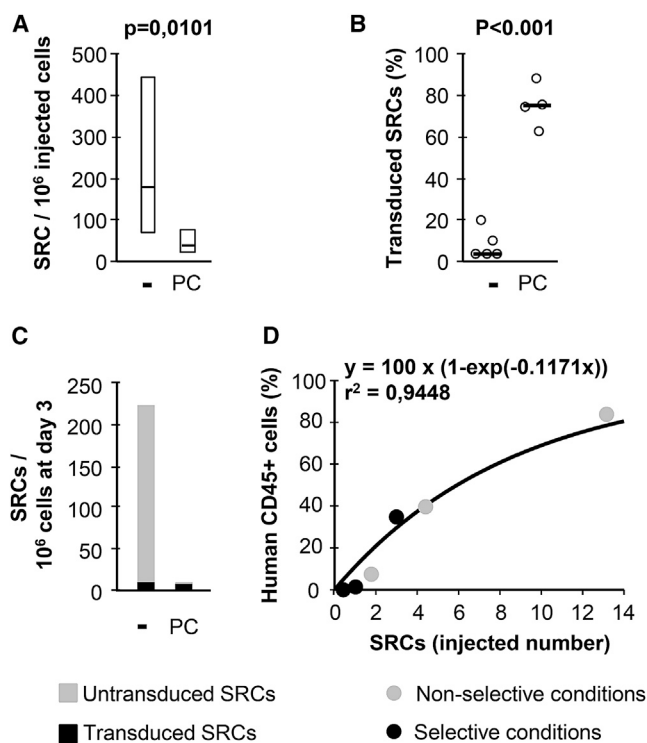
**Figure 4. Selection of Transduced SRCs in the Presence of MDR1 Inhibitors**

(A–C) Human cord blood CD34<sup>+</sup> cells were transduced with LTGCPU7 and left untreated (–Pu) or treated 2 days later with 5 μg/mL puromycin, in the absence (–) or presence of MDR inhibitors (verapamil [V, 20 μM], reserpine [R, 20 μM], mifepristone [M, 5 μM], or cyclosporine A [C, 2 μM]). Cells were plated on semi-solid medium or injected into NSG mice (65,000 cells per mouse in the treated groups and 225,000 cells per mouse in the non-selected group, corresponding to identical numbers of input cells before selection). (A and B) Mean VCN was determined for pooled cells from *in vitro* culture (A) and from human CD45<sup>+</sup> cells sorted from the bone marrow of individual mice 3 months after transplantation (B). (C) The human chimerism in the bone marrow of individual mice was also assessed by flow cytometry. The black bars indicate the median value for mean VCN and the level of human chimerism. Levels of chimerism <0.01% were below the limit of quantification. Mice with such low levels of chimerism were not included in the calculation of median VCN in (B). (D–F) The experiment was repeated with puromycin and cyclosporine A (2 μM). Mean VCN was determined in human CD45<sup>+</sup> cells (D) and in human erythroid progenitors retrieved from individual mice (E and F). (E and F) The proportion of vector-bearing progenitors (E) and the VCN in individual colonies (F) from NSG mice are indicated. The vector was detected in 12 of 63 colonies in the non-selected group and 33 of 33 colonies in the selected group ( $p = 10^{-16}$  in Fisher’s exact test). The black bars indicate the median and interquartile values. (G–I) Mobilized peripheral blood CD34<sup>+</sup> cells were transduced with LTGCPU7 and were left untreated (–Pu) or treated 2 days later with 5 μg/mL puromycin, in the presence (C) or absence (–) of cyclosporine A (2 μM). Cells were plated on semi-solid medium, and 109,000 (untreated), 98,000 (puromycin), or 68,000 (puromycin + cyclosporin A) cells were injected into NSG mice. (G and H) The mean VCN was determined for pooled cells from *in vitro* culture (G) and for human CD45<sup>+</sup> cells sorted from bone marrow from individual mice 4 months after transplantation (H). (I) The human chimerism in bone marrow from individual mice was also assessed by flow cytometry. Black bars indicate the median value of mean VCN and human chimerism. Levels of chimerism below 0.01% are below the limit of quantification. The mice concerned were not included in the calculation of median VCN in (H).

(Figure S5). The lower percentage of human cells in mice receiving selected cells therefore resulted primarily from the elimination of non-transduced SRCs.

**Transduced Hematopoietic Cells Are Sensitive to Ganciclovir**

CB CD34<sup>+</sup> cells were transduced with LTGCPU7, selected with puromycin, and treated with ganciclovir (Figures 6A and 6B). The number



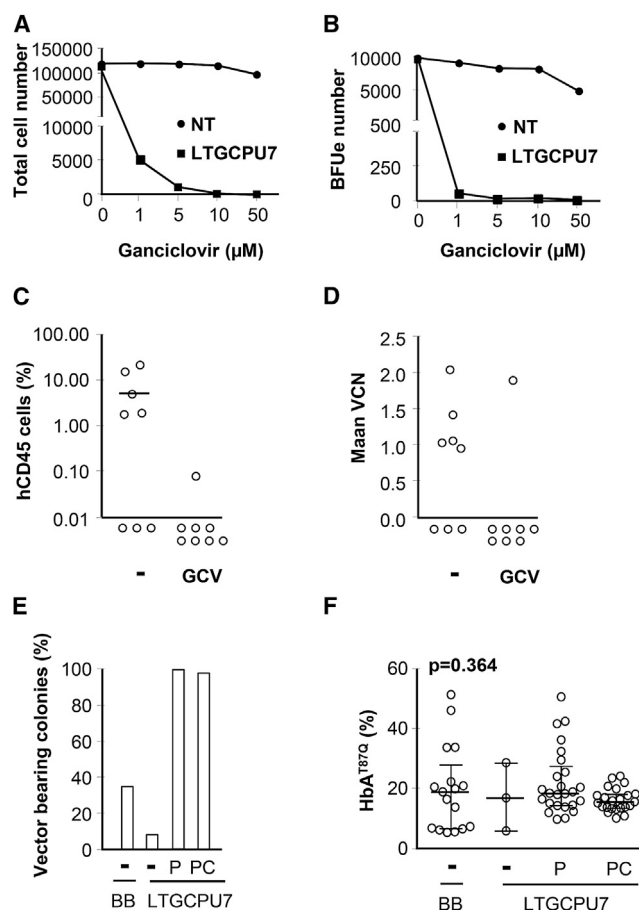
**Figure 5. Recovery of Transduced and Selected SRCs**

Human cord blood CD34<sup>+</sup> cells were transduced with LTGCPU7 and either left untreated (–) or treated with puromycin + cyclosporine A (PC). On day 4, six groups of mice received 10,000, 25,000, and 75,000 treated or untreated cells. (A) Three months later, the proportion of hCD45<sup>+</sup> cells in the bone marrow was evaluated by flow cytometry, to determine the frequency of SRCs. (B) The proportion of transduced SRCs is based on the proportion of their erythroid progeny (24 colonies per mouse) carrying the vector. (C and D) These data were used to calculate the relative numbers of transduced and non-transduced repopulating cells (SRCs) (C) and the reconstitution capacity of SRCs (median proportion of human cells) in NSG mice (D). SRC is here defined as a cell giving >1% hCD45<sup>+</sup> cells among all CD45<sup>+</sup> cells, 3 months after human cell infusion.

of transduced erythroid progenitors decreased by a factor of 200 in the presence of 1  $\mu$ M ganciclovir. The transduced cells were injected into 16 NSG mice. Four weeks later, eight of the mice received ganciclovir. Two months later, hCD45<sup>+</sup> cells were detected in five of the eight mice in the control group and in one of the eight mice of the treated group, although the number of human cells was very small in this mouse (Figure 6C). The absence of hCD45<sup>+</sup> cells in mice with levels below the limit of sensitivity (<0.01%) was confirmed by the lack of enrichment observed on hCD45<sup>+</sup> cell sorting (Figure S6). The vector was detected at a similar copy number in ganciclovir-resistant cells than in cells from mice that did not receive ganciclovir (Figure 6D), indicating that ganciclovir treatment eliminated most, but not all, transduced cells.

#### $\beta$ -Globin Gene Is Expressed at the Expected Level

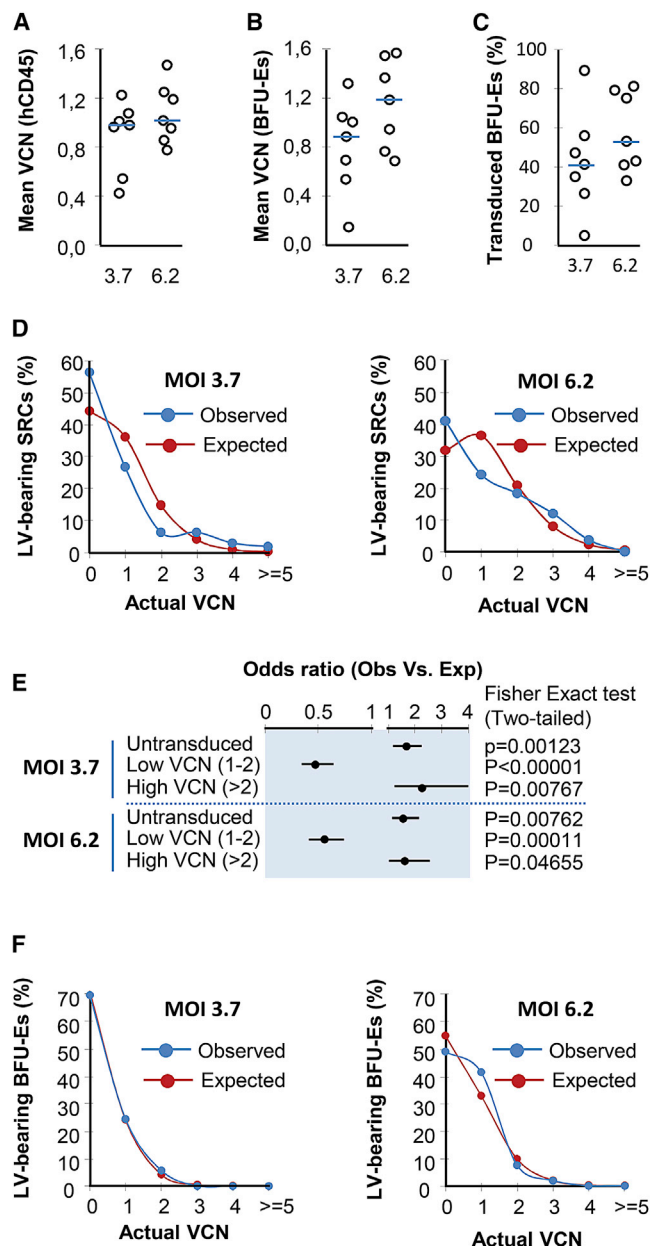
Because transcriptional interference or competition between the two promoters may be suppressive,<sup>28</sup> the  $\beta$ -globin expression level was



**Figure 6. Conditional Suicide and Expression of the  $\beta$ -globin Therapeutic Gene**

(A–D) Human cord blood CD34<sup>+</sup> cells were transduced with LTGCPU7 and treated 2 days later with puromycin and cyclosporine A for 24 hr. (A and B) Non-transduced (NT) and LTGCPU7-transduced cells were then treated with ganciclovir (GCV) for 3 days, counted (A), and plated on semi-solid medium for the evaluation of erythroid progenitor cell eradication (B). Selected cells were also injected into 16 NSG mice, half of which were treated with GCV 4 weeks later (50 mg/kg, 6 days per week, for 2 weeks). (C and D) Human chimerism (C) and VCN in human CD45<sup>+</sup> cells (D) were determined 2 months later. Levels of chimerism <0.01% were below the limit of quantification. (E and F) Human CD34<sup>+</sup> cells from a patient with sickle cell anemia (producing sickle hemoglobin HbS) were transduced with BB305 or LTGCPU7. LTGCPU7-transduced cells were left untreated (–), were selected with puromycin (P), or selected with puromycin + cyclosporine A (PC). Transduced cells were plated on semi-solid medium. The presence of the integrated vector was assessed in 48 individual erythroid colonies per condition (E), and the amount of transgenic hemoglobin (HbA<sup>T87Q</sup>/HbA<sup>T87Q</sup>+HbS) was then compared between vector-carrying BFU-Es (F). The black bars show the median and interquartile values.

compared with that of the parental vector. CD34<sup>+</sup> cells from an individual with SCD were transduced with BB305 and LTGCPU7. Transduction occurred in 35% and 10% of the erythroid progenitor cells, respectively (Figure 6E), so that VCNs are expected to be no more than one in about 80% and 95% of transduced progenitors for BB305 and LTGCPU7 vectors, respectively. LTGCPU7-transduced



**Figure 7. Lentiviral Transduction and Comparison with a Random Distribution**

onies per MOI and compared with the expected values based on a Poisson distribution. The bars shown in (A)–(C) represent the median values. Dots and bars in (E) indicate the odds ratio and 95% confidence intervals, respectively.

progenitor cells were efficiently selected (Figure 6E), and similar amounts of therapeutic  $\beta^{\text{AT87Q}}$ -globin protein were produced with the two vectors (Figure 6F), suggesting an absence of transcriptional interference or competition between the two promoters.

### Efficient Transduction of Hematopoietic Cells upon Vector Purification

In the above experiments, vectors were produced and concentrated by ultracentrifugation. LTGCPU7 vector titers were between  $5 \times 10^7$  and  $2 \times 10^8$  transducing units per milliliter. All transduction protocols were carried out with 2 million CD34<sup>+</sup> cells per milliliter and a 10% vector preparation (v/v), giving an MOI of 2.5–10. The LTGCPU7 vector was purified to increase transduction efficacy. In this case a batch of LTGCPU7 vector was made, and the supernatant was split in two. One half was ultracentrifuged, whereas the other half was subjected to purification by ion exchange chromatography before ultracentrifugation. The titers were determined and the two vector preparations were used to transduce CB CD34<sup>+</sup> cells. The average VCN did not exceed  $\approx 0.5$  with the crude extract, whereas it reached values of  $\approx 1$  with the purified vector used at a lower MOI (Figure S7A). At the highest MOIs used, some of the cells were cultured with or without puromycin for an additional day and plated in methylcellulose. The percentages of transduced erythroid progenitors were 18% for the crude vector and 56% for the purified vector, consistent with the differences in mean VCN observed (Figure S7B).

CB CD34<sup>+</sup> cells were transduced at MOIs of 3.7 and 6.2 with the purified LTGCPU7 preparation, and transduced cells were subsequently infused into NSG mice. The mean VCNs measured in hCD45<sup>+</sup> cells from individual mice (Figure 7A) or in the erythroid progeny of SRCs (Figure 7B) were higher than those obtained in experiments performed with crude extracts at similar MOIs and in the absence of selection (Figures 3C and 4B; Figure S4C). Transduction efficiency in the SRC progeny reached 43.4% and 58.5% (Figure 7C) and was higher than observed with the non-purified vector (Figure 5B).

### Distribution of VCN Is Variable between Cell Subtypes

If all cells were equally susceptible to transduction, the distribution of vector integration into individual cells should follow Poisson statistics. The distribution and the percentage of transduced cells may then be deduced from the mean VCN of the cell concerned (Figure S8). We compared, in the absence of selection, the distribution of vector integration into individual cells with the expected values calculated from the mean VCN in the populations concerned, based on a Poisson distribution of single events. In erythroid progenitors derived from transduced CD34<sup>+</sup> cells, the distribution of vector integration events did not differ significantly from expectations (Figure 7F). Details are given in Figure S9 (experiment 1). Conversely, in the progeny of SRCs, the observed distribution of vector integration events differed significantly from that predicted on the basis of



**Table 1. Integration Site Frequency in the Vicinity of Genes and Oncogenes**

	Unique Sites	% Sites within 50 kb TSS	% Sites within 50 kb TSS Onco	% Sites within 100 kb TSS	Sites within 100 kb TSS Onco	Sites within 50 kb TSS, % within 50 kb TSS Onco	Sites within 100 kb TSS, % within 100 kb TSS Onco	% Sites where Nearest TSS Is TSS Onco
MRC	3,000	47.7	8.0	64.8	13.9	16.8	21.5	10.3
	NS	94	73.4***	19.1**	85.1***	27.7**	26.1	14.9
<i>In Vitro</i>	Sel	227	78.4***	16.3***	93.0***	35.7***	20.8	38.4***
	p value <sup>b</sup>		0.382	0.503	0.035 <sup>a</sup>	0.194	0.396	0.414
	NS	147	84.4***	17.7***	93.9***	29.9***	21.0	31.9*
<i>In Vivo</i>	Sel	196	79.1***	19.9***	90.8***	32.1***	25.2*	35.4***
	p value <sup>b</sup>		0.263	0.677	0.320	0.724	0.477	0.748
	p value <sup>c</sup>		0.047 <sup>a</sup>	0.864	0.041 <sup>a</sup>	0.772	0.475	0.758
	p value <sup>d</sup>		0.906	0.375	0.642	0.472	0.362	0.378

MRC, matched random control; NS, non-selected cells; onco, oncogene; Sel, selected cells. \* $p < 0.05$ , \*\* $p < 0.005$ , and \*\*\* $p < 0.0005$ , respectively, by Fisher's exact test (two-tailed) comparing test samples with MRC dataset.

<sup>a</sup> $p < 0.05$ .

<sup>b</sup>Fisher's exact test (two-tailed) comparing samples from non-selected and selected groups.

<sup>c</sup>Fisher's exact test (two-tailed) comparing *in vitro* and *in vivo* dataset of non-selected samples.

<sup>d</sup>Fisher's exact test (two-tailed) comparing *in vitro* and *in vivo* dataset of selected samples.

random distribution (Figures 7D and 7E; Figure S10), suggesting that SRCs with different susceptibility to transduction coexist, and that neither percentage of transduced cells nor vector distribution can be straightforwardly and accurately deduced from mean VCN measured in the progeny of repopulating cells.

We also compared the distribution of vector integration into individual cells after puromycin selection with values calculated in conditions in which all vector-bearing cells survive and all untransduced cells die. The proportion of cells with a VCN of 1 was slightly lower than expected on the basis of the vector distributions obtained before selection, in both erythroid progenitors derived from selected CD34<sup>+</sup> cells and in the progeny of selected SRCs (experiments 1 and 3, respectively; Figure S11). A small number of cells with insufficiently strong PAC expression may therefore have been eliminated during selection on puromycin. Accordingly, the mean VCN in the erythroid progeny of SRCs was higher ( $p < 0.001$ ) in selected cells ( $4.8 \pm 2.9$ ) than in non-selected SRCs ( $3.1 \pm 1.4$ ).

#### Preferred Integration Sites after Puromycin Selection

Cells in which the vector integrates into genomic features associated with high levels of gene expression may be favored during *in vitro* selection, with further enrichment occurring *in vivo* through insertional mutagenesis. We investigated this possibility by determining whether vector integration sites post-selection and post-transplantation presented any evidence of such enrichment close to genes, particularly those associated with cell growth, or chromatin features associated with high levels of gene expression. In this case, we analyzed human cells harvested from the bone marrow of NSG mice receiving selected and non-selected transduced cells (experiment 3 described in Figures S10 and S11). The mean VCN in hCD45<sup>+</sup> cells from mouse bone marrow was  $4.6 \pm 3.0$  in selected cells and  $2.5 \pm 1.4$  in non-selected cells. Before transplantation, aliquots of cells were cultured on meth-

ylcellulose. The mean VCN was  $1.68 \pm 0.01$  for selected cells and  $0.87 \pm 0.01$  for non-selected cells.

As expected, the proportions and distributions of insertion sites were significantly different from those obtained for a matched random control dataset (Figures S12 and S13; Table S1). No significant difference between non-selected and selected *in vivo* samples was observed for any of the features investigated including the distance of insertion sites to oncogenes (Table 1). We can, therefore, conclude that drug-mediated selection had not biased the distribution of insertion sites beyond the possible bias due to *in vivo* expansion. We evaluated the potential distribution distortions specifically due to *in vitro* drug selection or *in vivo* expansion through the following comparisons: (1) non-selected versus selected samples *in vitro*, to evaluate the effect of drug selection; (2) non-selected *in vitro* versus non-selected *in vivo* samples, to evaluate the role of cell expansion; and (3) *in vitro* selected versus *in vivo* selected samples, to determine whether any insertion biases observed were additive (Figure S14). Both drug selection and *in vivo* expansion slightly distorted insertion site distributions toward regions with a high transcription start site (TSS) density, regions with a higher proportion of CpG islands, and regions with a high density of DNase hypersensitive sites. Although it is surprising to find that the selection of cells with insertion sites within regions of high gene density was not associated with higher transcript abundance, this observation is consistent with the higher frequency of insertion sites upstream rather than downstream of the TSS, with these insertion sites thus being located in non-transcribed regions.

#### DISCUSSION

We describe here the properties of a relevant  $\beta$ -globin LV capable of transducing a high proportion of hematopoietic cells with a limited number of insertion hits.

We decided to use the EFS promoter to control PAC/ $\Delta$ TK expression because it is relatively short (230 bp), is potent enough to express clinically relevant genes,<sup>29</sup> and is associated with a low transformation potential.<sup>30</sup> The LTGCPU7 LV was produced almost as efficiently as the parental BB305 (<2-fold lower), and the vector expressed both operational PAC and functional TK genes. In our culture conditions, which are identical to those used in the clinical setting,<sup>31</sup> MDR1 levels were correlated with the characteristic phenotype of stem cells and increased over time, peaking at the time of selection. In the presence of MDR1 inhibitors,<sup>32</sup> we show here that the MDR1 substrate puromycin<sup>23</sup> is effective to select vector-bearing CB and adult mPB SRCs. Cyclosporine A has been shown to relieve lentiviral restriction blocks in hematopoietic cells, but at a higher concentration than used here.<sup>33</sup> Furthermore, none of the inhibitors tested here in the absence of puromycin was capable of altering the transduction rate at the concentrations indicated.

As expected, stem cell selection resulted in smaller numbers of human cells being recovered from NSG mice receiving treated SRCs than from animals receiving untreated SRCs. We investigated whether the smaller number of human cells recovered was due to an excessive toxic effect of the selective molecule cocktail, on both vector-bearing and untransduced cells, by determining the number of stem cells transduced. With selection, a minimal decrease in the number of transduced cells was observed following treatment, whereas most of the non-transduced cells were effectively removed. The slightly smaller number of transduced SRCs ( $\approx$ 25%) may be caused by MDR1 inhibition or the residual toxicity of cyclosporine A. However, although this calcineurin inhibitor has been shown to be cytotoxic to some primary cells, these concentrations were much higher than that used here.<sup>34,35</sup> Furthermore, the hematopoietic reconstitution potential of *Mdr1*<sup>-/-</sup> HSCs in mice is not affected.<sup>36</sup> The proportion of human cells after transplantation in NSG mice was correlated with the number of cells injected into the mouse and independent of treatment conditions, indicating that there was no effect on SRC fitness. Alternatively, the integration of the vector into regions of heterochromatin may have driven the production of a limited amount of the resistance protein in a minority of transduced cells, leading to the observed cell loss. Our observation that the proportion of cells with a low VCN is slightly lower than expected after cell selection and the preferential distribution of insertion sites upstream of the TSS in gene-dense regions is consistent with this notion. If this should prove to be the case, then this procedure would also select the cells with the highest probability of expressing their transgenes. The pretransplantation selection of transduced mouse cells by a fluorescence-based method has been shown to obviate the subsequent consequences of gene silencing.<sup>12</sup> This effect would have the advantage of destroying cells harboring vectors, but not expressing the therapeutic gene,<sup>12</sup> thereby increasing treatment efficacy. Importantly, drug-mediated selection had not biased the distribution of insertion sites beyond the possible bias due to *in vivo* expansion, and the distribution biases were not cumulative.

The ultracentrifugation of non-purified LVs increases the concentration of cellular debris, membrane fragments, and denatured proteins,

which are toxic and decrease the efficiency of target-cell transduction.<sup>37</sup> We therefore purified the vector by ion exchange chromatography. This made it possible to transduce hematopoietic cells without the plateau in transduction rate that is rapidly reached when hematopoietic cells are transduced with crude extracts.<sup>38,39</sup> In this setting, the LTGCPU7 vector preparation yielded a transduction efficacy close to the approximately 50% required here to preclude a high frequency of cells having a high VCN and to minimize the excessive removal of cells. Lentiviral transduction rate and the distribution of the vector in late progenitors were consistent with the expected values calculated from the mean copy number in the bulk population of CD34<sup>+</sup> cells based on Poisson statistics, as suggested in a previous study.<sup>40</sup> The significant departure from an ideal Poisson distribution observed in the progeny of SRCs, however, suggests that transduction efficiency is not equal in cells at the most immature stages, resulting in a large number of vector copies in the subpopulation most susceptible to transduction. These observations indicate that the initial transduction rate for individual cells or colonies should be carefully monitored so as to assess the generation of subsets of cells with a high VCN. We show here that adjusting the initial transduction rate to  $\approx$ 50% and/or the mean VCN to  $\approx$ 1 limits the frequency of SRC progenies with  $\geq$ 5 LV copies to about 1%. Conversely, the effort to transduce a higher proportion of HSCs without selection by increased LV delivery with elevation in the average VCN may enhance the risk of oncogenesis from multiple integrants per cell.

We included a suicide gene to decrease the risks related to the potential genotoxicity of integrative vectors and to make it possible to ablate cells with modified genes if a serious adverse event occurred. We investigated the use of the TK gene because this strategy has been used in many clinical studies.<sup>41</sup> We incorporated  $\Delta$ TK into our LV because its smaller size was advantageous. Transduced cells were sensitive to ganciclovir treatment, with a 1  $\mu$ M solution decreasing the number of hematopoietic progenitors by two orders of magnitude. This level of sensitivity is similar to that seen in transduced cell lines or primary T cells expressing  $\Delta$ TK,<sup>42</sup> or fusion proteins including TK or its hypersensitive mutant form, TKSR39.<sup>43,44</sup> Transduced SRCs were almost completely eradicated, a result that compares favorably with those of other *in vivo* studies.<sup>45</sup> We were nevertheless able to detect a few transduced human cells. The presence of these cells may reflect insufficient drug sensitivity, epigenetic silencing of the transgene, or mutations of the TK gene. Nevertheless, a large majority of the transduced cells remained effectively eradicated. Thus, although non-obligatory, the presence of  $\Delta$ TK provides a higher level of safety than the use of a vector without a suicide gene.

A potential obstacle to the persistence of modified hematopoietic cells *in vivo* is the development of host immune responses to non-human proteins. Viral TK has been reported to be immunogenic in adoptively transferred modified donor T cells, although conflicting data reflecting differences in the immune status of treated patients have been published.<sup>46,47</sup> Preserving recipient immunity may therefore render PAC/ $\Delta$ TK-expressing cells susceptible to an immune response. Future studies will need to evaluate the induction of systemic

tolerance to the PAC/ $\Delta$ TK gene product in the setting of temporary mild immune suppression with agents such as anti-thymocyte globulin,<sup>48</sup> rapamycin,<sup>49–51</sup> or other non-genotoxic molecules and before the induction of central immune tolerance.<sup>52</sup> Non-immunogenic genes such as NGFR<sup>19</sup> and inducible caspase-9<sup>53</sup> may also be considered, respectively, as alternatives to PAC for selection and  $\Delta$ TK for suicide switch strategies and similarly warrant future consideration in the context of our current approach to improve the efficacy and safety of lentiviral transduction of HSCs.

The cell transduction efficiency of LVs is difficult to control and predict, even with purified vectors, and is seemingly highly dependent on patient samples.<sup>8,54,55</sup> Finding agents to enhance the frequency of hematopoietic cell transduction is being considered,<sup>56,57</sup> but higher transduction efficacy on the fraction of cells highly permissive to transduction may considerably increase the VCN per transduced cells and the risk for insertional mutagenesis.<sup>58</sup> Heterozygous  $\beta^0/\beta^+$ -thalassemia carriers are healthy. We have shown that the therapeutic  $\beta$ -globin output on a per-gene basis is at least 70% of the normal endogenous  $\beta$ -globin.<sup>1</sup> Thus, patients with transfusion-dependent  $\beta$ -thalassemia transplanted with 100% transduced allogeneic stem cells, carrying between one and two copies of a LV expressing  $\beta$ -globin, should produce therapeutic levels of the protein. Indian and Arabian homozygous patients with SCD having fetal hemoglobin (HbF) levels of about 25% have all the symptoms of SCD but a milder clinical course.<sup>59,60</sup> In patients with compound heterozygosity for HbS and  $\beta$ -globin gene deletion, with an inherited persistence of HbF (S-HPFH), pancellular distribution of HbF (typically 30%), as opposed to heterocellular, is considered to be responsible for the benign syndrome.<sup>61,62</sup> The mutant  $\beta^{A187Q}$ -globin chain expressed by our LTGCPU7 vector inhibits HbS polymerization as efficiently as the  $\gamma$ -globin chain.<sup>63</sup> Thus, the delivery of  $\geq 1$  vector copy per erythroid cells should ensure that anti-sickling Hb levels are high enough to cure most patients with SCD.

The decrease in total cell number due to selection may make it necessary to increase the number of cells initially transduced, especially for patients with SCD, for which autologous CD34<sup>+</sup> cells are obtained by bone marrow harvest because of the detrimental effect of granulocyte colony-stimulating factor (G-CSF) that promotes acute complications. For subjects with  $\beta$ -thalassemia, the association of G-CSF and plerixafor has been shown to display strong synergy,<sup>64</sup> increasing the number of mPB CD34<sup>+</sup> cells, and is the preferred combination in ongoing clinical trials.<sup>2</sup> Plerixafor has also been shown to mobilize HSCs effectively without the detrimental effect of G-CSF. It is currently tested in humans with severe SCD (NCT02140554 and NCT02193191) and may enhance the number of collected hematopoietic (stem) cells prior to *ex vivo* culture. Additionally, one part of the drug product may be expanded *ex vivo* with agents such as StemRegenin-1 or UM171 before cell infusion, in order to support and accelerate neutrophil and platelet recovery.<sup>65,66</sup> Myeloablative conditioning with busulfan also needs to be optimally delivered to achieve maximal hematopoietic engraftment by meticulous monitoring by pharmacokinetics analyses and adjustment of the dosing

to a fully myeloablative target value, as currently conducted in ongoing clinical trials.<sup>2,5</sup>

The protocol described in this manuscript ensures efficient selection, ensures high recovery rate of transduced hematopoietic stem/progenitor cells after only 24 hr of exposure to selective agents already approved by medical agencies, and reduces the risks of multiple lentiviral insertions within susceptible hematopoietic cell subpopulations. The vector titer is close to that of a vector currently used in clinical trials, and the  $\beta$ -globin expression level is similar. The drug selection method is affordable, rapid, easily scalable, and does not bias the distribution of lentiviral insertion sites beyond the integration preferences observed after lentiviral transduction of HSCs and reconstitution of hematopoiesis. High transduction rates can be obtained with a low/medium VCN per cell, and the vector provides a means of eradicating cells containing the modified gene *in vivo* if a serious adverse event occurs. Toxicological studies, including evaluation of the genotoxic potential of the modified vector, will have to be performed before evaluation in subjects with  $\beta^0/\beta^0$ -thalassemia and severe SCD.

## MATERIALS AND METHODS

### LV Preparation, Titration, and CD34<sup>+</sup> Cell Transduction

The PAC and human PGK promoter sequences were synthesized by Genscript (Piscataway, NJ, USA) and introduced between the  $\beta$ -globin locus control region and the 3' polypurine tract of HPV524 (HPV569<sup>1</sup> minus insulators) to make LTGCPU1. LTGCPU7 is derived from the more recent  $\beta$ -globin LV BB305<sup>14</sup> (Figure 1). In the products of PAC/ $\Delta$ TK constructs, the two proteins are linked via a glycine-serine spacer (GS) and  $\Delta$ TK starts at the second ATG of the full-length HSV-TK sequence. The PAC/ $\Delta$ TK<sub>opt</sub> cassette was optimized and synthesized by DNA2.0 (Newark, NJ, USA) for optimal expression in human cells. Its sequence, together with that of the EFS promoter, is given in Figure S1. Lentiviral particles were produced as described previously<sup>67</sup> and concentrated by ultracentrifugation. Where indicated, purification was performed with a Mustang Q anion exchange membrane cartridge (Pall, Saint Germain-en-Laye, France) and a 40K ZebaSpin desalting column (Thermo Fisher Scientific, Villebon, France) before concentration. Infectious titers were determined in NIH 3T3 cells. Adult and CB CD34<sup>+</sup> cells were cultured in SCGM (CellGenix, Freiburg, Germany) and X-VIVO20 (Lonza, Basel, Switzerland) medium, respectively. Two days after cell transduction, cells were selected on puromycin hydrochloride (Sigma-Aldrich, Lyon, France). Where indicated, mifepristone, reserpine, cyclosporine A, and verapamil hydrochloride (all from Sigma-Aldrich) were used at final concentrations of 5, 20, 2, and 20  $\mu$ M respectively, together with puromycin (5  $\mu$ g/mL).

### Immunodeficient Mice

NOD.Cg-Prkdc<sup>scid</sup> Il2rg<sup>tm1Wjl</sup>/SzJ (NOD-*scid* IL2r $\gamma$ <sup>null</sup> abbreviated to NSG) mice were obtained from Charles River Laboratories (Saint Germain sur l'Arbresle, France) and kept in individually ventilated cages supplied with sterile food and autoclaved water. Animal care and handling conformed to European Union (EU) Directive

2010/63/EU on the protection of animals used for scientific purposes. Experimental protocols were approved by the ethics committee for animal experimentation of the CEA (French alternative energies and atomic energy commission), under notification number 12-033. Three- to four-month-old female recipients (at least six mice per group) were sublethally irradiated with a dose of 2.8 Gy (1 Gy/min) to the whole body (total body irradiation [TBI]). We then injected  $10^4$ – $10^5$  transduced and selected human cells intravenously into the mice on the day after irradiation. Where indicated, mice were treated, 1 month after cell infusion, with intraperitoneal ganciclovir (Cymevan, Roche Pharma), at a dosage of 50 mg/kg/day (or saline), 6 days per week for 2 weeks. Bone marrow hematopoietic cells were harvested from the left and right femurs and tibias, 3 months after the transplantation of human cells. They were used for human progenitor assays, flow cytometric analysis, and hCD45 cell sorting.

### Human Cells

The study was approved by the ethics evaluation committee of INSERM, institutional review board (IRB00003888) of the French institute of medical research and health (IORG0003254 FWA00005831). CB cells were obtained from Saint-Louis Hospital (Paris, France). mPB cells were collected from the residual cells in bags prepared for transplantation at Pitié Salpêtrière Hospital (Paris, France). Adult bone marrow cells were obtained from hip operations at the “Centre Hospitalier Intercommunal” (Créteil, France). The mononuclear cell fractions were isolated by gradient separation, and human CD34<sup>+</sup> cells were isolated with the CD34<sup>+</sup> progenitor cell isolation kit (Miltenyi Biotec, Paris, France) and the autoMACSPro instrument (Miltenyi Biotec).

### Progenitor Assays and VCN Determination in Human Progenitor Cells

For progenitor assays performed immediately after *in vitro* transduction and selection, 500–1,000 cells were plated in methylcellulose-based medium already supplemented with human cytokines H4434 (STEMCELL Technologies, Grenoble, France). For progenitor assays performed with human cells isolated from immunodeficient mice after transplantation, the growth of mouse cells was prevented by plating 100,000–1,000,000 cells on methylcellulose-based medium H4230 (STEMCELL Technologies) supplemented with human stem cell factor (50 ng/mL; PeproTech, Neuilly-sur-Seine, France), human interleukin-3 (10 ng/mL; PeproTech), and human erythropoietin (3 U/mL; Roche Pharma). The colonies were counted and collected after incubation for 11–14 days, washed with PBS, and stored. LTC-ICs were analyzed as previously described.<sup>3</sup> The percentage of vector-modified LTC-ICs during the readout phase of the assay was determined by subjecting no more than one clonogenic myeloid colony per well to qPCR-based scoring to ensure that only independent LTC-ICs were analyzed. DNA from hematopoietic colonies was recovered and amplified with the TaqMan Sample-to-SNP kit (Thermo Fisher Scientific) and specific primers and probes ([Supplemental Information](#)) using the 7300 ABI Prism Detection system (Applied Biosystems, Courtaboeuf, France). The results were

compared with those obtained with genomic DNA from a human cell line containing a single copy of the integrated LV per haploid genome. VCNs were determined by the comparative cycle threshold (Ct) method and set at integer values.

### Flow Cytometry and SCID-Repopulating Cell Assays

The antibodies used for cytometry were directed against murine CD45 (clone 30-F11; Thermo Fisher Scientific), human CD45 (clone 5B1; Miltenyi Biotec), human CD33 (clone WM53; BD Biosciences, Le Pont de Claix, France), human CD15 (clone VIMC6; Miltenyi Biotec), human CD20 (clone 2H7; BD Biosciences), human CD3 (clone BW264/56; Miltenyi Biotec), human CD133 (clone AC133; Miltenyi Biotec), human CD34 (clone 581; BD Biosciences), and human CD243 (clone UIC2; Thermo Fisher Scientific). Isotype control antibodies were used to set the negative gates for the *in vitro* identification of human cell subsets. Non-viable cells were excluded by SYTOX Blue nucleic acid staining (Thermo Fisher Scientific) or with 7AAD (Thermo Fisher Scientific). Data were acquired with a MACSQUANT instrument (Miltenyi Biotec) and analyzed with FlowJo software (Tree Star, Ashland, OR, USA). The percentages of human cells in NSG bone marrow were calculated as follows: the number of human CD45 cells was divided by the number of human plus mouse CD45 cells, and the result was multiplied by 100. For limiting dilution assays and the assessment of human stem cell frequency, mice were arbitrarily considered to be positive for human SRCs if the level of human chimerism 3 months after transplantation was >1%, with the presence of both myeloid and lymphoid cells. Mice that had not undergone the transplantation procedure were used as a negative control. Background levels of human cell detection in these negative control mice were  $\approx 0.01\%$  ([Figure S6](#)). The frequency of SRCs was calculated by using Poisson statistics and L-Calc software (STEMCELL Technologies).

### Human CD45 Cell Enrichment for VCN Determination

Human cells from NSG mouse bone marrow were enriched to facilitate VCN determination in these cells and to confirm the absence of human cells in immunodeficient mice as follows: single-cell suspensions of bone marrow cells (20–70 million cells) were labeled by incubation with 0.2  $\mu$ L of whole-blood CD45 microbeads (Miltenyi Biotec) in 50  $\mu$ L of PBS supplemented with bovine serum albumin (0.5%) and EDTA (2 mM) for 15 minutes, washed, resuspended in 1 mL, and sorted with an autoMACSPro instrument (Miltenyi Biotec) and the *possel\_s* program. Sorted cells were then resuspended in PBS for the evaluation of cell enrichment by FACS and DNA extraction with the NucleoSpin Blood kit (Macherey Nagel). Relative enrichment varied between 1- and 20-fold, depending on the initial proportion of human cells ([Figure S6](#)). In mice with levels of hCD45<sup>+</sup> cells below background levels (0.01%), no enrichment was observed, confirming the absence of human cells. Genomic DNA was extracted with the NucleoSpin Blood kit (Macherey Nagel, Hoerdt, France). Real-time PCR was performed with specific primers and probes ([Supplemental Information](#)) using the 7300 ABI Prism Detection system and a 2 $\times$  qPCR MasterMix containing Rox (Erogenec, Liege, Belgium). VCNs were determined from the results obtained

with genomic DNA from a human cell line containing one copy of the integrated LV per haploid genome.

### Transgenic Human Hemoglobin from Individual Erythroid Colonies

Cells from erythroid bursts were lysed by adding 100  $\mu$ L of H<sub>2</sub>O and centrifuging at 20,000  $\times$  g. The nuclei were retained for qPCR and the assessment of vector modification, whereas hemoglobin from individual erythroid colony supernatants was separated by ion exchange high-pressure liquid chromatography (HPLC) on a PolyCAT A column (PolyLC). Elution was achieved with a linear gradient of buffer A (40 mM Tris, 3 mM KCN; pH adjusted to 6.5 with acetic acid) and buffer B (40 mM Tris, 3 mM KCN, 200 mM NaCl; pH adjusted to 6.5 with acetic acid) of different ionic strengths at a flow rate of 0.4 mL/min at 40°C. The detection wavelength was 418 nm. HPLC analyses were performed with a Prominence chromatograph (Shimadzu) and LC Solution software.

### Poisson Statistics and Vector Distribution

#### Distribution of VCN per Cell and Percentage of Transduced Cells in Non-selective Conditions

If cells from a specific lineage are transducible at a known mean rate ( $\mu$ ), if they are equally transducible, and if integration events occur independently of each other, the number of vector copies per cell ( $X$ ) is a discrete random variable the scattering of which obeys a Poisson distribution. The probability mass function of  $X$  is given by  $f(i; \mu) = \Pr(X = i; \mu) = (e^{-\mu}) \times \mu^i / i!$ , where  $i$  is the discrete number of vectors integrated per cell and  $\mu$  is the observed mean VCN (Figure S8A). Thus, the probability of a cell being transduced is  $\Pr(X \geq 1; \mu) = 1 - \Pr(X = 0; \mu) = 1 - e^{-\mu}$ . Therefore, if the mean VCN/cell ( $\mu$ ) in a specific cell population is known, and if the distribution of VCN per cell follows a Poisson distribution, then the expected percentage of transduced cells in non-selective conditions is  $(1 - e^{-\mu}) \times 100$  (Figure S8B).

#### Expected Mean VCN after the Selection of Transduced Cells in Optimal Conditions

In conditions of optimal selection, all non-transduced cells are eradicated and all vector-bearing cells survive. If, first, vector distribution before cell selection follows a Poisson distribution, and second, selection occurs in optimal conditions, then the expected VCN after selection is:

$$\text{expVCN}_{\text{Sel}} = \frac{(\Pr(X = 1; \mu) \times 1) + (\Pr(X = 2; \mu) \times 2) + \dots + (\Pr(X = n; \mu) \times n)}{1 - \Pr(X = 0; \mu)} = \frac{e^{-\mu}}{1 - e^{-\mu}} \times \sum_{i=1}^{\infty} \frac{\mu^i}{i!}$$

when  $\mu \geq 5$ ,  $\text{expVCN}_{\text{Sel}} \approx \mu$  (Figure S8C). If the distribution of the vector before selection does not obey Poisson statistics, then the expected VCN after selection in optimal selective conditions cannot be predicted in this way.

#### Expected Distribution after the Selection of Transduced Cells in Optimal Conditions

This distribution is calculated based on the observed distribution before selection. For each discrete number of integrated vectors per cell,  $\Pr(X = i)_A = (\Pr(X = i)_B) / (1 - \Pr(X = 0)_B)$ , where the A and B indices relate to values measured after and before selection, respectively.

#### Insertion Site Retrieval

Human CD45<sup>+</sup> (hCD45<sup>+</sup>) cells were harvested from the bone marrow of mice 3 months after the infusion of transduced human CD34<sup>+</sup> cells. The numbers of insertion site recovery were increased by preparing genomic DNA from these cells and their progeny (obtained by culture in methylcellulose). The number of *in vivo* insertion sites was calculated as the sum of unique insertion sites retrieved from these two DNA preparations. Before transplantation, aliquots of cells were subjected to culture on methylcellulose from which genomic DNA was prepared. The number of *in vitro* insertion sites was determined with these cells. A detailed explanation of the methodology used for the processing and analysis of sequencing files is included in the [Supplemental Information](#).

#### Statistical Tests

Comparisons between two or more groups were performed using Student's t test, Mann-Whitney U tests (if the data were shown to be non-normal in Shapiro-Wilk tests and/or if equal variance test failed), and one-way ANOVA (possibly based on ranks). Degrees of significance were calculated with SigmaPlot 10.0 software. Best-fit curves rising to a maximum were constructed with SigmaPlot 10.0. We assessed the significance of differences between proportions of cells bearing various numbers of LV copies in experimental groups and expectations, by creating 2  $\times$  2 contingency tables assuming equivalent numbers of cells in the two groups. Non-random associations between variables were assessed by Fisher's exact test. Odds ratio and 95% confidence interval were calculated with GraphPad Prism 6. The frequency of LTC-ICs and SRCs and the significance of differences between groups were determined with L-Calc software (STEMCELL Technologies).

#### Insertion Site Retrieval and Analysis

See [Supplemental Information](#).

---

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Materials and Methods, fourteen figures, and two tables and can be found with this article online at <https://doi.org/10.1016/j.ymthe.2017.10.015>.

## AUTHOR CONTRIBUTIONS

K.B. and C.C. performed the *in vivo* experiments, with the assistance of B.G.-L., K.S.-F., and L.M. K.B., E.D., M.G., C.C., O.N., S.B., and J.D.D. performed the *in vitro* experiments, with the assistance of M.D., B.G.-L., and A.P. K.B. and J.C. prepared the lentiviral vectors. H.T.-N. provided human hematopoietic adult cells. K.B., E.D., M.G., and C.C. designed the experiments. J.D.D., P.L., and E.P. developed the concept and designed the experiments. K.B., E.D., and E.P. analyzed the data. E.P. supervised the work and wrote the manuscript.

## CONFLICTS OF INTEREST

O.N. is an employee of bluebird bio, Inc. (Cambridge, MA, USA). M.D., B.G.-L., J.C., A.P., and J.D.D. were employees of bluebird bio, Inc. E.P. had financial relationships with bluebird bio, Inc. P.L. has financial relationships with bluebird bio, Inc. None of the other authors have any competing interests to disclose.

## ACKNOWLEDGMENTS

This work was supported by an Industrial Chair (HemGenTher) from France's "Agence Nationale pour la Recherche" (ANR) awarded to P.L. We thank Manfred Schmidt and Cynthia Bartholomae (National Center for Tumor Diseases, Heidelberg, Germany) for teaching the linear amplification-mediated (LAM)-PCR technique to E.D., Inpyo Hwang for technical help, Che Serguera and Noëlle Dufour (MIRCen, CEA de Fontenay aux Roses, France) for their help in the implementation of the LAM-PCR protocol, and Benjamin Versier and Didier Thenadey (MIRCen, CEA de Fontenay aux Roses, France) for implementing the local Galaxy platform. Next-generation sequencing was performed at the high-throughput sequencing core facility of I2BC (Centre de Recherche de Gif-sur-Yvette, France; <http://www.i2bc.paris-saclay.fr>), with expert assistance from its staff. Hematopoietic cord blood cells, mobilized adult cells, and adult bone marrow cells from patients with sickle cell disease were donated by Prof. Jérôme Larghero (Saint-Louis Hospital, Paris, France), Dr. Françoise Norol (Pitié-Salpêtrière Hospital, Paris, France), and Prof. Philippe Hernigou (Henri-Mondor Hospital, Créteil, France), respectively. We thank the "Fondation Générale de Santé" (FGDS) and the "Assistance Publique-Hôpitaux de Paris (AP-HP)" for collecting and testing cord blood cells.

## REFERENCES

- 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  $\beta$ -thalassaemia. *Nature* 467, 318–322.
- Negre, O., Eggimann, A.V., Beuzard, Y., Ribeil, J.A., Bourget, P., Borwornpinyo, S., Hongeng, S., Hacein-Bey, S., Cavazzana, M., Leboulch, P., and Payen, E. (2016). Gene therapy of the  $\beta$ -hemoglobinopathies by lentiviral transfer of the  $\beta$ (A(T87Q))-globin gene. *Hum. Gene Ther.* 27, 148–165.
- Payen, E., and Leboulch, P. (2012). Advances in stem cell transplantation and gene therapy in the beta-hemoglobinopathies. *Hematology Am. Soc. Hematol. Educ. Program* 2012, 276–283.
- Ribeil, J.A., Hacein-Bey, S., Payen, E., Semeraro, M., Elisa, M., Caccavelli, L., Touzot, F., Lefrere, F., Suarez, F., Hermine, O., et al. (2016). Update from the Hgb-205 phase 1/2 clinical study of lentiglobin gene therapy: sustained clinical benefit in severe hemoglobinopathies. *Blood* 128, 2311.
- 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.
- Thompson, A.A., Kwiatkowski, J., Rasko, J., Hongeng, S., Schiller, G., Anurathapan, U., Cavazzana, M., Joy Ho, P., von Kalle, C., Kletzel, M., et al. (2016). Lentiglobin gene therapy for transfusion-dependent  $\beta$ -thalassemia: update from the Northstar Hgb-204 phase 1/2 clinical study. *Blood* 128, 1175.
- Walters, M.C., Rasko, J.E., Hongeng, S., Kwiatkowski, J.L., Schiller, G.J., Kletzel, M., Joy Ho, P., Vichinsky, E., von Kalle, C., Cavazzana, M., et al. (2015). Update of results from the Northstar study (HGB-204): A phase 1/2 study of gene therapy for  $\beta$ -thalassemia major via transplantation of autologous hematopoietic stem cells transduced ex vivo with a lentiviral  $\beta$ A-T87Q-globin vector (Lentiglobin BB305 drug product). *Blood* 126, 201.
- Kanter, J., Walters, M.C., Hsieh, M.M., Krishnamurti, L., Kwiatkowski, J., Kamble, R.T., von Kalle, C., Kuypers, F.A., Cavazzana, M., Leboulch, P., et al. (2016). Interim results from a phase 1/2 clinical study of lentiglobin gene therapy for severe sickle cell disease. *Blood* 128, 1176.
- Steinberg, M.H., Chui, D.H., Dover, G.J., Sebastiani, P., and Al Sultan, A. (2014). Fetal hemoglobin in sickle cell anemia: a glass half full? *Blood* 123, 481–485.
- 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.
- Pawliuk, R., Bachelot, T., Wise, R.J., Mathews-Roth, M.M., and Leboulch, P. (1999). Long-term cure of the photosensitivity of murine erythropoietic protoporphyria by preselective gene therapy. *Nat. Med.* 5, 768–773.
- Kalberer, C.P., Pawliuk, R., Imren, S., Bachelot, T., Takekoshi, K.J., Fabry, M., Eaves, C.J., London, I.M., Humphries, R.K., and Leboulch, P. (2000). Preselection of retrovirally transduced bone marrow avoids subsequent stem cell gene silencing and age-dependent extinction of expression of human beta-globin in engrafted mice. *Proc. Natl. Acad. Sci. USA* 97, 5411–5415.
- 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 pan-erythroid correction of murine beta thalassemia by multiple lentiviral integration in hematopoietic stem cells. *Proc. Natl. Acad. Sci. USA* 99, 14380–14385.
- 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  $\beta$ -thalassemia and sickle cell disease. *Curr. Gene Ther.* 15, 64–81.
- Kutner, R.H., Zhang, X.Y., and Reiser, J. (2009). Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat. Protoc.* 4, 495–505.
- Merten, O.W., Charrier, S., Laroudie, N., Fauchille, S., Dugué, C., Jenny, C., Audit, M., Zanta-Boussif, M.A., Chautard, H., Radrizzani, M., et al. (2011). Large-scale manufacture and characterization of a lentiviral vector produced for clinical ex vivo gene therapy application. *Hum. Gene Ther.* 22, 343–356.
- 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.
- 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.
- Deola, S., Scaramuzza, S., Birolo, R.S., Carballido-Perrig, N., Ficara, F., Mucchetti, C., Dando, J., Carballido, J.M., Bordignon, C., Roncarolo, M.G., et al. (2004). Mobilized blood CD34+ cells transduced and selected with a clinically applicable protocol reconstitute lymphopoiesis in SCID-Hu mice. *Hum. Gene Ther.* 15, 305–311.
- Tisdale, J.F., Hanazono, Y., Sellers, S.E., Agricola, B.A., Metzger, M.E., Donahue, R.E., and Dunbar, C.E. (1998). Ex vivo expansion of genetically marked rhesus peripheral blood progenitor cells results in diminished long-term repopulating ability. *Blood* 92, 1131–1141.
- Kennedy, D.R., McLellan, K., Moore, P.F., Henthorn, P.S., and Felsburg, P.J. (2009). Effect of ex vivo culture of CD34+ bone marrow cells on immune reconstitution of

- XSCID dogs following allogeneic bone marrow transplantation. *Biol. Blood Marrow Transplant* 15, 662–670.
22. Sellers, S., Gomes, T.J., Larochelle, A., Lopez, R., Adler, R., Krouse, A., Donahue, R.E., Childs, R.W., and Dunbar, C.E. (2010). Ex vivo expansion of retrovirally transduced CD34+ cells results in overrepresentation of clones with MDS1/EVI1 insertion sites in the myeloid lineage after transplantation. *Mol. Ther.* 18, 1633–1639.
  23. Theile, D., Staffen, B., and Weiss, J. (2010). ATP-binding cassette transporters as pitfalls in selection of transgenic cells. *Anal. Biochem.* 399, 246–250.
  24. Abonour, R., Williams, D.A., Einhorn, L., Hall, K.M., Chen, J., Coffman, J., Traycoff, C.M., Bank, A., Kato, L., Ward, M., et al. (2000). Efficient retrovirus-mediated transfer of the multidrug resistance 1 gene into autologous human long-term repopulating hematopoietic stem cells. *Nat. Med.* 6, 652–658.
  25. Moscov, J.A., Huang, H., Carter, C., Hines, K., Zujewski, J., Cusack, G., Chow, C., Venzon, D., Sorrentino, B., Chiang, Y., et al. (1999). Engraftment of MDR1 and NeoR gene-transduced hematopoietic cells after breast cancer chemotherapy. *Blood* 94, 52–61.
  26. Ward, M., Pioli, P., Ayello, J., Reiss, R., Urzi, G., Richardson, C., Hesdorffer, C., and Bank, A. (1996). Retroviral transfer and expression of the human multiple drug resistance (MDR) gene in peripheral blood progenitor cells. *Clin. Cancer Res.* 2, 873–876.
  27. Takahashi, M., Matsuoka, Y., Sumide, K., Nakatsuka, R., Fujioka, T., Kohno, H., Sasaki, Y., Matsui, K., Asano, H., Kaneko, K., and Sonoda, Y. (2014). CD133 is a positive marker for a distinct class of primitive human cord blood-derived CD34-negative hematopoietic stem cells. *Leukemia* 28, 1308–1315.
  28. Eszterhas, S.K., Bouhassira, E.E., Martin, D.I., and Fiering, S. (2002). Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol. Cell. Biol.* 22, 469–479.
  29. Carbonaro, D.A., Zhang, L., Jin, X., Montiel-Equihua, C., Geiger, S., Carmo, M., Cooper, A., Fairbanks, L., Kaufman, M.L., Sebire, N.J., et al. (2014). Preclinical demonstration of lentiviral vector-mediated correction of immunological and metabolic abnormalities in models of adenosine deaminase deficiency. *Mol. Ther.* 22, 607–622.
  30. Zychlinski, D., Schambach, A., Modlich, U., Maetzig, T., Meyer, J., Grassman, E., Mishra, A., and Baum, C. (2008). Physiological promoters reduce the genotoxic risk of integrating gene vectors. *Mol. Ther.* 16, 718–725.
  31. Payen, E., Colomb, C., Negre, O., Beuzard, Y., Hehir, K., and Lebouloch, P. (2012). Lentivirus vectors in  $\beta$ -thalassemia. *Methods Enzymol.* 507, 109–124.
  32. Varma, M.V., Ashokraj, Y., Dey, C.S., and Panchagnula, R. (2003). P-glycoprotein inhibitors and their screening: a perspective from bioavailability enhancement. *Pharmacol. Res.* 48, 347–359.
  33. Petrillo, C., Cesana, D., Piras, F., Bartolaccini, S., Naldini, L., Montini, E., and Kajaste-Rudnitski, A. (2015). Cyclosporin A and rapamycin relieve distinct lentiviral restriction blocks in hematopoietic stem and progenitor cells. *Mol. Ther.* 23, 352–362.
  34. Jennings, P., Koppelstaetter, C., Aydin, S., Abberger, T., Wolf, A.M., Mayer, G., and Pfaller, W. (2007). Cyclosporine A induces senescence in renal tubular epithelial cells. *Am. J. Physiol. Renal Physiol.* 293, F831–F838.
  35. Wolf, A., Trendelenburg, C.F., Diez-Fernandez, C., Prieto, P., Houy, S., Trommer, W.E., and Cordier, A. (1997). Cyclosporine A-induced oxidative stress in rat hepatocytes. *J. Pharmacol. Exp. Ther.* 280, 1328–1334.
  36. Uchida, N., Leung, F.Y., and Eaves, C.J. (2002). Liver and marrow of adult *mdr-1a/1b(-/-)* mice show normal generation, function, and multi-tissue trafficking of primitive hematopoietic cells. *Exp. Hematol.* 30, 862–869.
  37. Yamada, K., McCarty, D.M., Madden, V.J., and Walsh, C.E. (2003). Lentivirus vector purification using anion exchange HPLC leads to improved gene transfer. *Biotechniques* 34, 1074–1078, 1080.
  38. Griffin, D.O., and Goff, S.P. (2015). HIV-1 is restricted prior to integration of viral DNA in primary cord-derived human CD34+ cells. *J. Virol.* 89, 8096–8100.
  39. Griffin, D.O., and Goff, S.P. (2016). Restriction of HIV-1-based lentiviral vectors in adult primary marrow-derived and peripheral mobilized human CD34+ hematopoietic stem and progenitor cells occurs prior to viral DNA integration. *Retrovirology* 13, 14.
  40. Charrier, S., Ferrand, M., Zerbato, M., Précigout, G., Viornery, A., Bucher-Laurent, S., Benkhalifa-Ziyyat, S., Merten, O.W., Perea, J., and Galy, A. (2011). Quantification of lentiviral vector copy numbers in individual hematopoietic colony-forming cells shows vector dose-dependent effects on the frequency and level of transduction. *Gene Ther.* 18, 479–487.
  41. Greco, R., Oliveira, G., Stanghellini, M.T., Vago, L., Bondanza, A., Peccatori, J., Cieri, N., Markt, S., Mastaglio, S., Bordignon, C., et al. (2015). Improving the safety of cell therapy with the TK-suicide gene. *Front. Pharmacol.* 6, 95.
  42. Salomon, B., Maury, S., Loubière, L., Caruso, M., Onclercq, R., and Klatzmann, D. (1995). A truncated herpes simplex virus thymidine kinase phosphorylates thymidine and nucleoside analogs and does not cause sterility in transgenic mice. *Mol. Cell. Biol.* 15, 5322–5328.
  43. Chen, X., Gao, W., Gambotto, A., and Finn, O.J. (2009). Lentiviral vectors encoding human MUC1-specific, MHC-unrestricted single-chain TCR and a fusion suicide gene: potential for universal and safe cancer immunotherapy. *Cancer Immunol. Immunother.* 58, 977–987.
  44. Junker, K., Koehl, U., Zimmerman, S., Stein, S., Schwabe, D., Klingebiel, T., and Grez, M. (2003). Kinetics of cell death in T lymphocytes genetically modified with two novel suicide fusion genes. *Gene Ther.* 10, 1189–1197.
  45. Gschwend, E.H., McCracken, M.N., Kaufman, M.L., Ho, M., Hollis, R.P., Wang, X., Saini, N., Koya, R.C., Chodon, T., Ribas, A., et al. (2014). HSV-sr39TK positron emission tomography and suicide gene elimination of human hematopoietic stem cells and their progeny in humanized mice. *Cancer Res.* 74, 5173–5183.
  46. Riddell, S.R., Elliott, M., Lewinsohn, D.A., Gilbert, M.J., Wilson, L., Manley, S.A., Lupton, S.D., Overell, R.W., Reynolds, T.C., Corey, L., and Greenberg, P.D. (1996). T-cell mediated rejection of gene-modified HIV-specific cytotoxic T lymphocytes in HIV-infected patients. *Nat. Med.* 2, 216–223.
  47. Berger, C., Flowers, M.E., Warren, E.H., and Riddell, S.R. (2006). Analysis of transgene-specific immune responses that limit the in vivo persistence of adoptively transferred HSV-TK-modified donor T cells after allogeneic hematopoietic cell transplantation. *Blood* 107, 2294–2302.
  48. Kanaji, S., Fahs, S.A., Ware, J., Montgomery, R.R., and Shi, Q. (2014). Non-myeloablative conditioning with busulfan before hematopoietic stem cell transplantation leads to phenotypic correction of murine Bernard-Soulier syndrome. *J. Thromb. Haemost.* 12, 1726–1732.
  49. Moghimi, B., Sack, B.K., Nayak, S., Markusic, D.M., Mah, C.S., and Herzog, R.W. (2011). Induction of tolerance to factor VIII by transient co-administration with rapamycin. *J. Thromb. Haemost.* 9, 1524–1533.
  50. McMahon, G., Weir, M.R., Li, X.C., and Mandelbrot, D.A. (2011). The evolving role of mTOR inhibition in transplantation tolerance. *J. Am. Soc. Nephrol.* 22, 408–415.
  51. Bhatt, K.H., Rudraraju, R., Brooks, J.F., Jung, J.W., Galea, R., Wells, J.W., and Steptoe, R.J. (2017). Short-course rapamycin treatment enables engraftment of immunogenic gene-engineered bone marrow under low-dose irradiation to permit long-term immunological tolerance. *Stem Cell Res. Ther.* 8, 57.
  52. Fehr, T., and Sykes, M. (2004). Tolerance induction in clinical transplantation. *Transpl. Immunol.* 13, 117–130.
  53. Barese, C.N., Felizardo, T.C., Sellers, S.E., Keyvanfar, K., Di Stasi, A., Metzger, M.E., Krouse, A.E., Donahue, R.E., Spencer, D.M., and Dunbar, C.E. (2015). Regulated apoptosis of genetically modified hematopoietic stem and progenitor cells via an inducible caspase-9 suicide gene in rhesus macaques. *Stem Cells* 33, 91–100.
  54. Cavazzana, M., Ribeil, J.A., Payen, E., Touzot, F., Neven, B., Lefrere, F., Suarez, F., Magrin, E., Beuzard, Y., Chretien, S., et al. (2016). Clinical outcomes of gene therapy with BB305 lentiviral vector for sickle cell disease and  $\beta$ -thalassemia. *Mol. Ther.* 24 (Suppl 1), S111–S112.
  55. Thompson, A.A., Kwiatkowski, J.L., Rasko, J., Hongeng, S., Schiller, G.J., Anurathapan, U., Cavazzana, M., Joy Ho, P., von Kalle, C., Kletzel, M., et al. (2016). Lentiglobin gene therapy for transfusion-dependent  $\beta$ -thalassemia: update from the Northstar Hgb-204 phase 1/2 clinical study. *Blood* 128, 1175.
  56. Heffner, G., Bonner, M., Campbell, J., Christiansen, L., Pierciey, F.J., Zhang, W., Lewis, G., Smurnyy, Y., Hamel, A., Shah, S., et al. (2016). PGE2 increases lentiviral vector transduction efficiency of human HSC. *Mol. Ther.* 24 (Suppl 1), S89–S90.

57. Timberlake, N., Ozog, S., D'souza, S., Slukvin, I., Catz, S.D., Snyder, S., and Torbet, B.E. (2016). b-deliverin: a small molecule for improving gene transfer to hematopoietic stem cells and probing mechanisms of lentiviral vector restriction. *Mol. Ther* 24 (Suppl 1), S2–S3.
58. Woods, N.B., Muessig, A., Schmidt, M., Flygare, J., Olsson, K., Salmon, P., Trono, D., von Kalle, C., and Karlsson, S. (2003). Lentiviral vector transduction of NOD/SCID repopulating cells results in multiple vector integrations per transduced cell: risk of insertional mutagenesis. *Blood* 101, 1284–1289.
59. Brittenham, G., Lozoff, B., Harris, J.W., Mayson, S.M., Miller, A., and Huisman, T.H. (1979). Sickle cell anemia and trait in southern India: further studies. *Am. J. Hematol.* 6, 107–123.
60. Pembrey, M.E., Wood, W.G., Weatherall, D.J., and Perrine, R.P. (1978). Fetal haemoglobin production and the sickle gene in the oases of Eastern Saudi Arabia. *Br. J. Haematol.* 40, 415–429.
61. Murray, N., Serjeant, B.E., and Serjeant, G.R. (1988). Sickle cell-hereditary persistence of fetal haemoglobin and its differentiation from other sickle cell syndromes. *Br. J. Haematol.* 69, 89–92.
62. Ngo, D.A., Aygun, B., Akinsheye, I., Hankins, J.S., Bhan, I., Luo, H.Y., Steinberg, M.H., and Chui, D.H. (2012). Fetal haemoglobin levels and haematological characteristics of compound heterozygotes for haemoglobin S and deletional hereditary persistence of fetal haemoglobin. *Br. J. Haematol.* 156, 259–264.
63. 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.
64. Yannaki, E., Karponi, G., Zervou, F., Constantinou, V., Bouinta, A., Tachynopoulou, V., Kotta, K., Jonlin, E., Papayannopoulou, T., Anagnostopoulos, A., and Stamatoyannopoulos, G. (2013). Hematopoietic stem cell mobilization for gene therapy: superior mobilization by the combination of granulocyte-colony stimulating factor plus plerixafor in patients with  $\beta$ -thalassemia major. *Hum. Gene Ther.* 24, 852–860.
65. Wagner, J.E., Jr., Brunstein, C.G., Boitano, A.E., DeFor, T.E., McKenna, D., Sumstad, D., Blazar, B.R., Tolar, J., Le, C., Jones, J., et al. (2016). Phase I/II trial of StemRegenin-1 expanded umbilical cord blood hematopoietic stem cells supports testing as a stand-alone graft. *Cell Stem Cell* 18, 144–155.
66. Zonari, E., Desantis, G., Petrillo, C., Boccalatte, F.E., Lidonnici, M.R., Kajaste-Rudnitski, A., Aiuti, A., Ferrari, G., Naldini, L., and Gentner, B. (2017). Efficient ex vivo engineering and expansion of highly purified human hematopoietic stem and progenitor cell populations for gene therapy. *Stem Cell Reports* 8, 977–990.
67. Westerman, K.A., Ao, Z., Cohen, E.A., and Leboulch, P. (2007). Design of a trans protease lentiviral packaging system that produces high titer virus. *Retrovirology* 4, 96.



## **Supplemental Information**

### ***Ex Vivo* Selection of Transduced Hematopoietic Stem Cells for Gene Therapy of $\beta$ -Hemoglobinopathies**

**Kanit Bhukhai, Edouard de Dreuzy, Marie Giorgi, Charlotte Colomb, Olivier Negre, Maria Denaro, Béatrix Gillet-Legrand, Joëlle Cheuzeville, Anaïs Paulard, Hélène Trebeden-Negre, Suparerk Borwornpinyo, Karine Sii-Felice, Leila Maouche, Julian D. Down, Phillippe Leboulch, and Emmanuel Payen**

## **SUPPLEMENTAL SECTION**

- Supplemental Figure Legends
- Supplemental Figures 1-14
- Supplemental Tables 1-2
- Supplemental Methods
  - Insertion site retrieval and bioinformatics
  - Primer list
- Supplemental References

## SUPPLEMENTAL FIGURE LEGENDS

### Supplemental Figure 1: Sequence of the optimized expression cassette EFS-PAC/ $\Delta$ TK

Yellow: intron-less short EF-1 $\alpha$  promoter (EFS); blue: puromycin acetyl-transferase (PAC) from *Streptomyces alboniger*; green: deleted thymidine kinase gene of HHV1 ( $\Delta$ TK).

### Supplemental Figure 2: Cell culture and MDR1 expression in human hematopoietic CD34<sup>+</sup> cells

Five different cord blood CD34<sup>+</sup> cell samples (CB1 to CB5) were transduced with LTGCPU7 and analyzed for the presence of CD243 (MDR1), CD34, and CD133 antigens on the day of transduction (**A**: day 1) and two (**B**: day 3) and three days (**C**: day 4) later. Top panels: identification of CD34<sup>+</sup>CD133<sup>+</sup>, CD34<sup>+</sup>CD133<sup>-</sup> and CD34<sup>-</sup>CD133<sup>-</sup> cell subpopulations. Bottom panels: Proportion of cells expressing MDR1 (CD243) and mean fluorescence intensity (MFI) in the different cell populations. (**D**) Correlation of MDR1 and CD133 expression in CD34<sup>+</sup> cells on days 1, 3 and 4.

### Supplemental Figure 3: Puromycin selection on sorted MDR1<sup>-</sup> and MDR1<sup>+</sup> progenitor cells

CD34<sup>+</sup> cells from two different human cord blood samples were transduced with LTGCPU7 and sorted into CD34<sup>+</sup>CD243<sup>-</sup> (MDR<sup>-</sup>) and CD34<sup>+</sup>CD243<sup>+</sup> (MDR<sup>+</sup>) cell fractions. Cells were analyzed for the presence of CD243 (MDR1), CD34, and CD133 before cell sorting and in the enriched cell fractions (A1 and B1). They were treated with either 2  $\mu$ g/mL (**A**) or 1  $\mu$ g/mL, 0.3  $\mu$ g/mL, 0.1  $\mu$ g/mL puromycin (**B**), with (+) or without (-) cyclosporine A (2  $\mu$ M). The mean VCN values were determined in pooled progenitor cells (A2 and B2). Except with 0.1  $\mu$ g/mL puromycin, VCN obtained in puromycin treated cells were significantly higher than in the absence of the antibiotic (not shown). Fold-increases and p values are given for comparison of samples grown with and without cyclosporine A.

### Supplemental Figure 4: Effect of experimental conditions on SRC fitness

Human cord blood CD34<sup>+</sup> cells were transduced with LTGCPU7 and treated with puromycin + cyclosporine A (PC) or left untreated (-). On day 4, six groups of mice received 15,000, 40,000 and 140,000 treated or untreated cells. Three months later, the proportion of human cells present was evaluated by flow cytometry, to determine SRC frequencies (**A**) and the reconstitution capacity of SRCs in NSG mice (**B**). The mean VCN values were determined for human CD45<sup>+</sup> cells from individual mice (**C**). As in the experiment presented **Figure 5**, VCN was higher in mice receiving selected cells than in mice receiving non-selected cells and treated and non-treated cells had similar reconstitution capacities in NSG mice.

### Supplemental Figure 5: Effect of MDR1 inhibitors on repopulating cells

Three days after thawing, human cord blood CD34<sup>+</sup> cells were left untreated (-) or treated with verapamil (20  $\mu$ M) or cyclosporine A (2  $\mu$ M) for 24 hours. On day 4, three groups of 7 NSG mice received 70,000 treated or untreated cells. Ten weeks later, the proportion of human cells present was evaluated by flow cytometry to detect human leukocytes, human CD34<sup>+</sup> cells, lymphoid and myeloid cells. The proportions of human CD45<sup>+</sup> cells, human CD34<sup>+</sup> cells and the relative proportions of human CD20<sup>+</sup>, human CD3<sup>+</sup>, and human CD33<sup>+</sup>/CD15<sup>+</sup> were not statistically different between the three groups.

### Supplemental Figure 6: Enrichment of NSG bone marrow in human CD45<sup>+</sup> cells

Representative flow cytometry analyses on four bone marrow samples with initial proportions of human cells ranging from 0.09% to 80.5% of leukocytes and a negative control mouse (bottom). FSC versus SSC plots showing the typical positions of mouse and human cells, mostly granulocytes for mouse cells and lymphocytes for human cells, in humanized mice. The graph shows the inverse correlation between the initial percentage of human cells and enrichment. An absence of enrichment is associated with the smallest numbers of cells detected in the gate for human CD45<sup>+</sup> cells, in the range of that found in non-injected control mice (<0.01%).

### Supplemental Figure 7: VCN and lentiviral vector purification

A batch of LTGCPU7 vector was made, and the supernatant was split in two. One half was ultracentrifuged, whereas the other half was subjected to purification by ion exchange chromatography, followed by ultracentrifugation to obtain the same volumes. The two vector preparations were used to transduce CB CD34<sup>+</sup>

cells at three different MOI, with non-purified (MOI=0.18, 1.8, 8.9) and purified (MOI=0.05, 0.5, 2.5) vector preparations, and plated on semi-solid medium. VCN was determined for pooled cells, two weeks later (A). At the highest MOI, transduced cells were left untreated (-) or treated (+) with puromycin for 24 hours, plated on semi-solid medium, and the presence of the integrated vector was assessed in individual colonies (B).

### Supplemental Figure 8: Poisson statistics

Probabilities of vector copy number per cell and probability mass function (A), and expected percentages of transduced cells (B) if vector distribution follows Poisson statistics; expected mean VCN after selection of transduced cells (C) if selection occurs through optimal conditions and if vector distribution before selection follows Poisson statistics.  $\mu$  is the average value of vector copy number per cell;  $i$  is the discrete number of vector copy per cell. The probability mass function is defined only at integer values of  $i$ . The connecting lines are guides for the eye.

### Supplemental Figure 9: The rate of transduction of mature progenitor cells follows a Poisson distribution

Two cord blood CD34<sup>+</sup> cell samples (Experiments #1 and #2) were transduced at several MOIs (a,b,c) and cultured in semi-solid medium. Results from experiment 1 are those summarized Figure 7F, where CD34<sup>+</sup> cells were transduced at two different MOIs (3.7 and 6.2). VCN was assessed by qPCR in >50 individual erythroid colonies per MOI (A1). The expected numbers of cells carrying discrete numbers of vectors (A3) were deduced from the average VCN in erythroid colonies (A2), on the basis of Poisson statistics. Here is an example of calculation: for experiment#2b, where the mean VCN is 0.702 in 57 erythroid colonies, the probability for cells to carry 3 vector copies is:  $[e^{-\mu} \times \mu^i]/i! = [e^{-0.702} \times (0.702)^3]/3! = 0.029$ ; the expected number of erythroid colonies with 3 copies is therefore  $0.029 \times 57 \approx 2$  (A3). The observed numbers of vector-bearing cells were compared with the expected numbers (B). The statistical significance of differences was assessed with Fisher's exact test, for 0, intermediate (1-2), and larger numbers ( $\geq 3$ ) of VCNs (C).

### Supplemental Figure 10: The rate of transduction of SRCs does not follow a Poisson distribution

Two cord blood CD34<sup>+</sup> cell samples (Experiments #1 and #3) were transduced at several MOIs and transplanted into NSG mice. Experiment #1 of this figure is part of experiment #1 in Figure S9; results have been summarized Figure 7D. VCN was assessed by qPCR in  $\geq 299$  individual erythroid colonies derived from mouse bone marrow hCD45<sup>+</sup> cells (A1). The expected numbers of cells bearing discrete numbers of vectors (A3) were deduced from the mean VCN in SRC-derived erythroid progenitors (A2), according to Poisson statistics (A). The expected numbers were compared with the observed numbers of vector-bearing cells (B). The statistical significance of differences was determined in Fisher's exact tests on populations with a low, intermediate or larger VCN (C).

### Supplemental Figure 11: The selection of transduced progenitors counterselects cells at VCN of 1

Two cord blood CD34<sup>+</sup> cell samples (Experiment #1 and #3) were transduced at several MOIs. Experiments #1 and #3 are part of the experiments #1 and #3 depicted in Figures S9 and S10. In experiment #1, transduced cells were selected and cultured on semi-solid medium. In experiment #3, transduced cells were selected and transplanted into NSG mice. The VCNs of CD34<sup>+</sup> cell-derived erythroid progenitors (Experiment #1) and of SRC-derived erythroid progenitors (Experiment #3) were determined for the indicated number of colonies (A1). The theoretical numbers of vector-bearing cells after selection (A2) were calculated if all transduced cells survived and none of the non-transduced cells survived. They are based on the numbers of vector-bearing cells observed before selection (indicated Figure S9A1 and Figure S10A1). Here is an example of calculation: for experiment#3, where the number of colonies carrying 3 copies is 40 and the number of transduced colonies is 285 before selection (Figure S10A1), the theoretical number of colonies carrying 3 copies among 161 (after selection) is  $40/285 \times 161 = 23$  (A2). The distributions of vector-bearing cells were compared with the theoretical distribution (B). The statistical significance of differences was assessed using Fisher's exact test (C).

### Supplemental Figure 12: Integration site analysis *in vitro* and in primary transplants

Percentages of insertion sites within introns and exons, and outside genes (A); percentage of insertion sites upstream and downstream from the TSS of the closest gene (B), percentage of insertion sites upstream and downstream from the closest TSS (C), percentage of insertion sites upstream and downstream from the TSS of the closest oncogene (D). The  $p$ -values given are for the comparison of non-selected and selected conditions using Fisher's exact test (two-tailed). The significance of comparisons between the matched random control (MRC) data set and other groups is indicated as follows: \*  $p < 0.05$ , \*\*  $p < 0.005$ , \*\*\*  $p < 0.0005$ .

### **Supplemental Figure 13: Distribution of integration sites with respect to specific genomic features**

Distribution of unique integration sites with respect to TSS density (**A**), transcript abundance (**B**), DNase I HS density (**C**) and CpG island density (**D**) within a window around integration sites. See the materials and methods section for an explanation of genomic features. The  $p$  value for the comparison between non-selected (NS) and selected (Sel) conditions was obtained in a Mann-Whitney U test. All comparisons between the matched random control (MRC) data set and other groups yielded a  $p$ -value  $<0.0001$  (not shown).

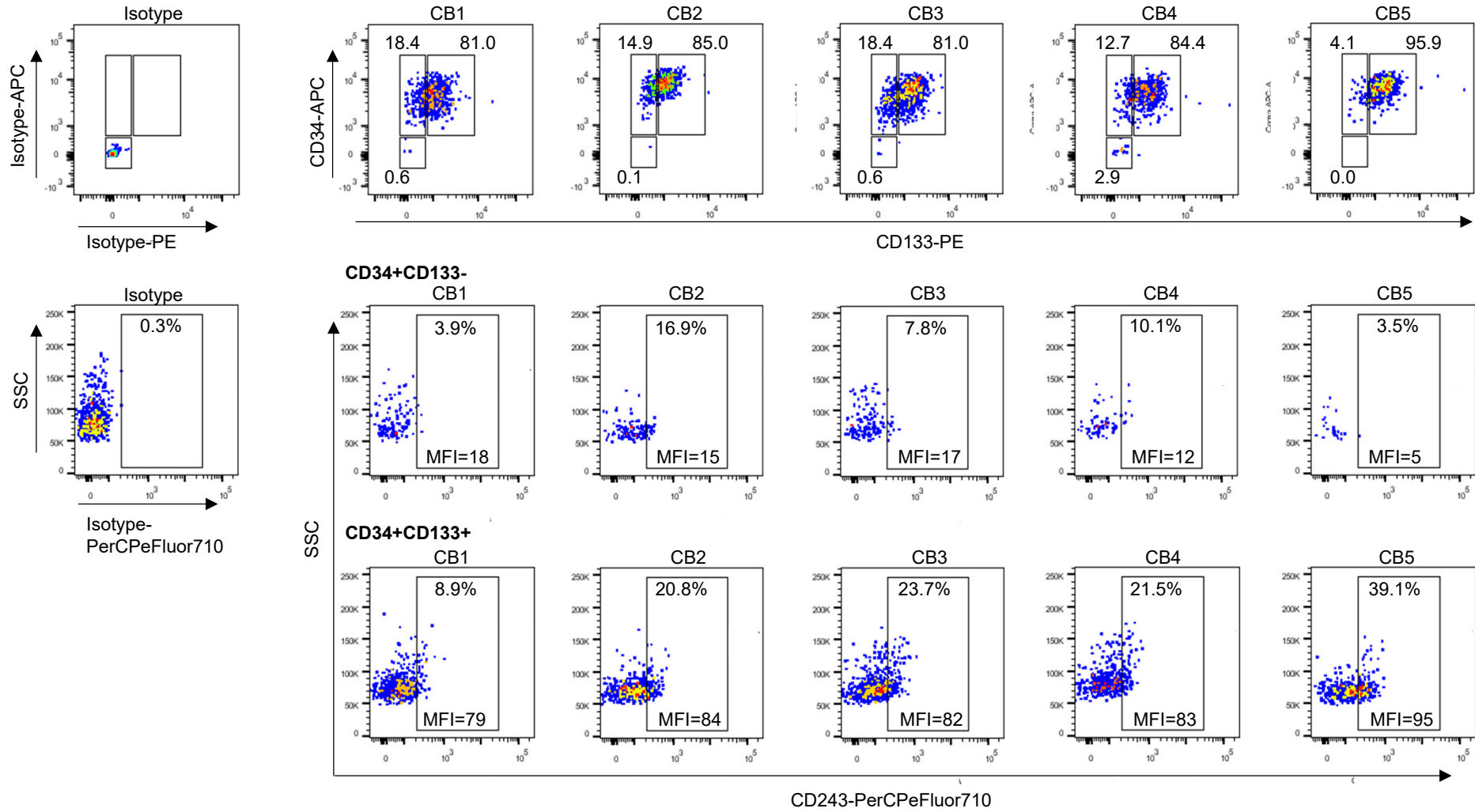
### **Supplemental Figure 14: Insertion site distribution biases**

Distortion of the proportion or distribution of unique integration sites with respect to TSS position (**A**), oncogene TSS position (**B**), position in the gene (**C**), transcript abundance (**D**), TSS density (**E**), CpG island density (**F**), and DNase I HS density (**G**). The  $p$  values were calculated as indicated in **Figures S12 and S13**.

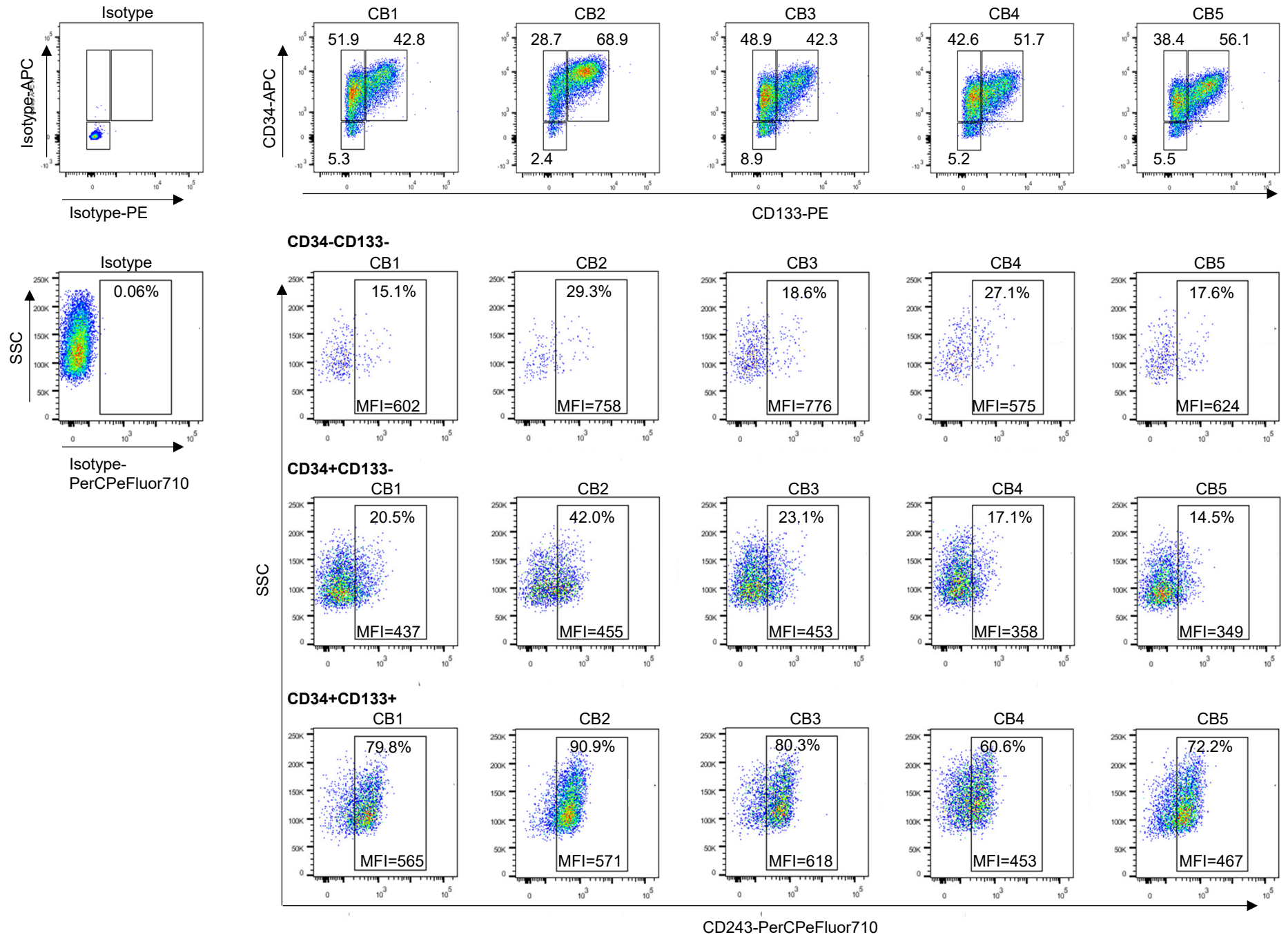
# Supplemental Figure 1

ggctccggtgcccggtcagtgggcagagcgcacatcgcccacagtccccgagaagtggggggg  
aggggtcggcaattgaaccggtgcctagagaagggtggcgcggggtaaaactgggaaagtgatg  
tcgtgtactggctccgcctttttcccaggggtgggggagaaccgtatataagtgcagtagtc  
gccgtgaacgttctttttcgcaacggggttgccgcccagaacacaggctagccaccatgacgg  
agtataagccgactgtcaggctggccacaagggacgatgtgccaggggccggtcagaactctg  
gcagcagctttcgcagactaccgggcgaccagacatacgggtggaccctgacagacatattga  
gcgcgtgacggagttgcaagaacttttcttgaccggagtcggactcgacatcgggaagggtg  
gggtggcggatgatggagcagccgtcgcgggtatggacaacgccgaatcagtcgaggctgga  
gcggtctttgcggaatcggccctagaatggcggaaactttccgggagccggctggcggcaca  
gcagcagatggagggactgttggcacctcaccgcccgaagagccggcctggttcctggcga  
cagtcggtgtatcaccgaccaccaggggaaagggtctgggttccgcctggtggtgcttggga  
gtggaggcagccgagagagcaggggtcccagcgtttttggaaaccagcgcgccacggaacct  
ccattctacgagcggctcggattcaccgtaacggccgacgtagaagtaccggaaggacctc  
ggacttgggtgcatgaccagaaagcccgggtgcggggtcgatgccactcttttgcgggatac  
attgatgggcccgcattggaatgggaaagaccacaacaacgcagctccttgtcgcgctgggtag  
ccgggacgacattgtctatgtgcccgagccgatgacttactggagggtattgggtgagcgcg  
aaacgattgccaacatctacactacacagcaccgcttggatcagggcgagatttcagcgggt  
gatgccgcagtggtgatgacgtcggcgcagatcacaatgggaatgccgtatgagggtgacgga  
tgaggacttgcgctcacatcgggggagaggcgggctcctcgcattgccctccccagcac  
tcacgctcatctttgatcgccatccaatcgccgctctcctttgctaccgggtgcccgatac  
ctgatgggttcgatgacgccccaaagctgtattggcgtttgctcgcgcttgatcccgccacgct  
gccagggaccaatatcgtcttgggggcatgcccgaggataggcacattgacagactggcca  
aaaggcagaggcccggagagcgttggacttggccatgctcgcggccattagacgggtgat  
ggattgctcgcaaacggtgagctatctgcagtggtggatcatggcgcgaggactgggg  
tcagctttcgggactgcagtacccccgcaagggtgcagaaccccagagcaacgcgggacccc  
gaccacacatcggagacacgctgttcacattgttttagagccccgaacttctcgcgccta  
ggggacctctacaacgtattcgcattggcgcttgatgtcttggcgaaaaggcttaggtcgat  
gcacgtgttcatcttggattatgaccagtcgcccggcaggatgcagggatgctctgctgcaac  
tgacgtcggggatggtgcagactcatgtgaccacccccggatcaattcctactatctgtgat  
ctggcgcgcacctttgcgcgagaaatgggggaagccaattga

# Supplemental Figure 2A (Day1)

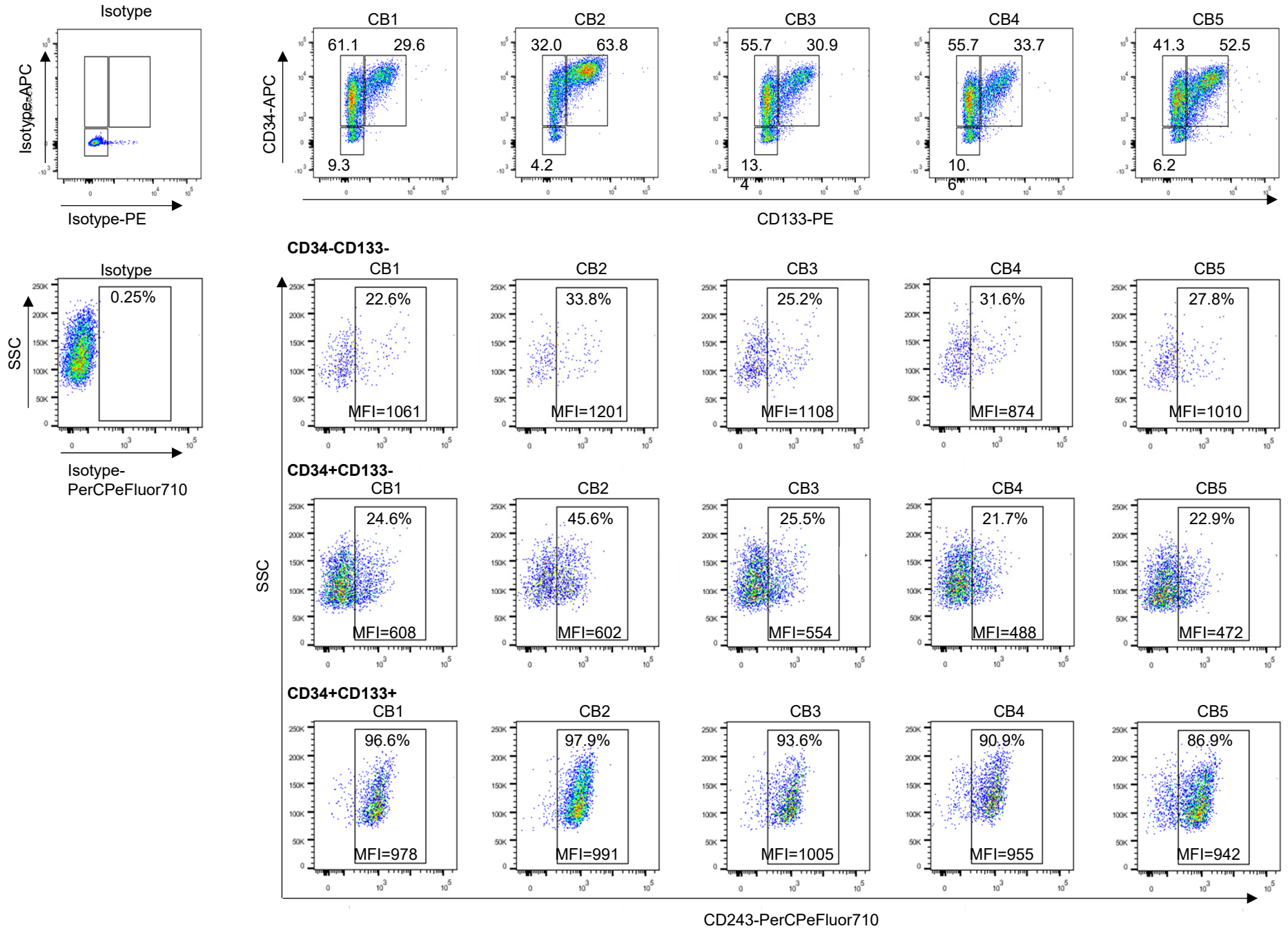


# Supplemental Figure 2B (Day3)

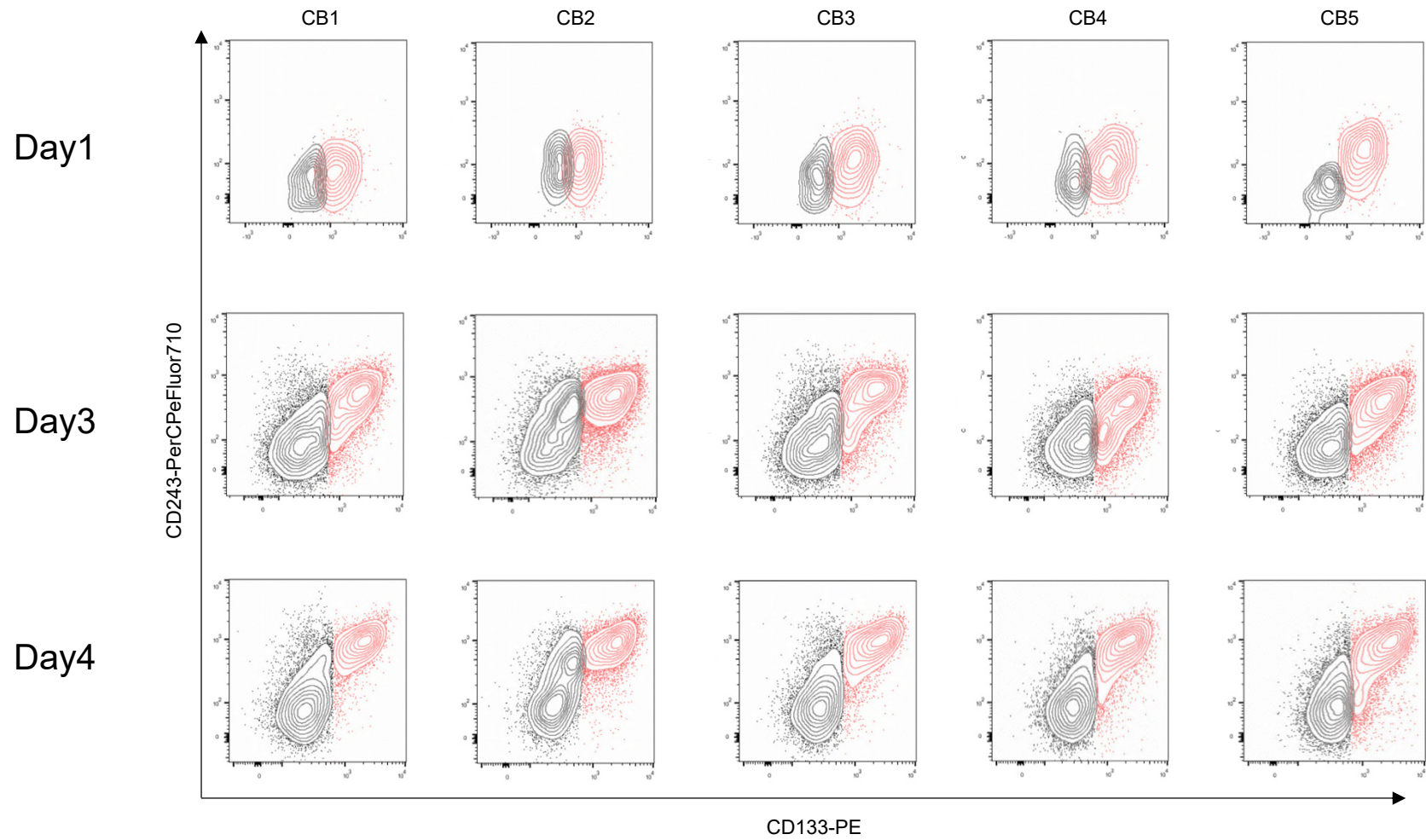




# Supplemental Figure 2C (Day4)

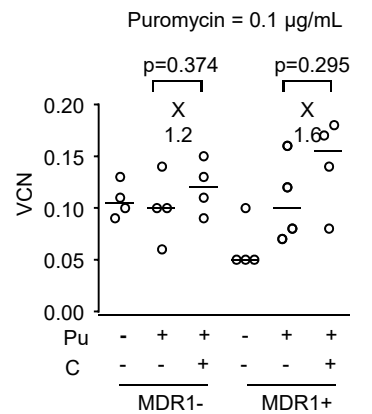
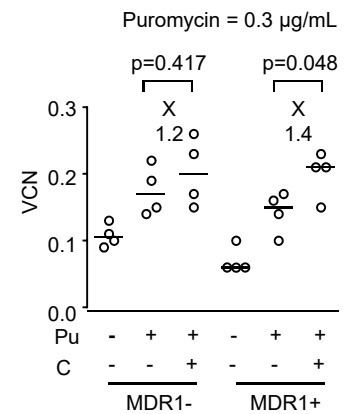
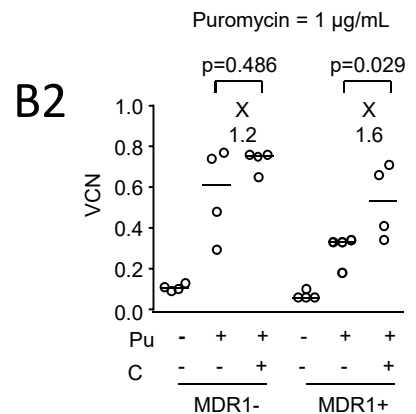
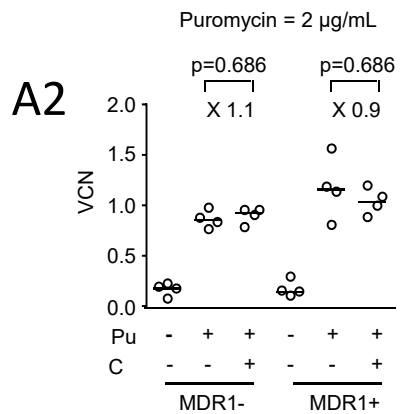
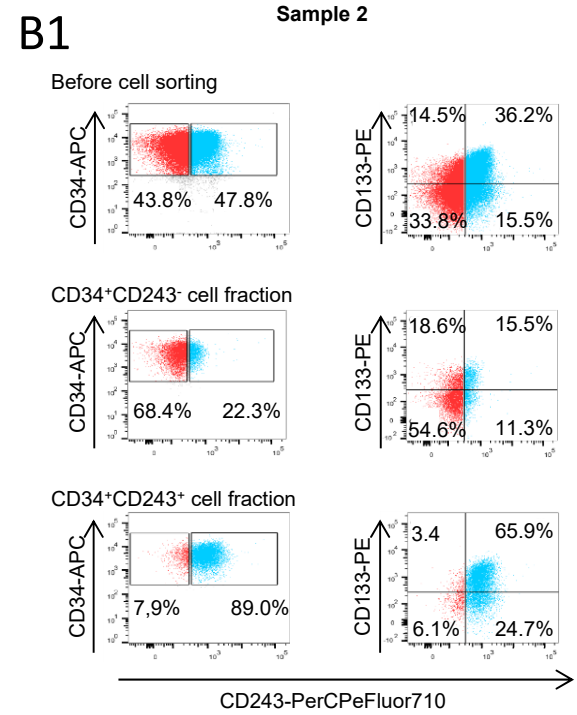
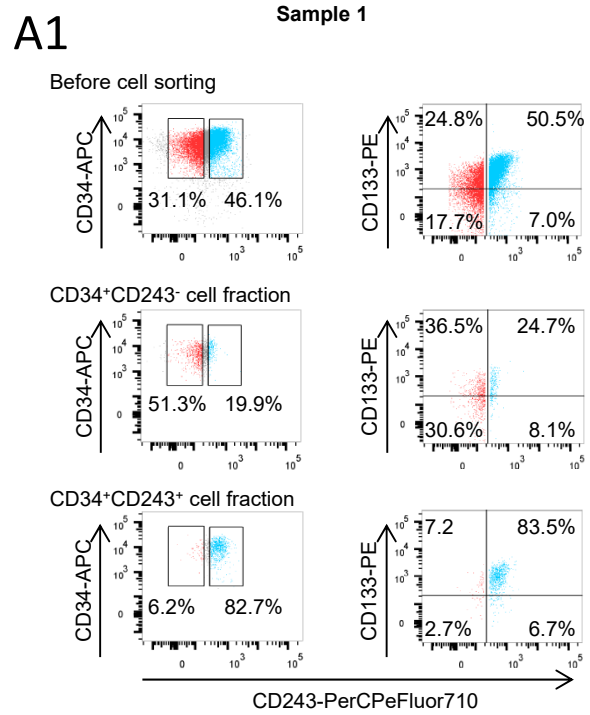


# Supplemental Figure 2D (Density plots on CD34+ cells)

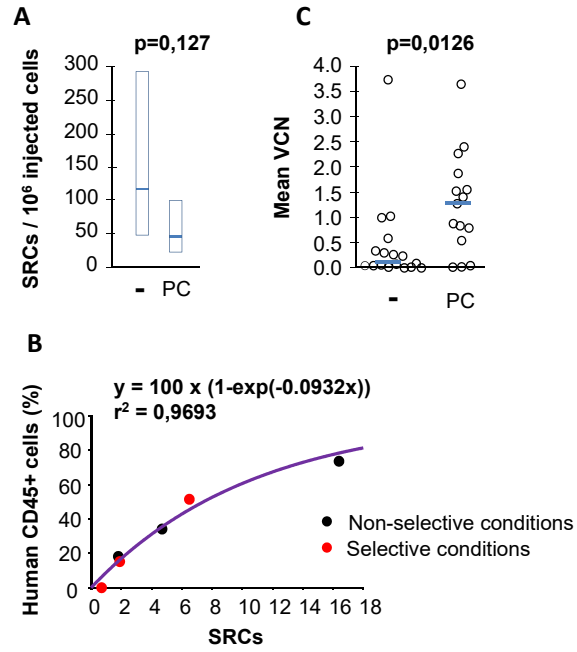


■ CD34<sup>+</sup>CD133<sup>-</sup>  
■ CD34<sup>+</sup>CD133<sup>+</sup>

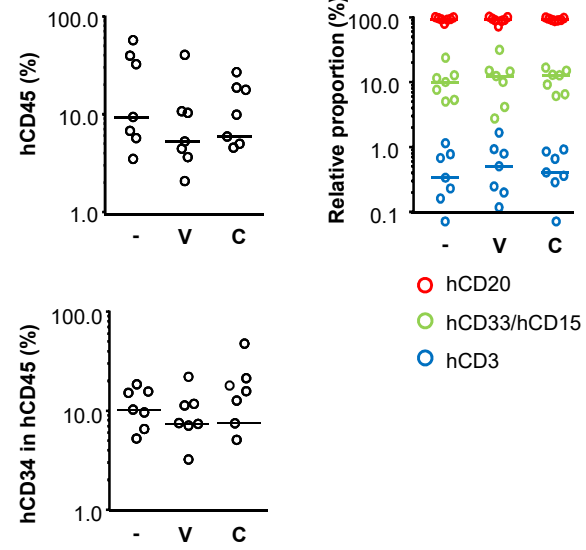
# Supplemental Figure 3



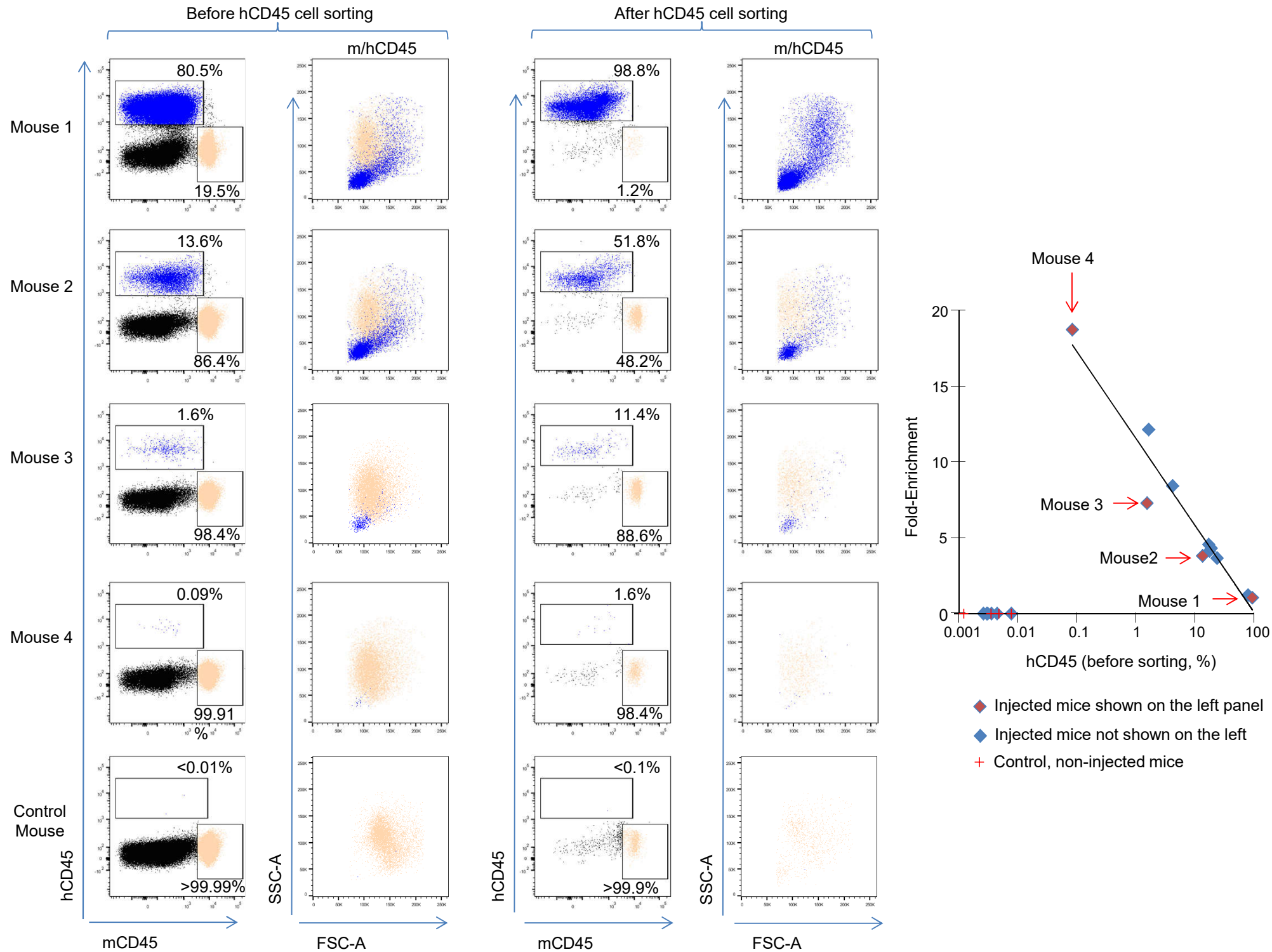
# Supplemental Figure 4



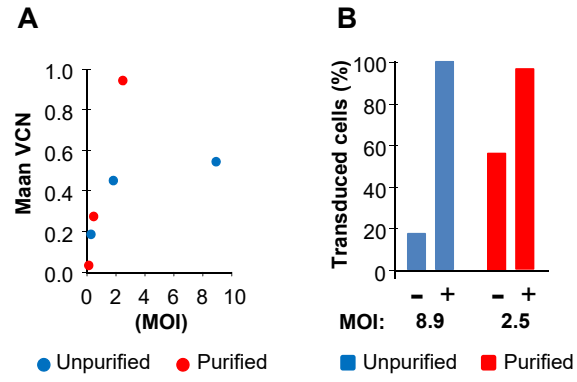
# Supplemental Figure 5



# Supplemental Figure 6



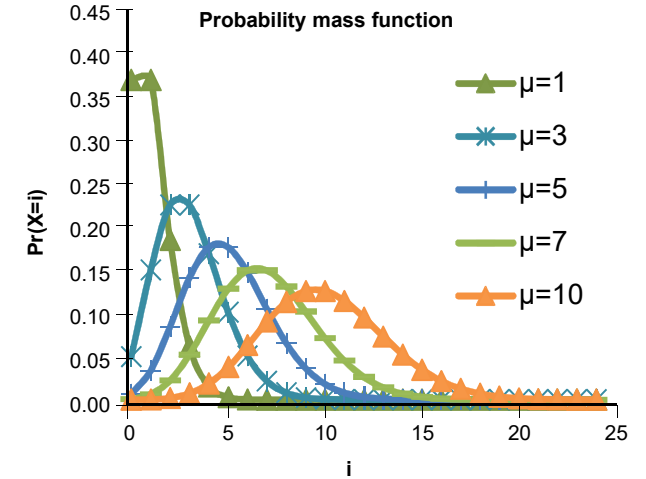
# Supplemental Figure 7



# Supplemental Figure 8

**A**  $\Pr(X=i) = (e^{-\mu}) * (\mu^i) / i!$  ; range of average VCN values: 0.1 - 10

		$\mu$											
		0,1	0,5	1	2	3	4	5	6	7	8	9	10
i	0	9,05E-01	6,07E-01	3,68E-01	1,35E-01	4,98E-02	1,83E-02	6,74E-03	2,48E-03	9,12E-04	3,35E-04	1,23E-04	4,54E-05
	1	9,05E-02	3,03E-01	3,68E-01	2,71E-01	1,49E-01	7,33E-02	3,37E-02	1,49E-02	6,38E-03	2,68E-03	1,11E-03	4,54E-04
	2	4,52E-03	7,58E-02	1,84E-01	2,71E-01	2,24E-01	1,47E-01	8,42E-02	4,46E-02	2,23E-02	1,07E-02	5,00E-03	2,27E-03
	3	1,51E-04	1,26E-02	6,13E-02	1,80E-01	2,24E-01	1,95E-01	1,40E-01	8,92E-02	5,21E-02	2,86E-02	1,50E-02	7,57E-03
	4	3,77E-06	1,58E-03	1,53E-02	9,02E-02	1,68E-01	1,95E-01	1,75E-01	1,34E-01	9,12E-02	5,73E-02	3,37E-02	1,89E-02
	5	7,54E-08	1,58E-04	3,07E-03	3,61E-02	1,01E-01	1,56E-01	1,75E-01	1,61E-01	1,28E-01	9,16E-02	6,07E-02	3,78E-02
	6	1,26E-09	1,32E-05	5,11E-04	1,20E-02	5,04E-02	1,04E-01	1,46E-01	1,61E-01	1,49E-01	1,22E-01	9,11E-02	6,31E-02
	7	1,80E-11	9,40E-07	7,30E-05	3,44E-03	2,16E-02	5,95E-02	1,04E-01	1,38E-01	1,49E-01	1,40E-01	1,17E-01	9,01E-02
	8	2,24E-13	5,88E-08	9,12E-06	8,59E-04	8,10E-03	2,98E-02	6,53E-02	1,03E-01	1,30E-01	1,40E-01	1,32E-01	1,13E-01
	9	2,49E-15	3,26E-09	1,01E-06	1,91E-04	2,70E-03	1,32E-02	3,63E-02	6,88E-02	1,01E-01	1,24E-01	1,32E-01	1,25E-01
	10	2,49E-17	1,63E-10	1,01E-07	3,82E-05	8,10E-04	5,29E-03	1,81E-02	4,13E-02	7,10E-02	9,93E-02	1,19E-01	1,25E-01
	11	2,27E-19	7,42E-12	9,22E-09	6,94E-06	2,21E-04	1,92E-03	8,24E-03	2,25E-02	4,52E-02	7,22E-02	9,70E-02	1,14E-01
	12	1,89E-21	3,09E-13	7,68E-10	1,16E-06	5,52E-05	6,42E-04	3,43E-03	1,13E-02	2,63E-02	4,81E-02	7,28E-02	9,48E-02
	13	1,45E-23	1,19E-14	5,91E-11	1,78E-07	1,27E-05	1,97E-04	1,32E-03	5,20E-03	1,42E-02	2,96E-02	5,04E-02	7,29E-02
	14	1,04E-25	4,25E-16	4,22E-12	2,54E-08	2,73E-06	5,64E-05	4,72E-04	2,23E-03	7,09E-03	1,69E-02	3,24E-02	5,21E-02
	15	6,92E-28	1,42E-17	2,81E-13	3,39E-09	5,46E-07	1,50E-05	1,57E-04	8,91E-04	3,31E-03	9,03E-03	1,94E-02	3,47E-02
	16	4,32E-30	4,42E-19	1,76E-14	4,24E-10	1,02E-07	3,76E-06	4,91E-05	3,34E-04	1,45E-03	4,51E-03	1,09E-02	2,17E-02
	17	2,54E-32	1,30E-20	1,03E-15	4,99E-11	1,81E-08	8,85E-07	1,45E-05	1,18E-04	5,96E-04	2,12E-03	5,79E-03	1,28E-02
	18	1,41E-34	3,61E-22	5,75E-17	5,54E-12	3,01E-09	1,97E-07	4,01E-06	3,93E-05	2,32E-04	9,44E-04	2,89E-03	7,09E-03
	19	7,44E-37	9,51E-24	3,02E-18	5,83E-13	4,76E-10	4,14E-08	1,06E-06	1,24E-05	8,54E-05	3,97E-04	1,37E-03	3,73E-03
	20	3,72E-39	2,38E-25	1,51E-19	5,83E-14	7,14E-11	8,28E-09	2,64E-07	3,73E-06	2,99E-05	1,59E-04	6,17E-04	1,87E-03
	21	1,77E-41	5,66E-27	7,20E-21	5,56E-15	1,02E-11	1,58E-09	6,29E-08	1,06E-06	9,97E-06	6,06E-05	2,64E-04	8,89E-04
	22	8,05E-44	1,29E-28	3,27E-22	5,05E-16	1,39E-12	2,87E-10	1,43E-08	2,90E-07	3,17E-06	2,20E-05	1,08E-04	4,04E-04
	23	3,50E-46	2,80E-30	1,42E-23	4,39E-17	1,81E-13	4,99E-11	3,11E-09	7,57E-08	9,65E-07	7,66E-06	4,23E-05	1,76E-04
	24	1,46E-48	5,83E-32	5,93E-25	3,66E-18	2,27E-14	8,31E-12	6,47E-10	1,89E-08	2,82E-07	2,55E-06	1,59E-05	7,32E-05



The function is defined only at integer values of  $i$ . The connecting lines are guides for the eye.

**B**  $\text{Exp}\%Tdcells = (1 - e^{-\mu}) * 100$ ; range of average VCN values: 0.1 - 10

$\mu$	0,1	0,5	1	2	3	4	5	6	7	8	9	10
Exp%Tdcells	9,5	39,3	63,2	86,5	95,0	98,2	99,3	99,8	99,9	100,0	100,0	100,0

=> When  $\mu \geq 5$ , Exp%Tdcells  $\approx$  100 %

**C**  $\text{ExpVCNsel} = S(\Pr(X=i)*i) / (1 - \Pr(X=0))$  ; range of average VCN values (before selection): 0.1 - 10

$\mu$	0,1	0,5	1	2	3	4	5	6	7	8	9	10
ExpVCNsel	1,1	1,3	1,6	2,3	3,2	4,1	5,0	6,0	7,0	8,0	9,0	10,0

=> When  $\mu \geq 5$ , ExpVCNsel  $\approx$   $\mu$



# Supplemental Figure 9

**A1** Observed numbers of non-selected erythroid progenitors with variable integration events (in vitro)

Expt:	1a(n=53)	1b(n=53)	2a(n=51)	2b(n=57)	2c(n=52)
copies	Observed colony numbers				
0	37	26	32	27	11
1	13	22	14	25	25
2	3	4	5	4	12
3	0	1	0	0	4
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	1	0

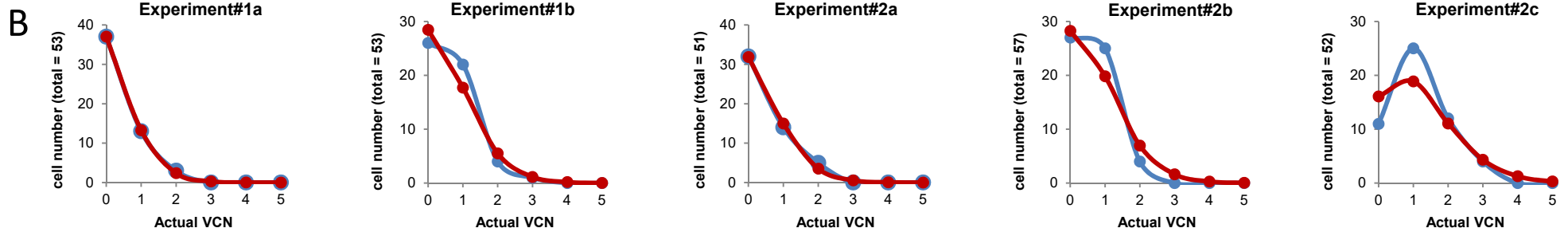
**A2** Average VCN in erythroid progenitors

1a	1b	2a	2b	2c
0.358	0.623	0.471	0.702	1.173

● Observed numbers  
● Expected numbers

**A3** Expected numbers of non-selected erythroid progenitors according to the average VCN and Poisson statistic

Expt:	1a(n=53)	1b(n=53)	2a(n=51)	2b(n=57)	2c(n=52)
Copies	Expected colony number				
0	37	28	32	28	16
1	13	18	15	20	19
2	2	6	4	7	11
3	0	1	1	2	4
4	0	0	0	0	1
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0



**C**

Experiment#1a					Experiment#1b					Experiment#2a					Experiment#2b					Experiment#2c				
VCN	Obs	Exp	Tot	p-value (2-sided)	VCN	Obs	Exp	Tot	p-value (2-sided)	VCN	Obs	Exp	Tot	p-value (2-sided)	VCN	Obs	Exp	Tot	p-value (2-sided)	VCN	Obs	Exp	Tot	p-value (2-sided)
0	37	37	74		0	26	28	54		0	32	32	64		0	27	28	55		0	11	16	27	
not 0	16	16	32	1.00000	not 0	27	25	52	0.84607	not 0	19	19	38	1.00000	not 0	30	29	59	1.00000	not 0	41	36	77	0.37122
Total	53	53	106		Total	53	53	106		Total	51	51	102		Total	57	57	114		Total	52	52	104	
1	13	13	26		1	22	18	40		1	14	15	29		1	25	20	45		1	25	19	44	
not 1	40	40	80	1.00000	not 1	31	35	66	0.54803	not 1	37	36	73	1.00000	not 1	32	37	69	0.44360	not 1	27	33	60	0.32104
Total	53	53	106		Total	53	53	106		Total	51	51	102		Total	57	57	114		Total	52	52	104	
2	0	2	2		2	4	6	10		2	5	4	9		2	4	7	11		2	12	11	23	
not 2	53	51	104	1.00000	not 2	49	47	96	0.74147	not 2	46	47	93	1.00000	not 2	53	50	103	0.52773	not 2	40	41	81	1.00000
Total	53	53	106		Total	53	53	106		Total	51	51	102		Total	57	57	114		Total	52	52	104	
1 or 2	16	16	32		1 or 2	26	23	49		1 or 2	19	19	38		1 or 2	29	27	56		1 or 2	37	30	67	
not (1 or 2)	37	37	74	1.00000	not (1 or 2)	27	30	57	0.69703	not (1 or 2)	32	32	64	1.00000	not (1 or 2)	28	30	58	0.85150	not (1 or 2)	15	22	37	0.21887
Total	53	53	106		Total	53	53	106		Total	51	51	102		Total	57	57	114		Total	52	52	104	
3	0	0	0		3	1	1	2		3	0	1	1		3	0	2	2		3	4	4	8	
not 3	53	53	106	1.00000	not 3	52	52	104	1.00000	not 3	51	50	101	1.00000	not 3	57	55	112	0.49557	not 3	48	48	96	1.00000
Total	53	53	106		total	53	53	106		Total	51	51	102		total	57	57	114		total	52	52	104	
4	0	0	0		4	0	0	0		4	0	0	0		4	0	0	0		4	0	1	1	
not 4	53	53	106	1.00000	not 4	53	53	106	1.00000	not 4	51	51	102	1.00000	not 4	57	57	114	1.00000	not 4	52	51	103	1.00000
Total	53	53	106		total	53	53	106		Total	51	51	102		total	57	57	114		total	52	52	104	
>=3	0	0	0		>=3	1	1	2		>=3	0	1	1		>=3	1	2	3		>=3	4	6	10	
not (>=3)	53	53	106	1.00000	not (>=3)	52	52	104	1.00000	not (>=3)	51	50	101	1.00000	not (>=3)	56	55	111	1.00000	not (>=3)	48	46	94	0.74121
Total	53	53	106		Total	53	53	106		Total	51	51	102		Total	57	57	114		Total	52	52	104	

# Supplemental Figure 10

**A1** Observed numbers of non-selected SRCs with variable integration events

Expt:	1a (n=355)	1b (n=350)	3 (n=299)
copies			
0	201	144	13
1	94	85	66
2	22	64	86
3	22	42	40
4	10	13	29
5	3	2	17
6	0	0	19
7	0	0	9
8	0	0	9
9	1	0	9
10	1	0	0
11	0	0	1
12	1	0	0
13	0	0	1

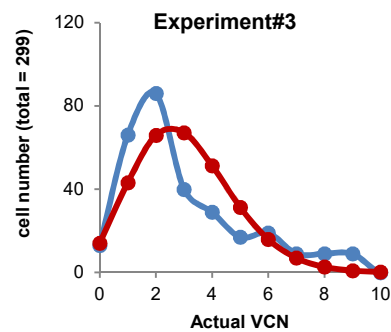
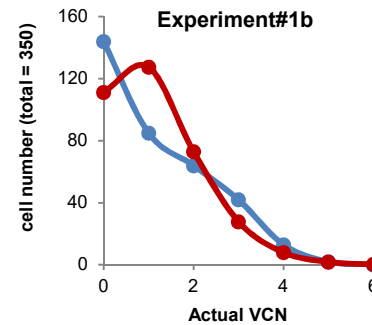
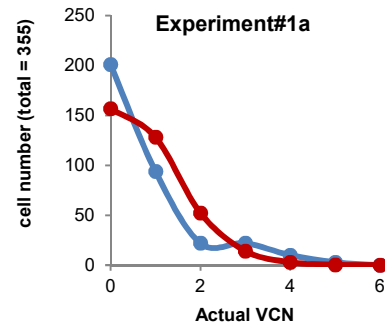
**A2** Average VCN in erythroid progenitors

1a	1b	3
0.817	1.146	3.050

**A3** Expected numbers of non-selected SRCs according to the average VCN and Poisson statistic

Expt:	1a (n=355)	1b (n=350)	3 (n=299)
copies			
0	157	111	14
1	128	128	43
2	52	73	66
3	14	28	67
4	3	8	51
5	0	2	31
6	0	0	16
7	0	0	7
8	0	0	3
9	0	0	1
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0

**B**



● Observed numbers  
● Expected numbers

**C**

Experiment#1a						
VCN	Obs	Exp	Tot	p-value (2-sided)	Odds ratio	95%
0	201	157	358			
not 0	154	198	352	0.00123	1.65	1.22 - 2.21
Total	355	355	710			
1-2	116	180	296			
Not (1-2)	239	175	414	< 0.00001	0.47	0.35 - 0.64
Total	355	355	710			
≥ 3	38	18	56			
not (≥ 3)	317	337	654	0.00767	2.24	1.25 - 4.01
Total	355	355	710			

Experiment#1b						
VCN	Obs	Exp	Tot	p-value (2-sided)	Odds ratio	95%
0	144	111	255			
not 0	206	239	445	0.00762	1.54	1.13 - 2.10
Total	350	350	700			
1-2	149	201	350			
Not (1-2)	201	149	350	0.00011	0.55	0.41 - 0.74
Total	350	350	700			
≥ 3	57	38	95			
not (≥ 3)	293	312	605	0.04655	1.60	1.03 - 2.48
Total	350	350	700			

Experiment#3						
VCN	Obs	Exp	Tot	p-value (2-sided)	Odds ratio	95%
0-2	165	124	289			
not (0-2)	134	175	310	0.00078	1.76	1.27 - 2.44
Total	299	299	598			
3-5	86	150	236			
Not (3-5)	214	149	362	< 0.00001	0.40	0.36 - 0.59
Total	299	299	598			
≥ 6	48	27	75			
not (≥ 6)	251	272	523	0.01312	1.92	1.17 - 3.18
Total	299	299	598			

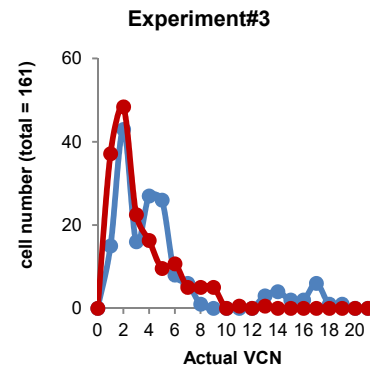
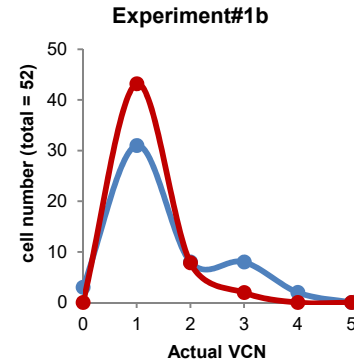
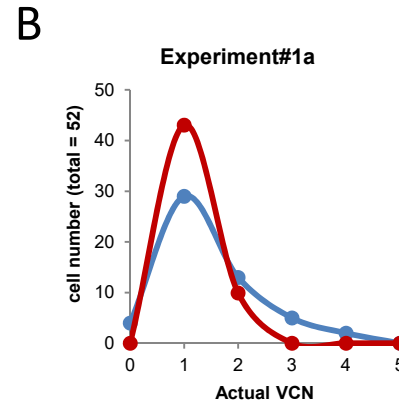
# Supplemental Figure 11

**A1** Observed numbers of selected cells with variable integration events

	Erythroid progenitors (in vitro)		SRCs
	Expt: 1a (n=53)	1b (n=53)	3 (n=161)
copies			
0	4	3	0
1	29	31	15
2	13	8	43
3	5	8	16
4	2	2	27
5	0	0	26
6	0	0	8
7	0	0	6
8	0	0	1
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	3
14	0	0	4
15	0	0	2
16	0	0	2
17	0	0	6
18	0	0	1
19	0	0	1

**A2** Expected numbers of selected cells with variable integration events, according to proportions determined in non-selected cells

	Erythroid progenitors (in vitro)		SRCs
	Expt: 1a (n=53)	1b (n=53)	3(n=161)
copies			
0	0	0	0
1	43	43	37
2	10	8	48
3	0	2	23
4	0	0	16
5	0	0	10
6	0	0	11
7	0	0	5
8	0	0	5
9	0	0	5
10	0	0	0
11	0	0	1
12	0	0	0
13	0	0	1
14	0	0	0
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0



● Observed numbers  
● Expected numbers

**C** Experiment#1a

VCN	Obs	Exp	Tot	p-value (2-sided)	Odds ratio	95%
1	29	43	27			
not 1	24	10	77	0.00638	0.28	0.12-0.67
Total	53	53	104			
2	13	10	27			
Not 2	40	43	77	0.63810		
Total	53	53	104			
≥ 3	7	0	10			
not (≥ 3)	46	53	94	0.01265	17.26	0.96-310.6
Total	53	53	104			

Experiment#1b

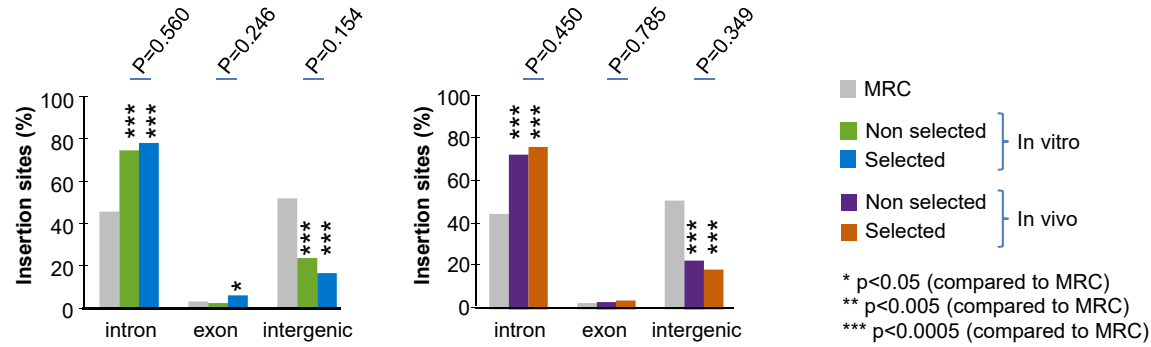
VCN	Obs	Exp	Tot	p-value (2-sided)	Odds ratio	95%
1	31	43	27			
not 1	22	10	77	0.01919	0.33	0.14-0.79
Total	53	53	104			
2	8	10	27			
Not 2	45	43	77	0.79654		
Total	53	53	104			
≥ 3	10	2	10			
not (≥ 3)	43	51	94	0.02833	5.93	1.23-28.56
Total	53	53	104			

Experiment#3

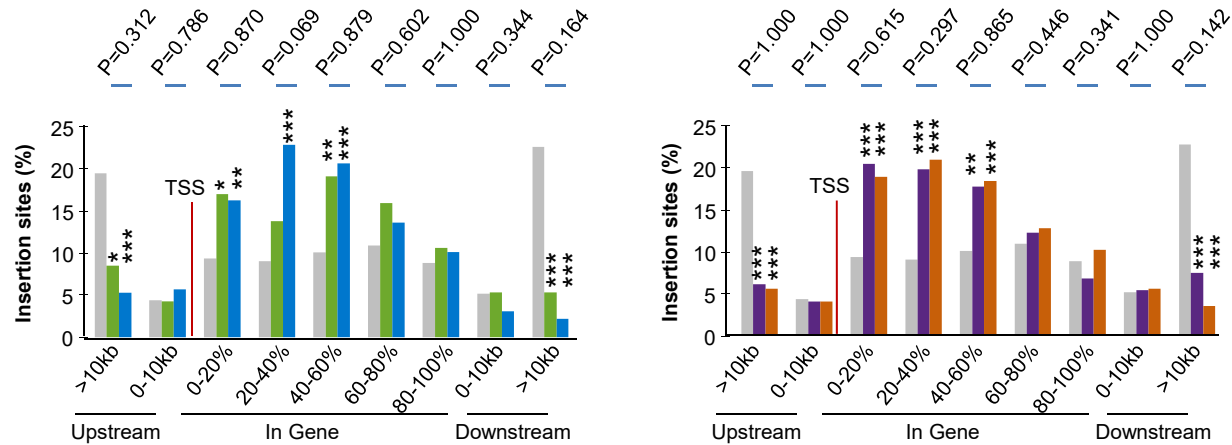
VCN	Obs	Exp	Tot	p-value (2-sided)	Odds ratio	95%
1	15	37	52			
not 1	146	124	270	0.00130	0.34	0.18 - 0.66
Total	161	161	322			
2	43	48	91			
Not 2	118	113	231	0.60691		
Total	161	161	322			
≥ 3	103	75	178			
not (≥ 3)	58	86	144	0.02422	2.04	1.30 -3.18
Total	161	161	322			

# Supplemental Figure 12

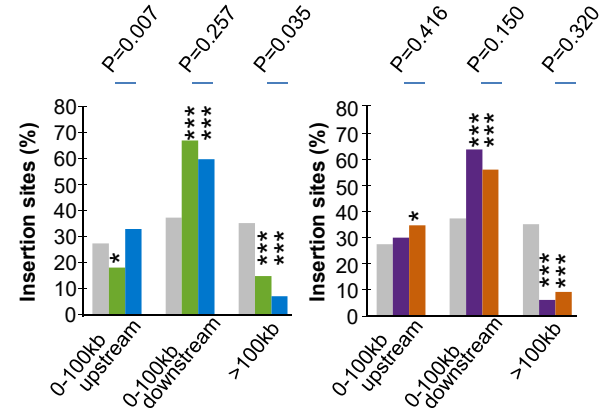
## A Position relative to genes



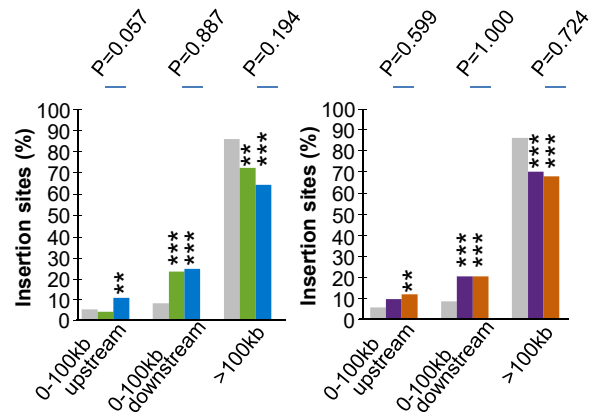
## B Distance to the closest gene



## C Distance to the closest TSS

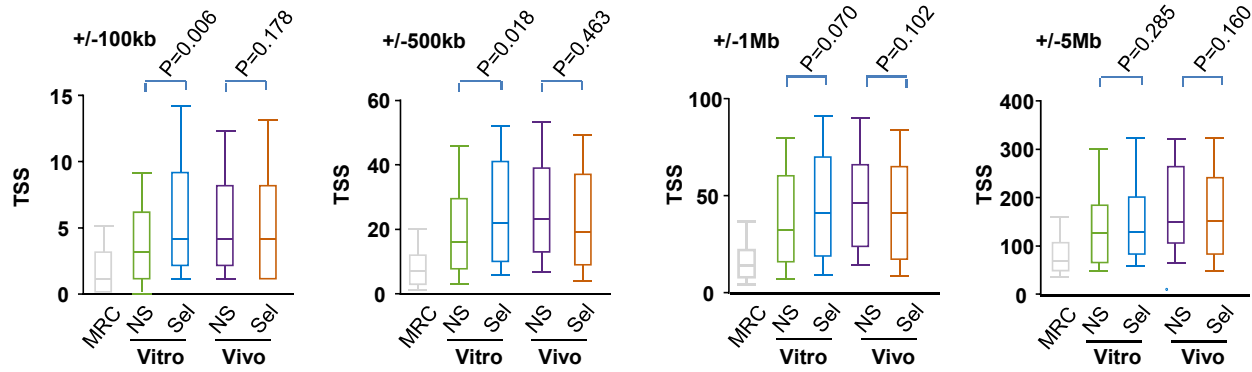


## D Distance to the closest TSS (oncogenes)

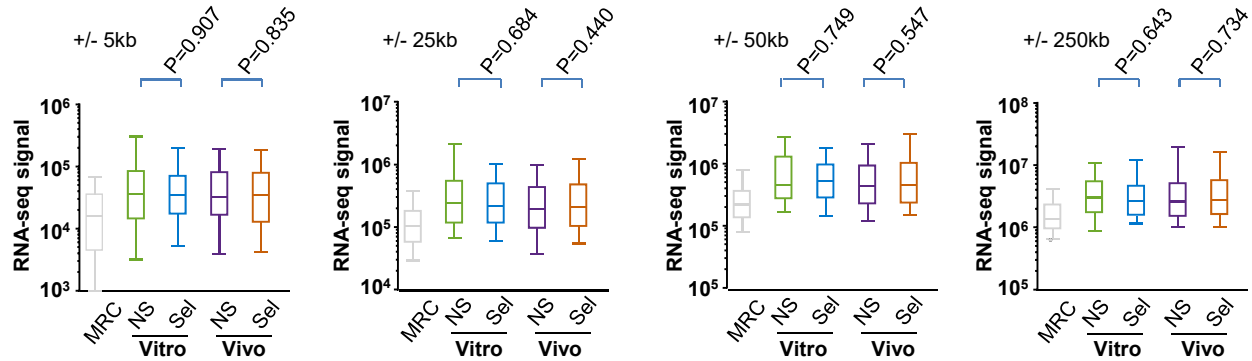


# Supplemental Figure 13

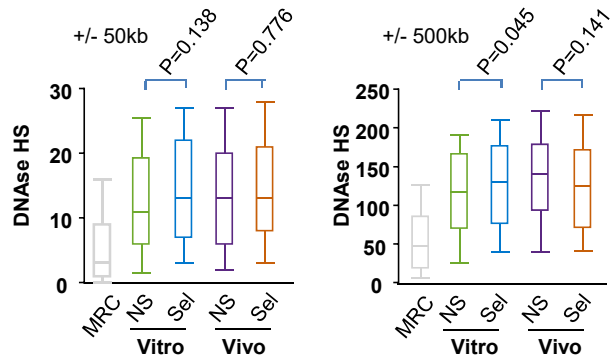
## A TSS density



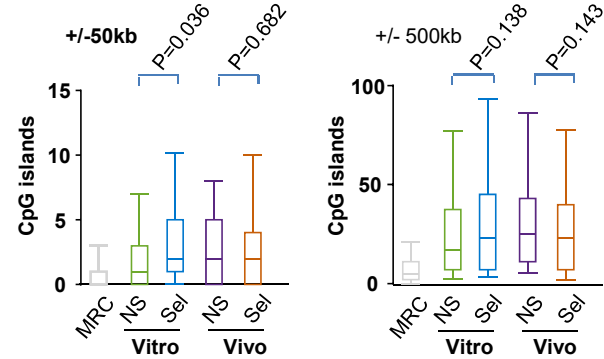
## B Transcript abundance



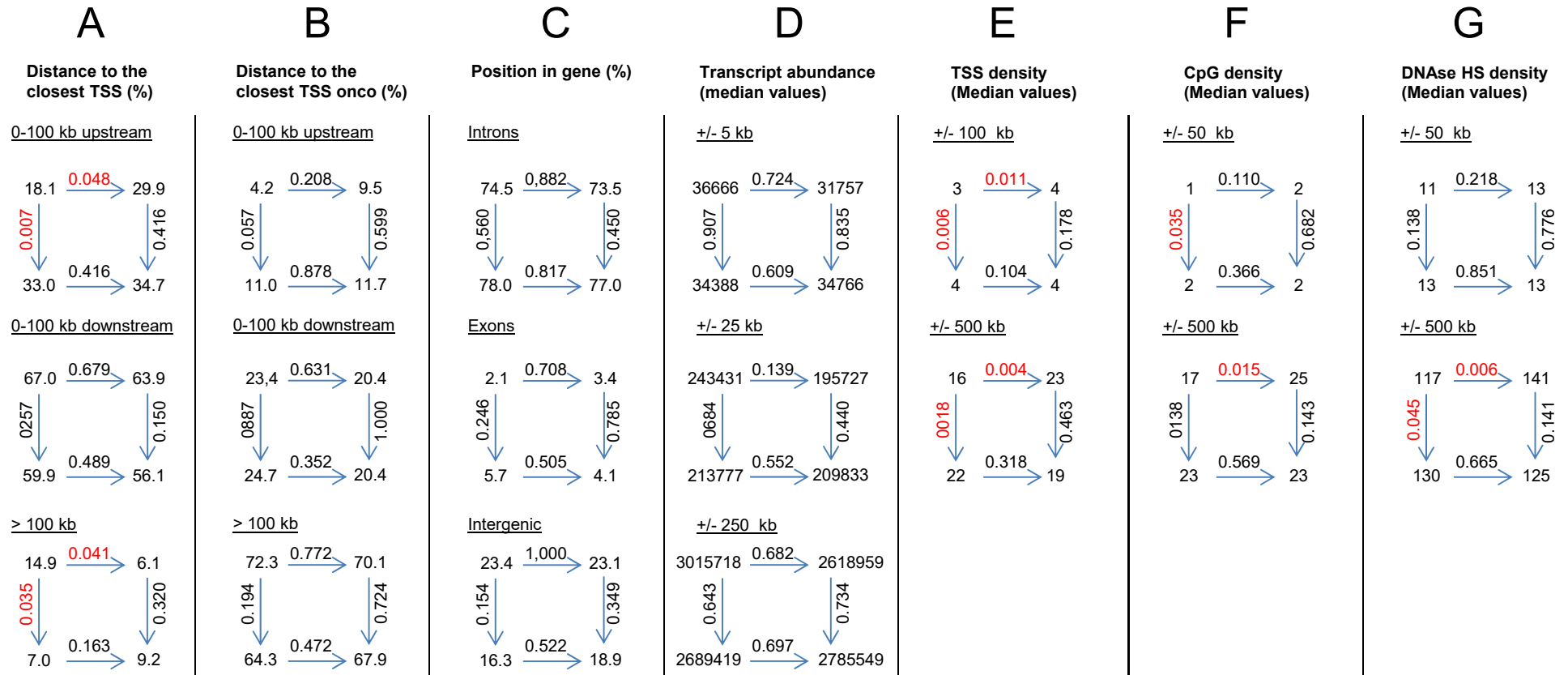
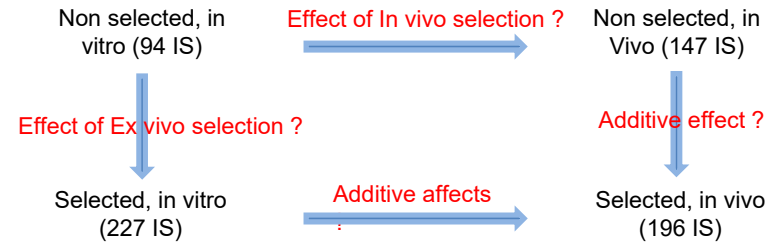
## C DNase HS density



## D CpG Islands density



# Supplemental Figure 14



**SUPPLEMENTAL TABLE 1**

Position relative to genes; p-values: Fisher exact test (2-sided)

	Samples vs. Matched Random Control				Selected vs. non selected		Vitro vs. vivo	
	Vitro		Vivo		Vitro	Vivo	Non selected	Selected
	Non selected	Selected	Non selected	Selected				
Intron	<b>2.2x10<sup>-8</sup></b>	<b>4.6x10<sup>-22</sup></b>	<b>2.5x10<sup>-11</sup></b>	<b>2.4x10<sup>-18</sup></b>	0.560	0.450	0.882	0.817
Exon	1.000	<b>0.025</b>	0.614	0.614	0.246	0.785	0.708	0.505
Intergenic	<b>4.5x10<sup>-8</sup></b>	<b>8.2x10<sup>-25</sup></b>	<b>4.9x10<sup>-12</sup></b>	<b>3.5x10<sup>-20</sup></b>	0.154	0.349	1.000	0.522

**Bold font: p<0.05**

Distances to the closest gene; p-values: Fisher exact test (2-sided)

	Samples vs. Matched Random Control				Selected vs. non selected		Vitro vs. vivo	
	Vitro		Vivo		Vitro	Vivo	Non selected	Selected
	Non selected	Selected	Non selected	Selected				
In the closest gene	<b>4.5x10<sup>-8</sup></b>	<b>8.2x10<sup>-27</sup></b>	<b>4.9x10<sup>-12</sup></b>	<b>3.5x10<sup>-20</sup></b>	0.154	0.349	1.000	0.522
0-10kb upstream the closest gene	1.000	0.318	1.000	1.000	0.786	1.000	1.000	0.505
>10kb upstream the closest gene	<b>5.1x10<sup>-3</sup></b>	<b>2.7 x10<sup>-9</sup></b>	<b>1.2x10<sup>-5</sup></b>	<b>7.5x10<sup>-8</sup></b>	0.312	1.000	0.607	1.000
0-10kb downstream the closest gene	0.815	0.206	0.848	0.740	0.344	1.000	1.000	0.232
>10kb downstream the closest gene	<b>1.2x10<sup>-5</sup></b>	<b>5.2x10<sup>-18</sup></b>	<b>3.0x10<sup>-6</sup></b>	<b>3.0x10<sup>-13</sup></b>	0.164	0.142	0.603	0.559

**Bold font: p<0.05**

Distances to the closest transcription start site; p-values: Fisher exact test (2-sided)

	Samples vs. Matched Random Control				Selected vs. non selected		Vitro vs. vivo	
	Vitro		Vivo		Vitro	Vivo	Non selected	Selected
	Non selected	Selected	Non selected	Selected				
0-100kb upstream the closest TSS	<b>0.045</b>	0.077	0.509	<b>0.032</b>	<b>0.007</b>	0.416	<b>0.048</b>	0.758
0-100kb downstream the closest TSS	<b>1.5 x10<sup>-8</sup></b>	<b>4.4x10<sup>-11</sup></b>	<b>3.0x10<sup>-10</sup></b>	<b>3.x10<sup>-7</sup></b>	0.257	0.150	0.679	0.489
>100kb the closest TSS	<b>2.2x10<sup>-5</sup></b>	<b>3.6x10<sup>-22</sup></b>	<b>3.2x10<sup>-16</sup></b>	<b>2.7x10<sup>-16</sup></b>	<b>0.035</b>	0.320	<b>0.041</b>	0.475

**Bold font: p<0.05**

Distances to the closest oncogene transcription start site; p-values: Fisher exact test (2-sided)

	Samples vs. Matched Random Control				Selected vs. non selected		Vitro vs. vivo	
	Vitro		Vivo		Vitro	Vivo	Non selected	Selected
	Non selected	Selected	Non selected	Selected				
0-100kb upstream the closest TSS	0.817	<b>1.9x10<sup>-3</sup></b>	0.065	<b>1.4x10<sup>-3</sup></b>	0.057	0.599	0.208	0.878
0-100kb downstream the closest TSS	<b>1.3x10<sup>-5</sup></b>	<b>2.4x10<sup>-12</sup></b>	<b>9.3x10<sup>-6</sup></b>	<b>4.8x10<sup>-7</sup></b>	0.887	1.000	0.631	0.352
>100kb the closest TSS	<b>7.8x10<sup>-4</sup></b>	<b>4.2x10<sup>-15</sup></b>	<b>1.0x10<sup>-6</sup></b>	<b>3.7x10<sup>-10</sup></b>	0.194	0.724	0.772	0.472

**Bold font: p<0.05**

**SUPPLEMENTAL TABLE 2**

Unique lentiviral insertion sites, mapped on hG19 (Excel file).



## SUPPLEMENTAL METHODS:

### Insertion site retrieval and bioinformatics

DNA was extracted with the Nucleospin Blood kit (Macherey Nagel). All primers were purchased from Eurogentec (see **primer list**). Linear amplification-mediated (LAM) PCR was performed essentially as previously described.<sup>1</sup> Each DNA sample (500 ng) was linearly amplified for 50 cycles with the 5'-biotinylated primer LTR1b and Q5 high-fidelity DNA polymerase (New England Biolabs, Evry, France). Linear PCR products were captured on M-280 streptavidin-coupled Dynabeads (ThermoFisher Scientific), and complementary DNA strands were synthesized with the Klenow fragment of DNA polymerase I (NEB) and hexanucleotides. The resulting double-stranded DNAs were separately digested with *Mlu*CI, *Mse*I or *Hin*P1I, ligated to complementary linker cassettes (LK3, LK4 and LK5, respectively) with the Fast Link DNA ligase (Lucigen, Middleton WI, USA), and denatured with sodium hydroxide. *De novo* synthesized (non-biotinylated) single-stranded DNA fragments were purified with a magnetic stand and pooled by sample. Exponential PCR was performed on pooled samples with primers LTRIIIB and LCII, and the Q5 polymerase, for 35 cycles. PCR products were purified with NucleoMag NGS beads (Macherey Nagel), in a 1.1:1 ratio, and quantified using the Qubit fluorometric assay (ThermoFisher Scientific). Next-generation sequencing (NGS) libraries were generated by PCR amplification on 40 ng of PCR products with the LCII-SEQ and LTRIV-SEQ primers, for 12 cycles with Q5 polymerase, and were purified with nucleoMag NGS beads used in a 0.8:1 ratio. It was possible to sequence all amplicons in one run, because each library was prepared with a specific LTRIV-SEQ primer, these primers differing by 6 to 12 nucleotides introduced between the Illumina adapter sequence and the nucleotide sequence complementary to the vector LTR. PCR products were analyzed with the 2100 Bioanalyzer high-sensitivity DNA chip (Agilent), quantified with the Qubit assay, mixed in equimolar concentrations with genomic libraries to compensate for low diversity, and sequenced on a NextSeq 500 instrument (Illumina) by the NGS core facility of I2BC (I2BC, CNRS, Gif-sur-Yvette, France). Once the run was completed, the reads were demultiplexed with *bcl2fastq2* version 2.15.0 and *cutadapt* version 1.9.1, and FASTQ format files were generated for downstream analysis.

Demultiplexed reads were analyzed on a local Galaxy platform, with the Galaxy open-source application<sup>2</sup> and tools obtained from the Galaxy tool shed.<sup>3</sup> The linker cassette sequence was removed with Trim Galore! ([http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) and reads were processed with tools from the FASTX-toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). Low-quality base calls (Phred score < 20) were trimmed at the 3' end, and reads of overall poor quality were discarded (Phred score < 25 for at least 10% of their bases). Reads containing the internal vector sequence were identified with the Burrows-Wheeler Aligner (BWA)<sup>4</sup> and discarded. Actual vector-genome-junction reads were identified by mapping the lentiviral vector long terminal repeat (LTR) with BWA, imposing an overall minimal sequence homology of 89% on the last 42 nucleotides of the LTR and a perfect match on the last three bases. Selected reads were trimmed for the LTR sequence, and reads of >29 bp were aligned with the Hg19/GRCh37 human genome with BWA. The resulting BAM alignment files were filtered with SAMtools to retain only reads mapped with the highest mapping probability (mapping quality score of 37).<sup>5</sup> The integration site position for each read was retrieved, formatted in BED format,<sup>6</sup> and integration sites were considered to be identical if their relative alignments began within 3 bp of each other. Collision filtering was performed to assign insertion sites to their correct group of samples, based on a minimal read abundance >90%. Insertion site positions are presented **Table S2**. A matched random control (MRC) dataset was generated by the random extraction of 3000 insertion sites of DNA sequences mapped onto the Hg19/GRCh37 human genome, with distances to the closest cutting site for *Miu*CI, *Mse*I, or *Hin*P1I of 20 to 500 bp.

The degree of enrichment in integration events around genomic features was compared between the MRC dataset and the insertion sites retrieved from LTGCPU7-transduced cells. Using the BEDTools suite,<sup>6</sup> the distance to the closest transcription start sites (TSS) of refseq genes and oncogenes, the intersection with exons and introns, and the densities of transcripts, DNase I hypersensitive sites (HS) and CpG islands were determined. The genomic coordinates of CpG islands, refseq genes, and DNase I HS sites were obtained from the University of California Santa Cruz (UCSC) annotation database (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/>).<sup>7</sup> The list of oncogenes was retrieved from the allonco data set (<http://www.bushmanlab.org/links/genelists>) generated by Doctor Rick Bushman (University of Pennsylvania). The DNase I HS sites of CD34<sup>+</sup> cells were obtained from data generated through the ENCODE project<sup>8</sup> (Prof. Stamatoyannopoulos, University of Washington) and archived under UCSC accession number wgEncodeEH001885. The expression profile was generated from RNA-seq data generated by the BLUEPRINT Consortium for cord blood hematopoietic stem cells. It has been archived by the European Genome-phenome archive under the experiment number EGAX00001157677 as part of the EGAD00001002316 data set. A full list of the investigators contributing to data generation is available from [www.blueprint-epigenome.eu](http://www.blueprint-epigenome.eu). Funding for the project was provided by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 282510 – BLUEPRINT.

## Primer list

### Primers and probes used for real-time PCR

Assay	Amplicon	Primer and Probe	Sequence (5'-3')	modification	Concentration (nM)
NIH3T3 Titer	Lentiviral vector	GAGF	GGAGCTAGAACGATTTCGCAGTTA	-	720
		GAGR	GGTTGTAGCTGTCCCAGTATTTGTC	-	720
		GAGP1	ACAGCCTTCTGATGTCTCTAAAAGGCCAGG	5'FAM 3'TAMRA	140
	Mouse βActin	mbactF	ACGGCCAGGTCATCACTATTG	-	900
		mbactR	CAAGAAGGAAGGCTGGAAAAGA	-	900
		mbactP1	CAACGAGCGGTTCCGATGCCCT	5'FAM 3'TAMRA	250
Vector copy number	Lentiviral vector	HPV569F21	TGTGTGCCCGTCTGTTGTGT	-	900
		HPV569R19	CGAGTCCTGCGTCGAGAGA	-	900
		HPV569P2	CAGTGGCGCCCGAACAGGGA	5'FAM 3'BHQ1	200
	Human HCK	hHCKF1	TATTAGCACCATCCATAGGAGGCTT	-	900
		hHCKR1	GTTAGGGAAAGTGGAGCGGAAG	-	900
		hHCKP3	TAACGCGTCCACCAAGGATGCGAAT	5'YY 3'BHQ1	200

FAM, 6-carboxyfluorescein ester; TAMRA, tetramethyl-6-carboxyrhodamine; YY, yakima yellow; BHQ1, black hole quencher-1

### Primers used for LAM-PCR

step	Cassette	Primer	Sequence (5'-3')	modification	Concentration (nM)
Lin PCR		LTRIB	GCCTCAATAAAGCTTGCC	5'biotin-TEG	1.7
Ligation	LK3 (MluCI)	LC1	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG	-	
		LC3	AATTCCTAACTGCTGTGCCACTGAATTCAGATC	-	
	LK4 (MseI)	LC1	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG	-	
		LC4	TACCTAACTGCTGTGCCACTGAATTCAGATC	-	
	LK5 (HinPI)	LC1	GACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG	-	
		LC5	CGCCTAACTGCTGTGCCACTGAATTCAGATC	-	
Exp PCR		LTRIIIB	GTGTGACTCTGGTAACTAGAG	-	167
		LCII	AGTGGCACAGCAGTTAGG	-	167

Biotin-TEG, Biotin- triethylene glycol spacer

### Primers used for the constitution of Illumina libraries

Primer	Sequence (5'-3')	Concentration (nM)
LCII-SEQ	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTCCGATCTAGTG GCACAGCAGTTAGG	100
LTRIV-SEQ	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTCCGATCT- <u>BC</u> - GATCCCTCAGACCCTTTTAGTC	100
LTRIV-SEQ-BC1	BC = AGTCAA	
LTRIV-SEQ-BC2	BC = GGCTACCC	
LTRIV-SEQ-BC3	BC = ACAGTGCCCT	
LTRIV-SEQ-BC4	BC = TAGCTTTTATAG	
LTRIV-SEQ-BC5	BC = CTTGTA	
LTRIV-SEQ-BC6	BC = GATCAGCG	
LTRIV-SEQ-BC7	BC = ATCACGATCT	
LTRIV-SEQ-BC8	BC = TTAGGCCTCTAG	
LTRIV-SEQ-BC9	BC = CAGATC	

BC: 6 to 12 nucleotides introduced between the Illumina adapter sequence and the nucleotide sequence complementary to the vector LTR

## SUPPLEMENTAL REFERENCES

1. Schmidt, M, Schwarzwaelder, K, Bartholomae, C, Zaoui, K, Ball, C, Pilz, I, *et al.* (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nature methods* **4**: 1051-1057.
2. Afgan, E, Baker, D, van den Beek, M, Blankenberg, D, Bouvier, D, Cech, M, *et al.* (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*.
3. Blankenberg, D, Von Kuster, G, Bouvier, E, Baker, D, Afgan, E, Stoler, N, *et al.* (2014). Dissemination of scientific software with Galaxy ToolShed. *Genome biology* **15**: 403.
4. Li, H, and Durbin, R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.
5. Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
6. Quinlan, AR, and Hall, IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
7. Speir, ML, Zweig, AS, Rosenbloom, KR, Raney, BJ, Paten, B, Nejad, P, *et al.* (2016). The UCSC Genome Browser database: 2016 update. *Nucleic acids research* **44**: D717-725.
8. Consortium, EP (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.