Supplementary Table 1. Frequency of lentiviral genes in refseq genes and CpG regions

	HF-LV G0	HF-LV Stim	Random G0	Random stim
Total IS	6631	2930	19893	8751
IS in Refseq genes	74,71	68,81	39,67	40,65
IS in Refseq exons	4,45	4,54	1,84	1,83
IS in CpG	0,05	0,07	0,18	0,17

Supplementary figure 2 Vector copy number in different lineages in vivo upon reconstitution with VSVG-LV and H/F-LV transduced unstimulated $CD34^+$ cells.

Isolated human cell lineages[a]	**VSVG-LV**		**H/F-LV**	
	% GFP (mean +/- SD)	**VCN/cell** (mean +/- SD)	**% GFP** (mean +/- SD)	**VCN/cell** (mean +/- SD)
BM-CD34+ cells	3.4+/-3.6	0.16+/-0.15	94.5+/-4.2	12.8 +/- 4.4
BM-CD13+ cells	3.7+/-5.7	0.20+/-0.18	96.0+/-5.1	13.8 +/- 1.3
Spleen-CD19+ cells	4.5+/-5.4	0.20+/-0.12	92.9+/-7.7	12.1 +/- 1.2
Thymocytes	1.1+/-1.2	0.06+/-0.05	91,3+/-4,8	8.4 +/- 3.4

[a] NSG mice were reconstituted with HF-LV or VSVG-LV transduced unstimulated CD34+ cells. For VSV-G an MOI of 50 and for HF-LVs an MOI of 20 was used. 12 weeks post-reconstitution, mice were sacrificed and the different indicated hematopoietic lineages were isolated for VCN analysis.

Supplementary material and methods

Lentiviral vector concentration

The vectors were concentrated at low speed by overnight centrifugation of the viral supernatants at 3,000g at 4°C. This concentration method has now been broadly spread in many labs for H/F-LV and VSVG-LV production. We have compared this method with the classical method used for VSVG-LV concentration (25000 rpm, 2h, 4°C) and found no significant differences in titers Infectious titers (in transduction units [TU]/ml) were determined by fluorescence-activated cell sorting (FACS) of target cells using serial dilutions of the supernatants added to the appropriate target cell. Nor did we see a difference in transduction efficiencies for CD34+ cells between the vectors concentrated by these two methods. For all comparisons between VSV-G LVs and HF-LVs, vectors were titered on the same cell line, 293T cells (Figure Aa). Exclusively for the comparison of HD4F-LVs and HF-LVs, we verified these SLAM tropic vector on CHO-SLAM cells.

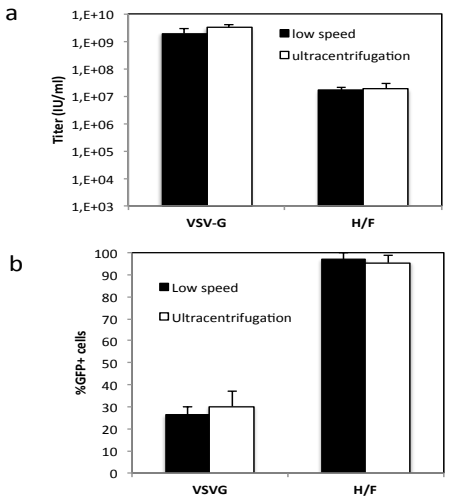


Figure A: Comparison of low speed and by ultracentrifugation concentrated vectors

HF-LVs and VSVG-LVs were concentrated 100-fold using low speed concentration (3000g, 18h, 4°C) or ultracentrifugation (2h, 25000 rpm, 4°C) and were titered in (a) on 293T cells by applying serial dilutions of the vectors. (b) The vectors concentrated by the two method were compared for transduction of cytokine pre-stimulated $CD34^{+}$ cells (VSVG-LV, MOI of 50 and H/F-LV at MOI of 10). 10 days post-transduction GFP+ cells were analyzed by FACS.

Titration of the lentiviral vectors

1. Titration by FACS analysis

To determine the infectious titers of the HIV vectors, 293T cells were incubated with serial dilutions of the vector supernatants overnight. Medium on the cells was replaced with 2 ml fresh DMEM and cells were incubated for further 72h. The percentage of GFP^+ cells was determined by FACS. To exclude GFP pseudo-transduction and thus verify that expression was only due to integrated vector copies we continued the culture of these 293T cells for 10 more days (4 passages) and determined again the percentage of GFP^+ cells by FACS and also confirmed in parallel the titer by Q-PCR on these same samples (see below).

2. Titration by Q-PCR

Genomic DNA was extracted from the 293T cells transduced with LVs as indicated above and kept in culture for 14 days after transduction, using the genomic DNA purification kit (Nucleospin Tissue XS, Macherey-Nagel, Hoerdt, France). Q-PCR to determine the titer was performed using Taqman gene expression master mix (Life Technologies, St Aubin, France) and following primer pairs and Taqman probes to amplify the lentiviral vector genome:

Lenti-F : 5'-TGT GTG CCC GTC TGT TGT GT-3'

Lenti-R : 5'-GAG TCC TGC GTC GAG AGA GC-3'

Lenti-Probe : 5'-CAG TGG CGC CCG AAC AGG GA-3'

Results were compared with those obtained after serial dilutions of genomic DNA from a 293T cell line containing one copy of the integrated lentiviral vector per haploid genome.

3. Titration for physical particles (HIV p24gag content)

Vectors were titered for physical particles by measuring p24 antigen using an enzyme-linkedimmunoabsorbant assay (ELISA) following manufacturer's instructions (RetroTek-ZeptoMetrix, Buffalo, NY)

Table A: Infectious and p24 titers of H/F LVs compared to VSV-G-LVs

Vector (100x-concentrated[a])	293T (TU/ml)[b]	p24 (ng/ml)[c]
H/F-LVs	1.6E7 ± 5,4E6	2550±670
VSVG -LVs	4.5E8± 1.8E8	2820±620

[a]Supernatants of 293T producer cells were concentrated 100-fold by low-speed centrifugation.
[b]Titers were assessed by adding serial dilutions of each vector preparation to 293T cells. For each vector the % of GFP^+-cells was analysed at day 3 post-transduction. Titers are expressed as transduction units per ml (TU/ml; mean±SD; n=6).
[c]Titers expressed as p24 level (ng/ml) determined by ELISA which represents a measure for total physical particles present in the vector preparation

DNA extraction and Quantitative PCR

Estimation of the mean vector copy number per transduced cell was determined by qPCR performed on human CB $CD34^+$-cells 14 days after transduction.

Genomic DNA was prepared using NucleoSpin Tissue XS kits (Macherey Nagel, Hoerdt, France). Average vector copy numbers were determined by qPCR with primers amplifying the packaging signal (Primers: Fwd: TGT GTG CCC GTC TGT TGT GT, Rev: GAG TCC TGC GTC GAG AGA GC, Probe: CAG TGG CGC CCG AAC AGG GA) after normalization for endogenous beta-actin genes (Primers: Fwd: TCC GTG TGG ATC GGC GGC TCC, Rev: CTG CTT GCT GAT CCA CAT CTG, Probe: CCT GGC CTC GCT GTC CAC CTT CCA). Results were compared with those obtained after serial dilutions of genomic DNA from a cell line containing one copy of the integrated lentiviral vector per haploid genome.

Cell cycle analysis

Cell-cycle analysis was performed by staining DNA and RNA with 7-amino-actinomycin-D (7AAD; Sigma, St Quentin-Fallavier, France) and pyronin Y, respectively. A total of $3x10^5$ cells were labeled with 7AAD at a final concentration of 20 μM for 30 minutes at room temperature followed by 5 μM pyronin Y (Sigma) for 10 minutes on ice. Cells were immediately analyzed by FACS.

FACS analysis

Receptor surface expression was detected with anti-CD46-PE, anti-SLAM-PE (BD Biosciences) and anti-Nectin-4-PE (R and D Systems, Mineapolis, USA).

In vitro transduction of $hCD34^+$-cells was detected by APC-coupled anti-CD34 antibody. Transduced $hCD34^+$-cells were also stained with anti-CD34-Pacific blue, anti-CD90-APC and anti-CD38-PE. Cells dissociated from the CFC colonies were washed once in PBS and stained with anti-GlyA-PE, anti-CD14APC, anti-CD15PECy7, antiCD71PE. All these antibodies were purchased from Myltenyi (Germany).

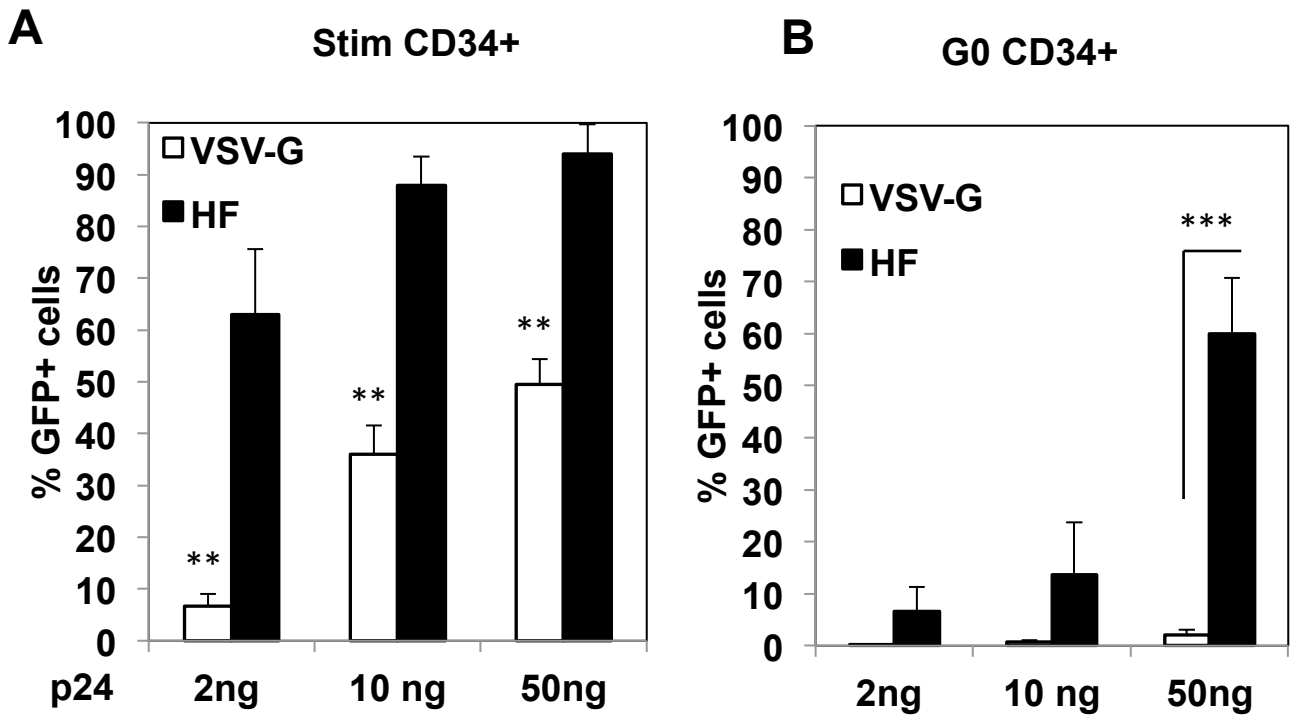
For the detection of LV transduction of the engrafted human cells in NSG mice flow cytometry analysis was performed using APC-conjugated anti-hCD45 antibody for the detection of total human cell engraftment in the bone marrow, thymus, peripheral blood and spleen. APC-coupled antibodies were used for the detection of hCD4, hCD56, hCD19, hCD14, and hCD13. Staining of progenitors and early B-cell population in spleen and BM was performed using anti-hCD34-APC, anti-hCD19-PECy7, anti-CD10PE. Staining of thymocytes with anti-hCD3-PE, anti-hCD8-PECy7 and anti-hCD4-APC was performed to screen for thymic subpopulations. Blood samples were stained using anti-CD3APC, anti-hCD19-PECy7, CD14-APC. All antibodies were purchased from Myltenyi (Germany).

Secondary recipient NSG mice for BaEVTR-LV transduction were analysed for % of GFP^+ hB cells ($hCD19^+$), human immature progenitor cells ($CD34^+$) and myeloid progenitor cells ($CD13^+$) by FACS.

Integration site analysis

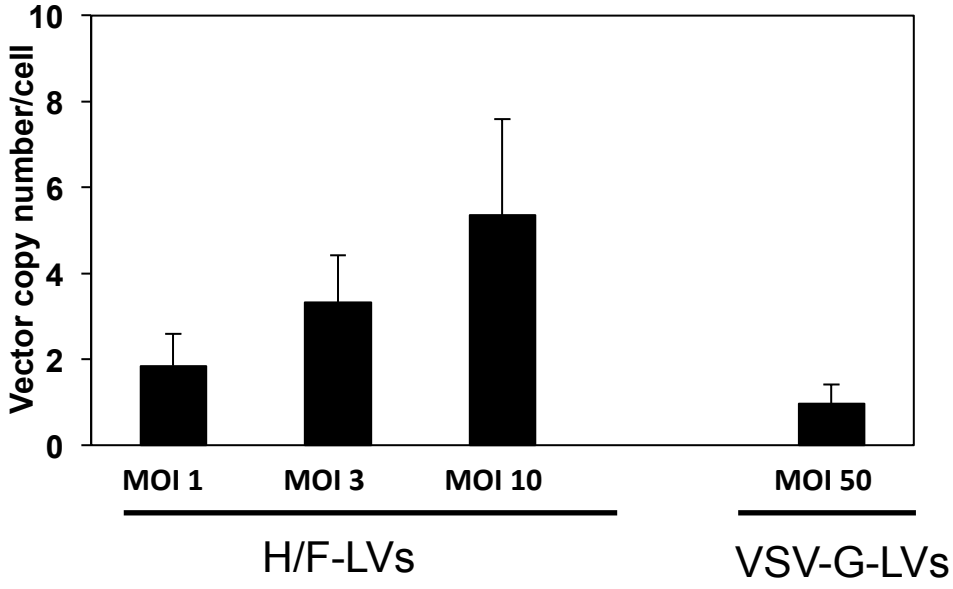
Freshly isolated hCD34+ cells or CD34+ cells prestimulated for 24h with a cytokine cocktail (TPO+SCF+Flk-3) were transduced with H/F-LVs at an MOI of 10. Seven and 8 independent transductions were performed for resting and prestimulated $CD34^+$-cells, respectively. After 24h transduction the cells were washed and cultures were continued for 7 days; then the samples were processed for genomic DNA isolation and integration site analysis by LAM-PCR. Briefly, linkers were ligated to restriction enzyme-digested (*Mse*I) genomic DNA isolated from transduced cells and virus-host DNA junctions were amplified by nested PCR. Samples were individually barcoded with the second pair of PCR primers to generate 454 libraries. PCR products were purified and sequenced using 454/Roche pyrosequencing (titanium technology). Reads were quality-filtered by requiring perfect matches to the long

terminal repeat (LTR) linker, barcode, and flanking LTR and subsequently mapped to the human genome. All sites were required to align to the reference genome within 3bp of the LTR edge. In order to control for possible biases in the datasets due to the choice of the *Mse*I restriction endonuclease in cloning integration sites, matched random control (MRC) sites were generated *in silico*. To do so, each experimental integration site was paired with three sites in the genome, locating with the same distance from a randomly selected *Mse*I site in the genome. Briefly, sites were judged to be authentic when the sequences contained a proper bar code and long terminal repeat (LTR) primer and had a best unique hit when aligned to the human genome as appropriate (hg18) using BLAT, and the alignment began within 3 bp of the viral LTR end and had >98% sequence identity. Integration site frequencies were compared with matched random controls (MRCs) by Fisher's exact test (where stated). Analysis was carried out using The R Project for Statistical Computing. Heat maps are developed to summarize many relationships using receiver operating characteristic (ROC) area method. For heat maps, comparisons were carried out with different interval sizes surrounding each integration site (10kb and 50kb), since previous studies have shown that the interval sizes chosen for comparison can influence the conclusions. In this study, results were similar for all interval size examined (data not shown), so only the data for 10kb intervals are shown. Results of statistical tests comparing the distributions of integration sites to the reference dataset are summarized as asterisks on each tile of the heat map.



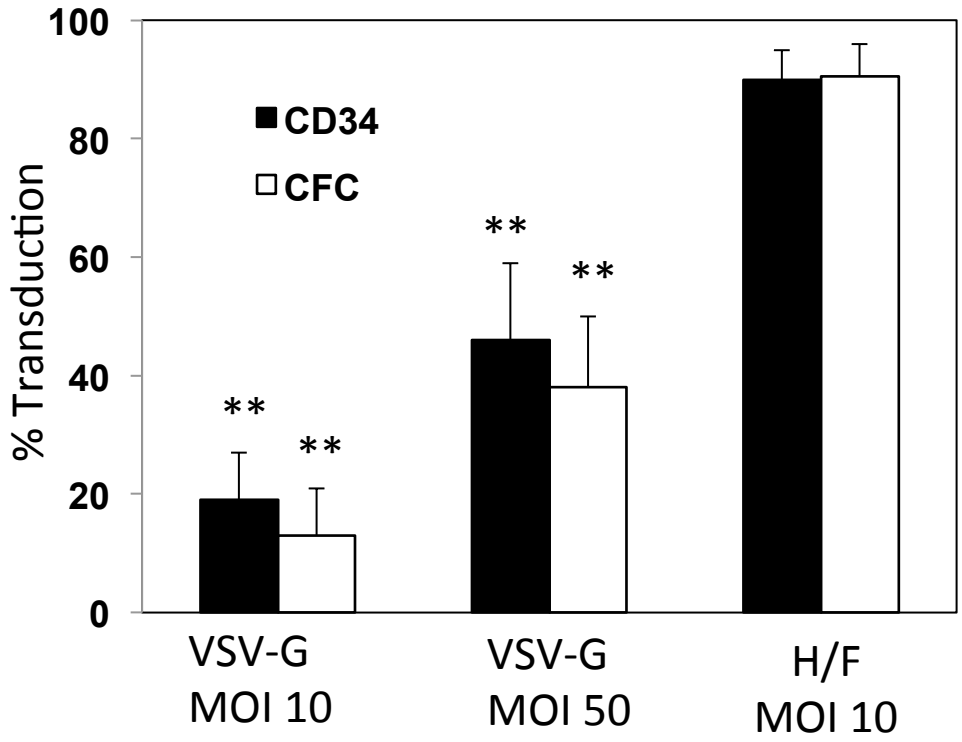
Supplementary Figure 1: hCD34$^+$-cell transduction with different LV pseudotypes at equivalent physical particle doses.

(A) CD34$^+$-cells were prestimulated for 24h (SCF (100ng/ml), TPO, (20ng/ml), Flt3-L, (100ng/ml)) and transduced at the indicated HIV p24 levels with VSV-G-LVs or HF-LVs. After 24h transduction, cells were washed and cultured in the presence of the same cytokine cocktail for 6 more days. The % of GFP$^+$ cells was determined by FACS analysis (means±SD, n=3); VSVG-LV was compared to HF-LV for same p24 dose by paired student t-test,** $p< 0.01$.
(B) Same transduction protocol as in (A) but using unstimulated (G0) hCD34$^+$-cells (means±SD, n=3); paired student t-test, *** $p<0.001$.



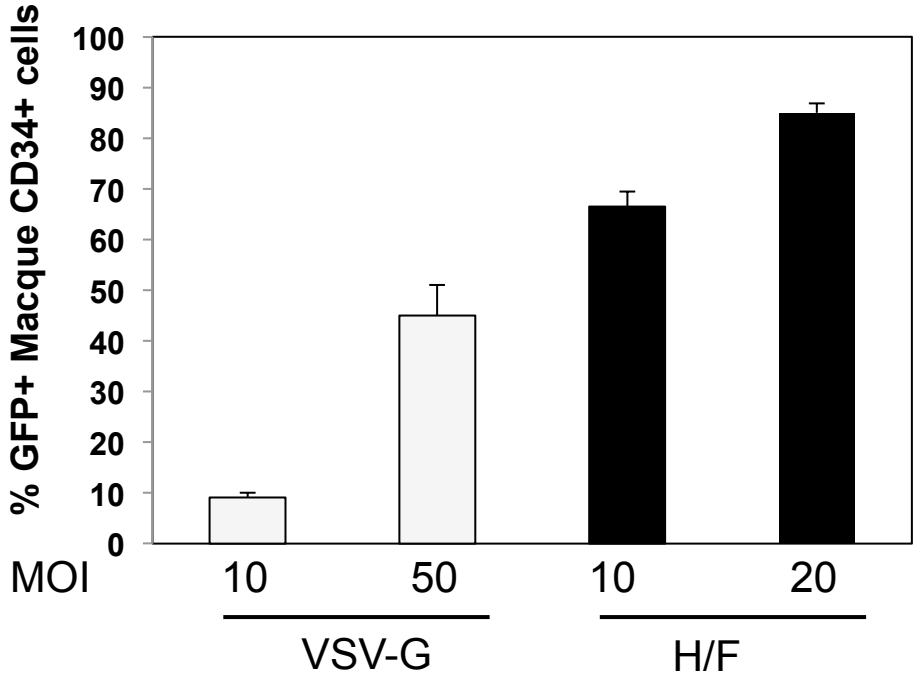
Supplementary figure 2: Copy number detection in HF-LV and VSVG-transduced cytokine pre-stimulated CD34+ cells.

hCD34^{+}-cells were pre-stimulated with a cytokine cocktail (TPO,SCF, Flt3) and transduced with H/F-LVs or VSV-G-LVs at indicated MOIs in the presence of Retronectin for 48h. Culture was continued till 14 days post-transduction; genomic DNA was isolated from these cells and the copy number of integrated vector/cell was determined by Q-PCR (means±SD, n=4).



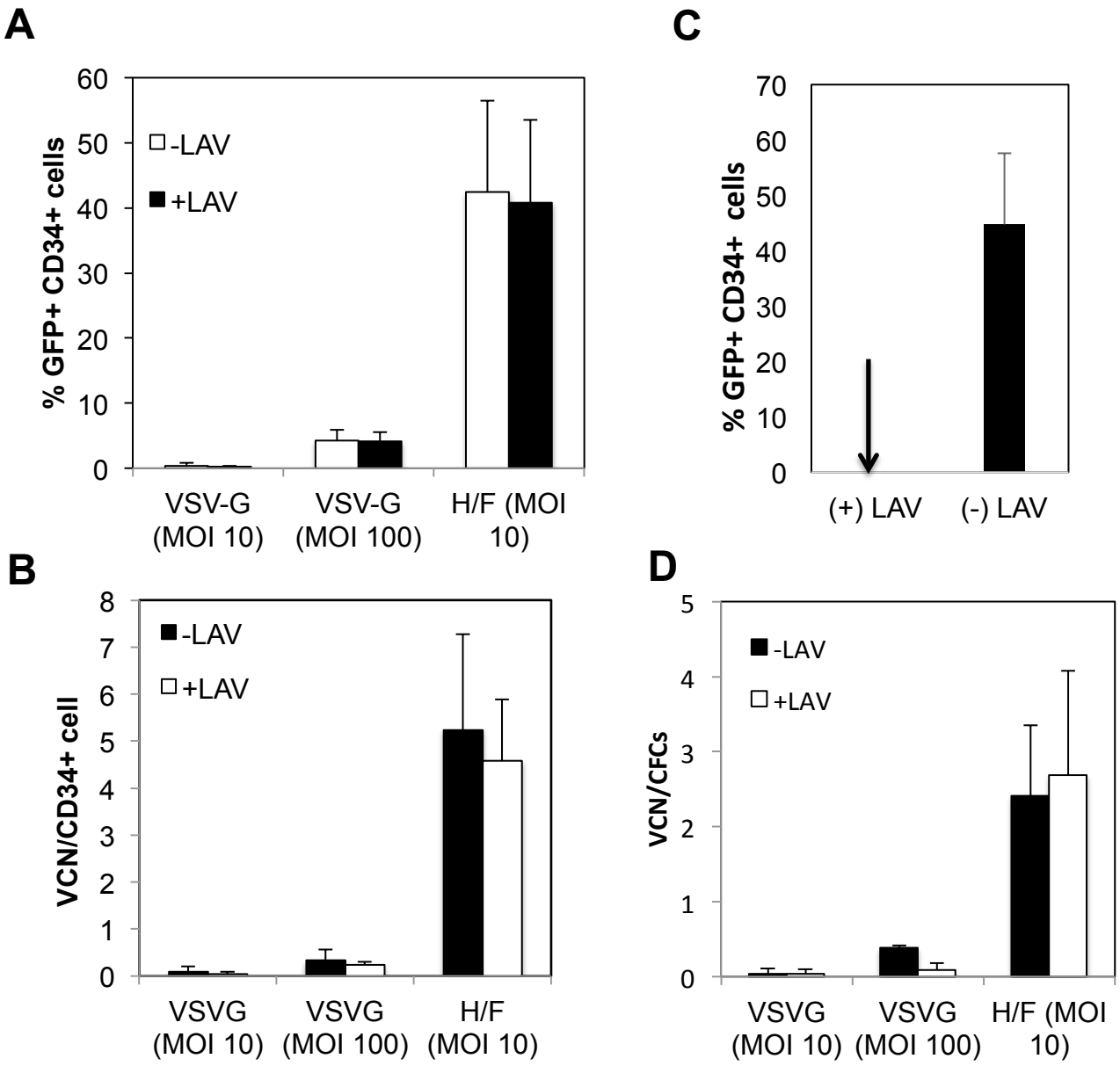
Supplementary Figure 3: H/F-LVs transduce efficiently mobilized hCD34$^+$-cells.

Mobilized hCD34$^+$-cells were prestimulated with SCF+TPO+Flt3-L+IL-3 and transduced with the indicated LV pseudotypes in the presence of retronectin. An MOI of 10 was applied except for VSV-G LVs for which in addition an MOI of 50 was used. Three days after transduction, the cells were seeded in myeloid differentiation medium for 14 days. The % of GFP$^+$ CD34$^+$-cells 3 days post transduction and the GFP$^+$ colony forming cells (CFCs) at day 14 of differentiation were determined by FACS (means±SD, n=3; $p<0.01$ when comparing CFC transduction levels of HF-LVs to the VSVG pseudotype). VSV-G-LV MOIi 10 and 50 was compared to HF-LVs by apaired student t-test. ** $p< 0.01$.



Supplementary Figure 4: SIV-derived LVs pseudotyped with H and F MV gps, allow efficient transduction of macaque BM CD34$^+$-cells

(A) Cynomolgus macaque BM CD34$^+$-cells were prestimulated with rSCF, rTPO, rFlt3-L and IL-3 for 24h and were transduced with SIV-derived LVs pseudotyped with H and F measles virus envelope glycoproteins (Black bars) or VSV-G-LVs (grey bars) at the indicated vector dosis (MOI). All transduction were performed in the presence of retronectin. GFP expression of the cells was analyzed 6 days after transduction by FACS (means±SD, n=3).

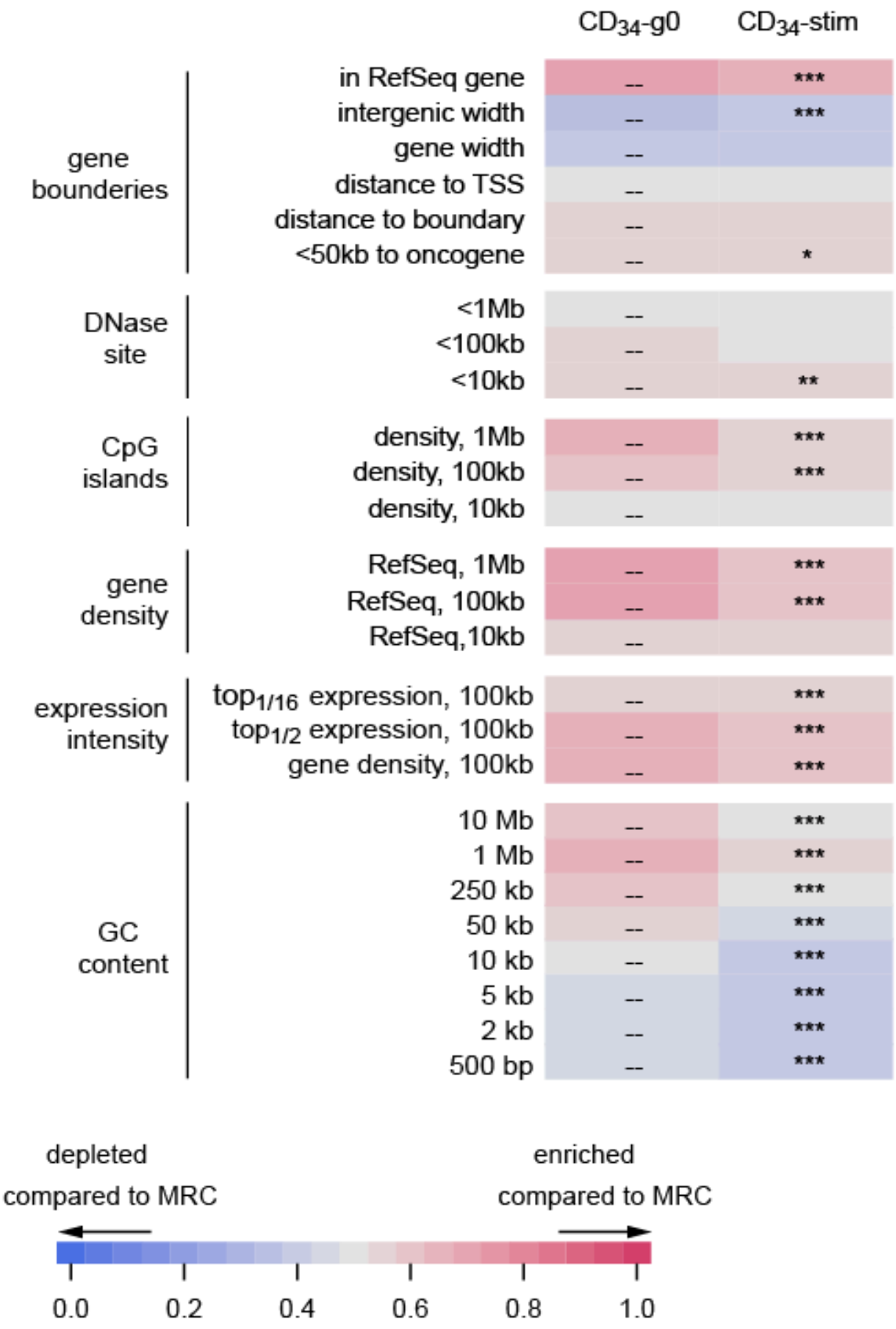


Supplementary figure 5: HF-LV and VSVG-LV transduction of unstimulated CD34+ cells in the presence and absence of a retrotranscriptase (RT) inhibitor

Unstimulated hCD34$^+$-cells (G0 CD34$^+$-cells) were transduced with H/F-LVs and VSVG-LVs at indicated MOIs in the presence of retronectin for 3 days. The cells were washed and continued in liquid culture supplemented with cytokines (TPO, SCF and FL-k3 L) for 10 more days without (-LAV) or with (+LAV) the RT inhibitor, Lamivudin (LAV). (A) The % of GFP$^+$ CD34$^+$-cells 10 days post transduction was determined by FACS. (B) Genomic DNA was isolated from these cells and the copy number of integrated vector/cell was determined by Q-PCR.

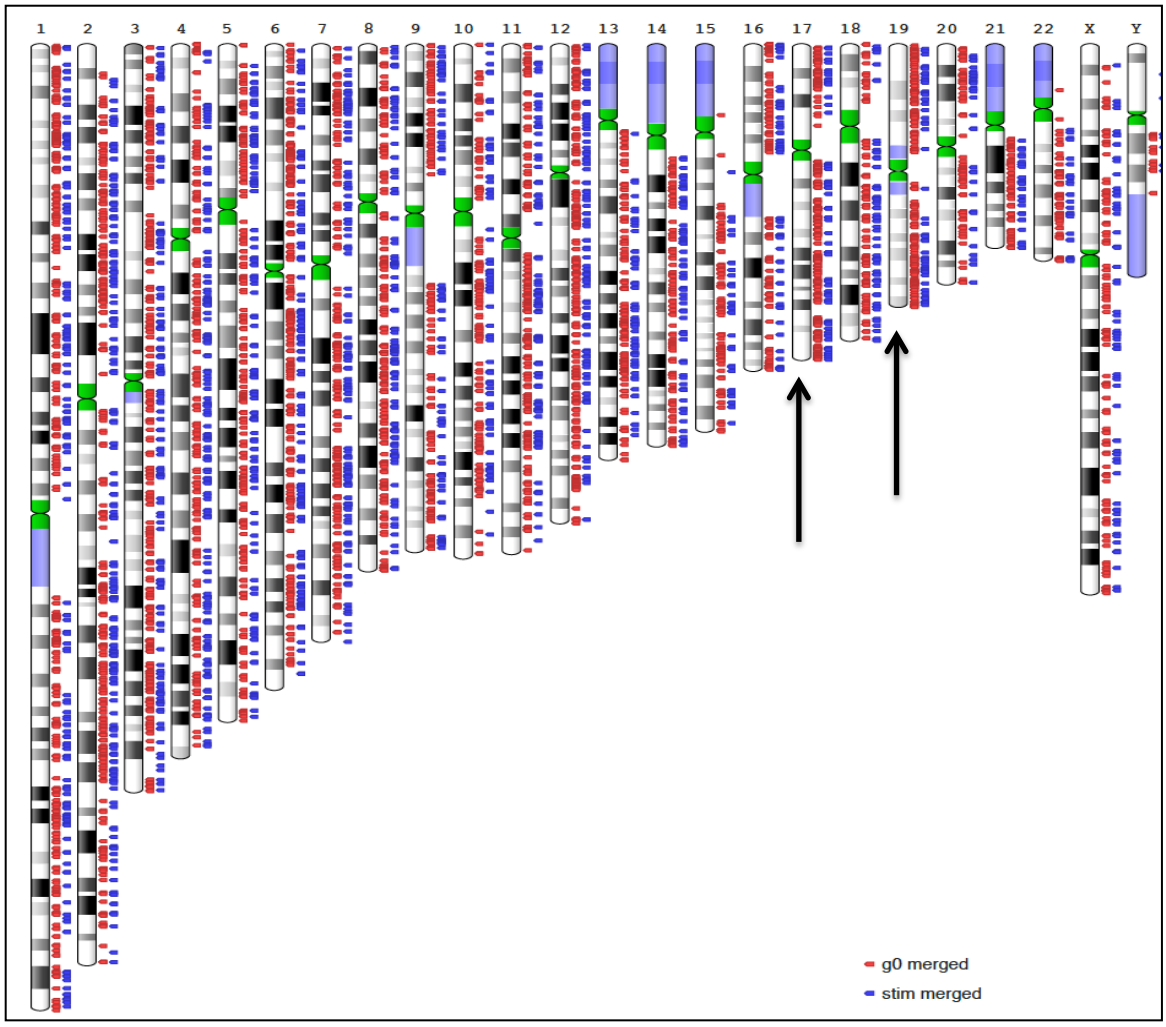
(C) Unstimulated CD34+-cells were preincubated 2h with RT inhibitor (+LAV) or not and then transducted with HF-LVs. The % of GFP$^+$ CD34$^+$-cells was determined at day 3 of transduction. (means ± SD, n=4).

In (D) the transduced G0 hCD34$^+$-cells were seeded for CFC assay, at 14 days of culture the cells were verified for GFP+ CFCs (Figure 2E) and genomic DNA was isolated from these CFCs and the copy number of integrated vector/cell was determined by Q-PCR.

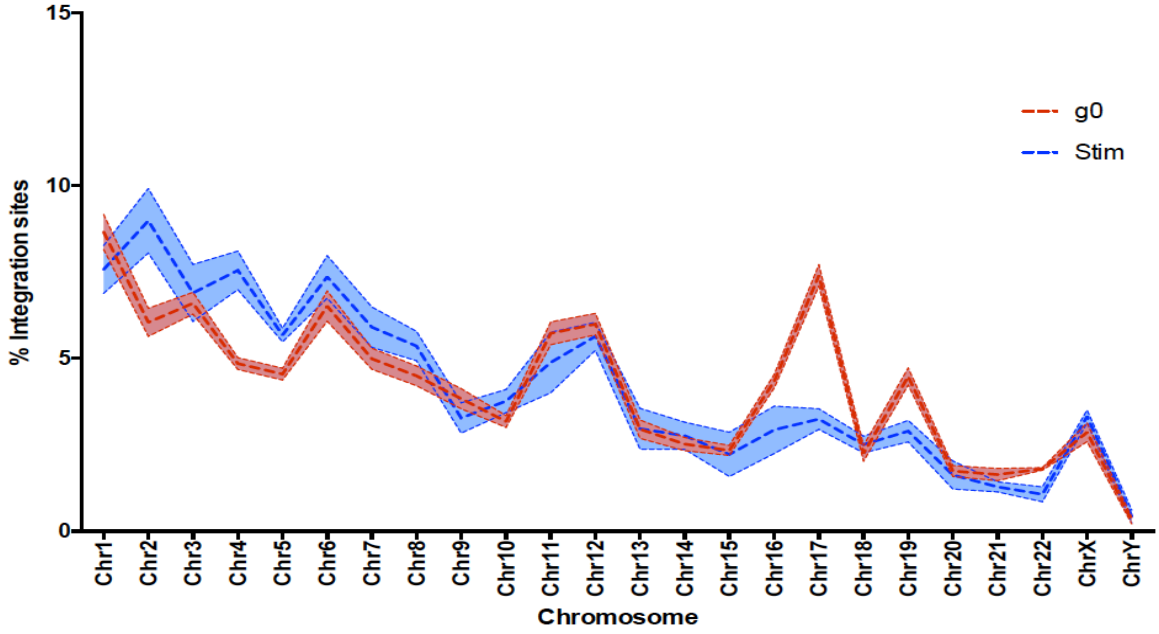


Supplementary Figure 6: Genomic heat map for LV-integration sites in human CD34^{+}-cells

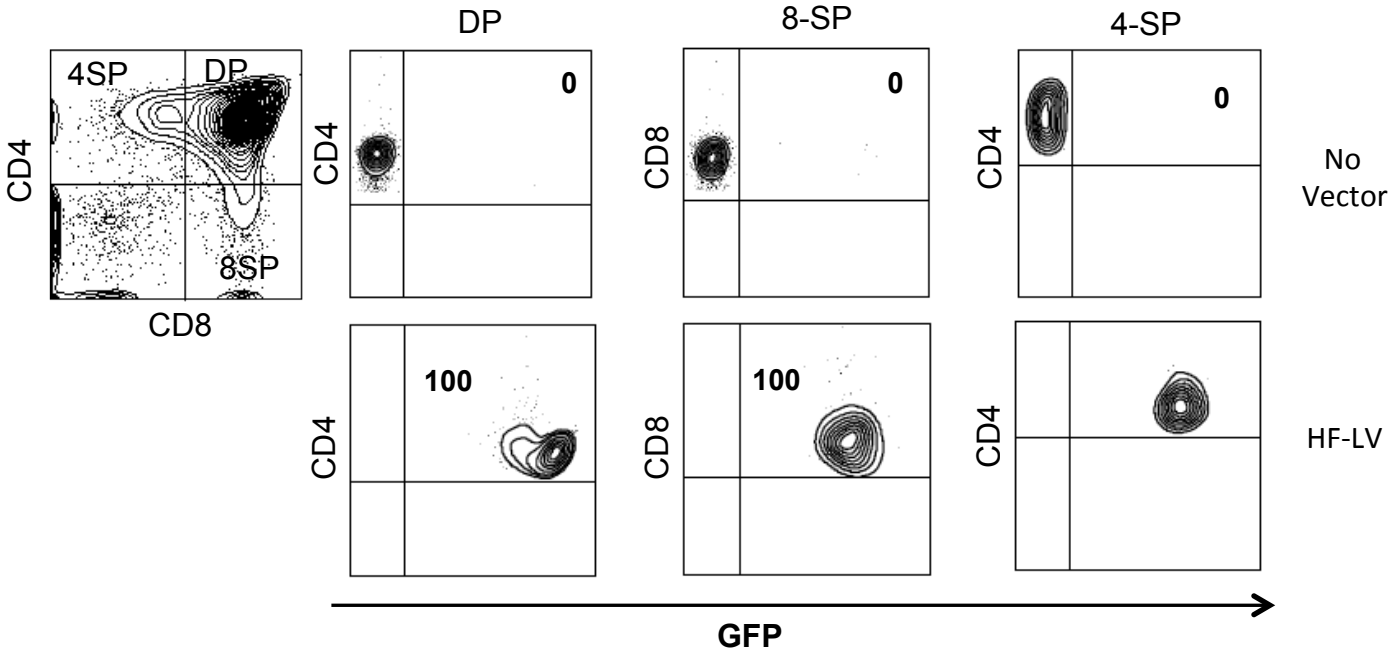
A



B



Supplementary Figure 7 : Chromosomal distribution of LV-integration sites in human CD34⁺-cells. Unstimulated (g0) or cytokine stimulated CD34⁺-cells (Stim) were transduced with HF-LVs for 24h, washed and kept in culture for 10 days and were then processed for LV genomic IS analysis. In (A) the IS density for each chromosome is indicated, while in (B) the chromosomal distribution of the LV-IS is summarized.



Supplementary Figure 8: High-level transduction in human thymocytes after engraftment in recipient NSG mice of transduced cytokine-stimulated hCD34$^+$-cells

hCD34$^+$-cells were prestimulated with a cytokine cocktail (TPO+SCF+Flk-3L) for 16h and transduced with HF-LVs at MOI=10 in presence of retronectin for 36 h. Subsequently the cells (2E5) were injected into the liver of irradiated newborn NSG mice. Upon reconstitution for 12 weeks, thymus of these primary engrafted mice were analyzed. The transduced cells in the thymic subpopulations are shown (SP-4 (CD4$^+$CD8$^-$), SP-8 (CD4$^-$CD8$^+$) and DP (CD4$^+$CD8$^+$). Upper pannels show the FACS data for untransduced cells (no vector), lower pannels show FACS data for HF-LV transduced cells. The FACS profiles are representative of 5 independent experiments.