SUPPLEMENTARY MATERIALS

Supplemental Methods

SUPPLEMENTAL METHODS

Cell sources, CD34$^+$ enrichment and *in vitro* culture

Pigtailed macaque steady-state bone marrow, primed bone marrow and umbilical cord blood, as well as rhesus macaque steady-state bone marrow CD34$^+$ cells were harvested and enriched as previously described (*11*). Briefly before enrichment of nonhuman primate CD34$^+$ cells, red cells were lysed in ammonium chloride lysis buffer, and white blood cells were incubated for 20 minutes with the 12.8 immunoglobulin-M anti-CD34 antibody, then washed and incubated for another 20 minutes with magnetic-activated cell-sorting anti-immunoglobulin-M microbeads (Miltenyi Biotech, Bergisch Gladbach, Germany). For enrichment of human CD34$^+$ cells, microbead conjugated anti-CD34 antibody (Miltenyi Biotech) was used. The cell suspension was run through magnetic columns enriching for CD34$^+$ cell fractions with a purity of 60-80% confirmed by flow cytometry.

Colony-forming cell assay

For colony-forming cell assays 1000-1200 sorted cells were seeded into 3.5 ml ColonyGEL 1402 (ReachBio, Seattle, WA). Hematopoietic colonies were scored after 12-14 days. Arising colonies were identified as colony forming unit- (CFU-) granulocyte (CFU-G), macrophage (CFU-M), granulocyte-macrophage (CFU-GM) and burst forming unit-erythrocyte (BFU-E). Colonies consisting of erythroid and myeloid cells were scored as CFU-MIX. For secondary colony-forming cell assays, all cells or individual colonies were harvested, dissociated, washed, and seeded into 3.5 ml ColonyGel 1402. Hematopoietic colonies of secondary colony-forming cell assays were analyzed/quantified after 12-14 days without discrimination of colony-subtypes.

Megakaryocyte assay

For megakaryocyte differentiation. 3000-5000 sorted cells were seeded into 2 ml MegaCult (StemCell Technologies) and performed according to the manufacturer's instructions. Myeloid (granulocytes and/or macrophages/monocytes) as well as megakaryocyte colonies were quantified microscopically after 10-12 days.

T cell assay

T cell differentiation potential was tested as previously described (*40*). Sort-purified CD34- subpopulations were co-cultured with murine stromal cells OP9 expressing the mouse Delta-like 1 gene (OP9-DL1) (*41*) in α-minimal essential medium (α MEM) (Gibco by Life Technologies) supplemented with 20% FBS, 100 U/ml penicillin, 100 U/ml streptomycin, 5 ng/ml FLT3-L, and 5 ng/ml IL-7 (R&D System, Minneapolis, MN). Half-media changes were performed on a bi-weekly basis. Obtained cells were stained with antibodies against CD2, CD3, CD4, CD5, CD8, and CD45 and analyzed/quantified by flow cytometry 5 weeks post-seeding.

MS-5 assay

Lympho-myeloid differentiation potentials were tested in the clonal MS-5 assay as previously described *(34)*. Single-sorted hematopoietic progenitors were deposited onto stromal MS-5 cells (*42*) in 96-well plates. Co-cultures were carried out in H5100 medium (StemCell Technologies) supplemented with 100 U/ml penicillin, 100 U/ml streptomycin, 100 ng/ml SCF, 50 ng/ml TPO, 20 ng/ml IL-2 (R&D System) and 10 ng/ml IL-7 for 5 weeks with weekly half-media changes. Colony-forming potential was evaluated microscopically. Cells were harvested by physical

dissociation, filtered through a 70-µm filter, and stained with antibodies against CD11b, CD14, CD20, CD34, CD45, and CD56 for analysis. The phenotype of progeny (granulocytes: $CD11b^{+}CD14^{-}$; monocytes: $CD11b^{+}CD14^{+}$; and NK cells: $CD11b^{-}CD14^{-}CD56^{+}$) was determined by flow-cytometry.

Autologous nonhuman primate transplants and *ex vivo* engineering of HSPCs

Autologous nonhuman primate transplants, priming (mobilization), collection of cells and genetic engineering were conducted consistent with our previously published protocols (*11*). Briefly, animals were pre-treated with G-CSF and SCF for 4 days to prime the bone marrow. On day 4 bone marrow aspirates were performed, $CD34^{+}$ fractions isolated, and cells plated into StemSpan SFEM containing 1% penicillin/streptomycin (P/S, Life Technologies) supplemented with SCF, TPO, and FLT3-L 100 ng/ml each. Cells were pre-stimulated overnight and re-plated in culture medium containing 4 µg/ml protamine sulfate, and 1 µg/ml Cyclosporine A onto flasks coated with CH296 (retronectin) for transduction with lentiviral vectors (multiplicity of infection [MOI] 10). A second dose of lentivirus (MOI 10) was added after 6-8 hours for overnight culture. Next day, cells were collected, residual virus removed, and washed cells were resuspended in culture medium supplemented with PGE2. After 2 hours, cells were collected and prepared for infusion.

Nonhuman primate animal housing and care / ethics statement

Healthy juvenile pigtailed macaques and juvenile rhesus macaques, were housed at the University of Washington National Primate Research Center (WaNPRC) under conditions approved by the American Association for the Accreditation of Laboratory Animal Care. All experimental procedures performed were reviewed and approved by the Institutional Animal Care and Use

Committee of the Fred Hutchinson Cancer Research Center and University of Washington (Protocol #3235-01). This study was carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health ("The Guide") and monkeys were randomly assigned to the study. This study included at least twice-daily observation by animal technicians for basic husbandry parameters (e.g. food intake, activity, stool consistency, and overall appearance) as well as daily observation by a veterinary technician and/or veterinarian. Animals were housed in cages approved by "The Guide" and in accordance with Animal Welfare Act regulations. Animals were fed twice daily, and were fasted for up to 14 hours prior to sedation. Environmental enrichment included grouping in compound, large activity, or run-through connected cages, perches, toys, food treats, and foraging activities. If a clinical abnormality was noted by WaNPRC personnel, standard WaNPRC procedures were followed to notify the veterinary staff for evaluation and determination for admission as a clinical case. Animals were sedated by administration of ketamine HCl and/or telazol and supportive agents prior to all procedures. Following sedation, animals were monitored according to WaNPRC standard protocols. WaNPRC surgical support staff are trained and experienced in the administration of anesthetics and have monitoring equipment available to assist: electronic monitoring of heart rate, respiration, and blood oxygenation; audible alarms and LCD readouts; monitoring of blood pressure, temperature, etc. For minor procedures, the presence or absence of deep pain was tested by the toe-pinch reflex. The absence of response (leg flexion) to this test indicates adequate anesthesia for this procedure. Similar parameters were used in cases of general anesthesia, including the loss of palpebral reflexes (eye blink). Analgesics were provided as prescribed by the Clinical Veterinary staff for at least 48 hours after the procedures, and could be extended at the discretion of the clinical veterinarian, based on clinical signs. Decisions to

euthanize animals were made in close consultation with veterinary staff and were performed in accordance with guidelines as established by the American Veterinary Medical Association Panel on Euthanasia (2013). Prior to euthanasia, animals were first rendered unconscious by administration of ketamine HCl.

Preconditioning and supportive care for autologous nonhuman primate transplants

In parallel to cell processing, macaques were conditioned with myeloablative total body irradiation (TBI) of 1020 cGy from a 6 MV x-ray beam of a single-source linear accelerator located at the Fred Hutchinson Cancer Research Center South Lake Union Facility (Seattle, Washington, USA); irradiation was administered as a fractionated dose over the 2 days before cell infusion. During irradiation, animals were housed in a specially modified cage that provided unrestricted access for the irradiation while simultaneously minimizing excess movement. The dose was administered at a rate of 7 cGy/min delivered as a midline tissue dose. G-CSF was administered daily from the day of cell infusion until the animals began to show onset of neutrophil recovery. Supportive care, including antibiotics, electrolytes, fluids, and transfusions, was given as necessary, and blood counts were analyzed daily to monitor hematopoietic recovery.

Neutrophil/platelet engraftment and engraftment failure

Neutrophils: The day of neutrophil engraftment was defined as a minimum of 500 neutrophils per µl peripheral blood (lower limit for engraftment: LLE) for a duration of at least 3 consecutive days. The duration of G-CSF administration was not considered in this definition and is indicated by the yellow bar in each graph. Lower limit of normal (LLN) neutrophil counts in untransplanted animals is 1,800 per µl.

Platelets: The day of platelet engraftment was defined as a minimum of 20,000-50,000 platelets per µl peripheral blood for a duration of at least 7 consecutive days, and trending toward a self-sustained increase in platelet counts without transfusion reaching greater than 50,000 platelets per µl. LLN for platelet counts in un-transplanted animals is 260,000 per µl.

RNA isolation and RNA-Seq

Total RNA was extracted from sort-purified cell populations with the Arcturus PicoPure RNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's protocol.

RNA quality control

Total RNA integrity was analyzed using an Agilent 2200 TapeStation (Agilent Technologies, Inc., Santa Clara, CA) and quantified using a Trinean DropSense96 spectrophotometer (Caliper Life Sciences, Hopkinton, MA).

RNA-Seq expression analysis

RNA-Seq libraries were prepared using the TruSeq RNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA) and a Sciclone NGSx Workstation (PerkinElmer, Waltham, MA, USA). Library size distribution was validated using an Agilent 2200 TapeStation (Agilent Technologies, Santa Clara, CA, USA). Additional library QC, blending of pooled indexed libraries, and cluster optimization was performed using Life Technologies Invitrogen Qubit® 2.0 Fluorometer (Life Technologies-Invitrogen, Carlsbad, CA, USA). RNA-Seq libraries were pooled and clustered onto

a flow cell lane. Sequencing was performed using an Illumina HiSeq 2500 in rapid mode employing a paired-end, 50 base read length (PE50) sequencing strategy.

Image analysis and base calling was performed using Illumina's Real Time Analysis v1.18 software, followed by 'demultiplexing' of indexed reads and generation of FASTQ files, using Illumina's bcl2fastq Conversion Software v1.8.4.

All original RNA-Seq data were uploaded to the NCBI database (BioProject Accession codes PRJNA320857 [nonhuman primate] and PRJNA320858 [human]).

RNA-Seq data analysis

Paired-end reads were aligned to the human genome assembly (hg38) (*43*) using *bwa aln* (*44*). The human genome was chosen for its higher degree of annotation. SAM (sequence alignment map) files were converted to BAM files and subsequently merged using the *samtools view* and *merge* options, respectively (*45*). The following RNA-Seq analysis relies on the *Bioconductor* package for *R* and follows the workflow outlined by Love *et al*. (2014) (*46, 47*). The gene transfer format (GTF) file (Homo_sapiens.GRCh38.83.gtf) used to define genomic features was obtained from Ensembl and read into *R* as a transcription database object using the *GenomicFeatures* package (*48*). Using the *GenomicFeatures* package, transcripts identified in the transcription database object were organized by exons as a *GRangesList* object (*48*). Reads for each exon were counted using the *summarizeOverlaps* function from the *GenomicAlignments* package in "IntersectionNotEmpty" mode (*48*). After replacing the column data of the generated *SummarizedExperiment* object with a custom, descriptive data frame, the *SummarizedExperiment* was converted to a *DESeqDataSet* object with experimental groups corresponding to cell type and

individual using the identically named function from the *DESeq2* package (*49*). Human samples were analyzed separately to avoid forcing false positive similarities in expression.

The data sets were filtered to only include transcripts with more than one sequence read across all samples and subsequently analyzed for differential expression using the *DESeq* function of the *DESeq2* package (*49*). This creates a *DESeqDataSet* object with estimated size factors and dispersions. Data frames characterizing the comparison of $CD34^+CD117^+CD45RA^-CD90^+$ cells for nonhuman primate and $CD34^+CD133^+CD45RA^-CD90^+$ cells for human to bulk $CD34^+$ cells from the same species containing relative expression values, p-values, and adjusted p-values of each gene were extracted from the *DESeqDataSet* objects using the *results* function of the *DESeq2* package (*49*).

For all genes with sequence read counts in both the human and the nonhuman primate datasets, the relative expression value of each human gene versus each nonhuman primate gene was plotted in the background of a scatterplot in grey. Genes with p-values less than 0.05 in either the human *or* the nonhuman primate dataset were plotted in black. Finally, genes with p-value less than 0.05 in both human *and* nonhuman primate datasets were plotted in red as the top level of the scatterplot. Filtering and merging of gene lists was done with a custom R script. The *biomaRt* R package was used to map Ensembl gene IDs to external gene names where available. Finally, scatterplots were generated using the R package *ggplot2* (http://ggplot2.org/).

Insertion site analysis

Processing of gDNA (genomic deoxyribonucleic acid) to amplify integration loci included either LAM (linear-amplification mediated)-PCR (polymerase chain reaction) (*50*) (animal T04228) or MGS (modified genomic sequencing)-PCR (*51*) methods (animals M05189, J02370, Z08103 and

Z09132). A variety of next generation sequencing platforms were used depending on date of transplant and sample collection. Platforms included single end 454 GS FLX Titanium (Roche), single end Ion Torrent PGM (Personal Genome Machine), and paired end Illumina Miseq. A detailed list of samples is included in the dataset available for download (*BioProject PRJNA357116*). Insertion sites were identified using a method similar to that described by Hocum *et al*. 2015 (*52*). For Illumina data, the forward and reverse reads were stitched using PEAR with the *-q 30* option to trim sequence reads after two bases with a quality score below 30 were observed (*53*). Stitched FASTQ files and raw FASTA files were filtered using custom python scripts. *Pairwise2* from the *Bio* module was used to confirm the presence of the LTR (long terminal repeat) primer at the start of the sequence read using a gap open penalty of -2, a gap extension penalty of -1, and requiring a total mapping score of 24 or greater, equivalent to two mismatches or one insertion or deletion (indel). Presence or absence of the LTR region was determined using *Pairwise2* to align a known 22 base pair sequence from the LTR region to the sequence read using the same gap penalties described previously and requiring a total mapping score of 20 or greater, corresponding to two mismatches or one indel. To remove reads representing vector sequences (as opposed to genomic sequences) a known 24 base pair sequence from the lentivirus backbone was aligned to the sequence read using *Pairwise2* and the same settings as described for primer alignment. Reads containing the primer sequence and the LTR sequence but not the vector sequence were then trimmed at the end of the aligned LTR sequence and output in FASTQ or FASTA format depending on the input format. The *Macaca mulatta* reference genome (rheMac3, GCA_000230795.1, Oct. 2010) provided by the Beijing Genomics Institute, Shenzhen was downloaded from the UCSC genome browser (http://genome.ucsc.edu/) (*54*). Filtered and trimmed sequence reads were aligned to the reference genome using BLAT with options -

out=blast8, *-tileSize=11*, *-stepSize=5*, and *-ooc=rh11-2253.ooc* (*55*). The rh11-2253.occ file contains a list of 11-mers occurring at least 2,253 times in the genome to be masked by BLAT and was generated using the following command: $blat rheMac3.2bit /dev/null /dev/null -tileSize=11 -stepSize=5 -makeOoc=rh11-2253.ooc -repMatch=2253 as recommended by UCSC (University of California, Santa Cruz) [http://genome.ucsc.edu/goldenpath/help/blatSpec.html and http://genome.ucsc.edu/FAQ/FAQblat.html#blat6]. Resulting blast8 files were parsed using a custom python script. The blast8 files contained multiple possible alignments for each sequence read, so any sequence read with a secondary alignment with percent identity up to 95% of the best alignment was discarded. Sequence reads were then grouped based on their genomic alignment positions and orientation (sense (+) vs. antisense (-)). Any alignments within 5 base pairs of one another with identical orientations were considered to originate from the same insertion site; the genomic position with the greatest number of contributing sequence reads is reported as the insertion site. Multiple sequencing runs of the same sample were combined using a custom python script by combining the number of sequence reads for each genomic position.

Persistence analysis

For each sample, clone contributions were normalized by converting the number of sequence reads to frequencies for each sample. A filter was applied to only retain clones with a contribution of three or more sequence reads, then only those clones detected <3 months after transplant. For each subsequent sample, the sum of early engrafting clone contributions was calculated. Finally, all samples collected within the time frame were grouped for each animal. An example analysis is as follows:

In animal J02370 there are three samples available from the first 3 months after transplant. We identified 3,946, 1,633, and 93,885 insertion site-associated sequence reads in samples from 29, 60, and 64 days after transplant respectively. These insertion site-associated reads correspond to 1,615, 620, and 5,055 unique insertion site respectively. Some insertion sites were identified in multiple samples, thus a total of 6,422 unique insertion site were identified between 0 and 3 months after transplant. These 6,422 unique insertion sitess account for 100% of the insertion site-associated sequence reads observed in samples from the first 3 months after transplant. However, only 527, 209, and 2,201 insertion sites are represented by three or more sequence reads in each sample respectively. Again, some insertion sites were identified in multiple samples, thus a total of 2,634 unique insertion sites are represented by three or more sequence reads in samples from the first 3 months after transplant. These 2,634 clones are deemed early repopulating clones, and account for 63.7%, 65.8%, and 96.4% of insertion sites associated sequence reads observed in the three samples from the first 3 months after transplant, making the average frequency of early repopulating clones in this time frame 75.3%. For the second time frame of interest (>3 months after transplant, but <12 months after transplant), there is one sample available (109 days after transplant). In this sample, we found 101,926 insertion site-associated reads representing 5,681 unique insertion sites. Of the 5,681 insertion site-associated reads, 885 were also early repopulating clones, however only 776 of those were represented by three or more sequence reads. The 776 clones that are early repopulating clones and represented by three or more sequence reads account for 76.3% of insertion site-associated reads from the sample 109 days after transplant, giving us the frequency of early repopulating clones observed in the second time frame. This process is repeated for the third and fourth time frames of interest (12-24 months after transplant

and >24 months after transplant) to obtain an average frequency of early repopulating clones of 87.2% and 94.6%, respectively.

Identifying HSC signatures

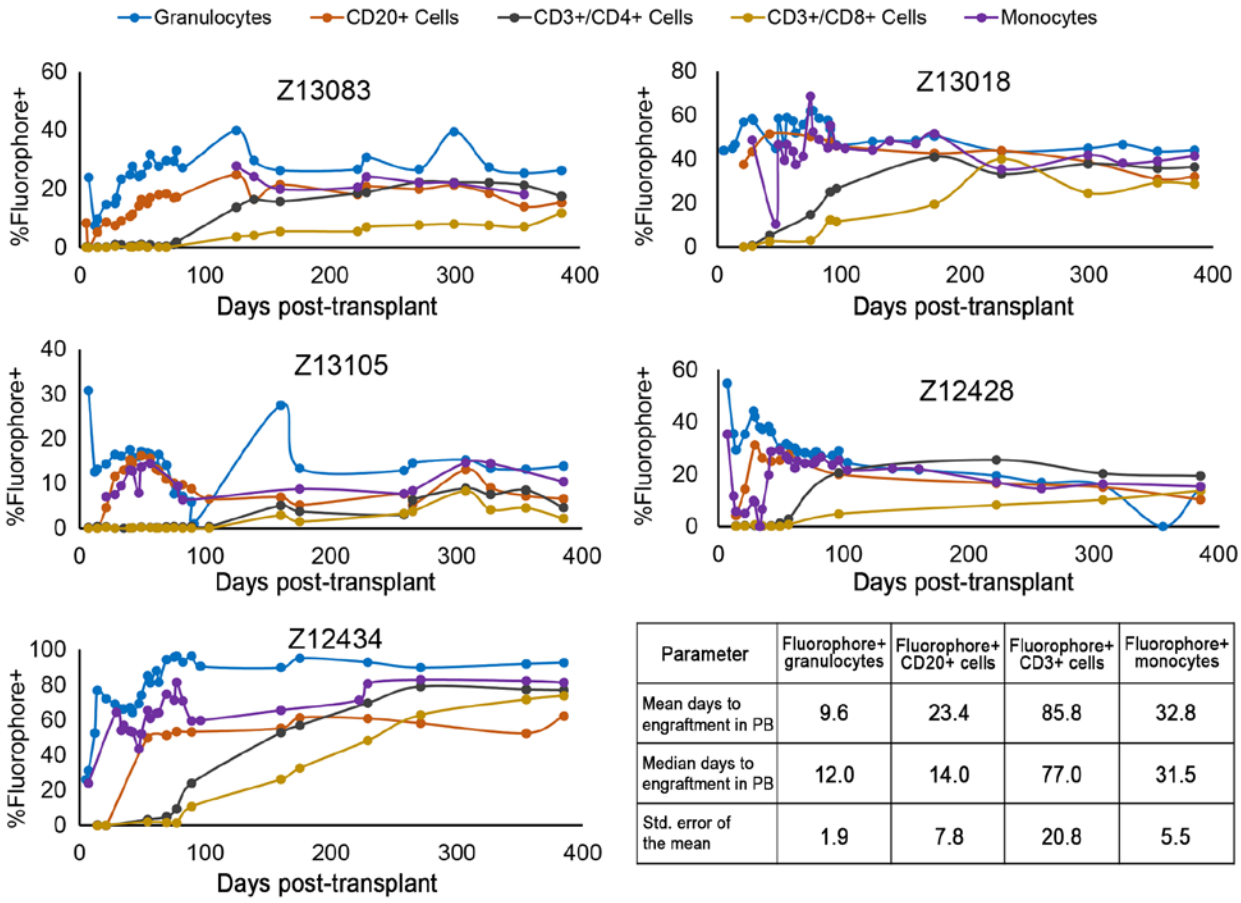
The HSC signature was defined as a clone signature present in more than one peripheral blood or bone marrow cell lineage at a single time point more than 6 months after transplant *and* detected in any other sample at any other time point.

A custom R script was used to identify HSC clone signatures, and to generate contribution graphs. All files for a given animal were read into R and sample clone contributions were normalized by converting to frequencies. Clones shared between subsets were identified as HSC clones and remaining samples were parsed to identify HSC clones present.

HSC clone contribution graphs

HSC clone contributions were determined by dividing the number of sequence reads for each clone by the total number of HSC clone sequence reads for the sample. Unallocated clones were noted. The mean maximum frequency of HSC clones was calculated. One animal, Z09132, had 676 clones contributing above the mean maximum frequency of HSC clones, thus only the top 100 most abundant HSC clones were displayed due to color limitations. A data frame was constructed including three columns: (1), days after transplant; (2), clone frequency in the sample; (3), clone identifier (genomic locus or barcode). The data frame was visualized using the R package *ggplot2*.

SUPPLEMENTAL FIGURES



Parameter	Fluorophore+ granulocytes	Fluorophore+ CD20+ cells	Fluorophore+ CD3+ cells	Fluorophore+ monocytes
Mean days to engraftment in PB	9.6	23.4	85.8	32.8
Median days to engraftment in PB	12.0	14.0	77.0	31.5
Std. error of the mean	1.9	7.8	20.8	5.5

Fig S1. ***(Related to Figure 1.)*** **Engraftment kinetics in transplanted nonhuman primates.** The frequency of fluorophore$^+$ peripheral blood (PB) subsets including granulocytes (blue), monocytes (purple), B cells (CD20$^+$ cells; orange) and T cells (CD3$^+$/CD4$^+$ cells (gray) and CD3$^+$/CD8$^+$ cells (yellow)) measured by flow cytometry over 1+ year after transplant in the myeloablative setting in five macaques. The table summarizes engraftment data as numeric values. The day of engraftment for each lineage is defined as the first day ≥1% fluorophore$^+$ cells were observed in peripheral blood with consecutive increases over the next three measurements.

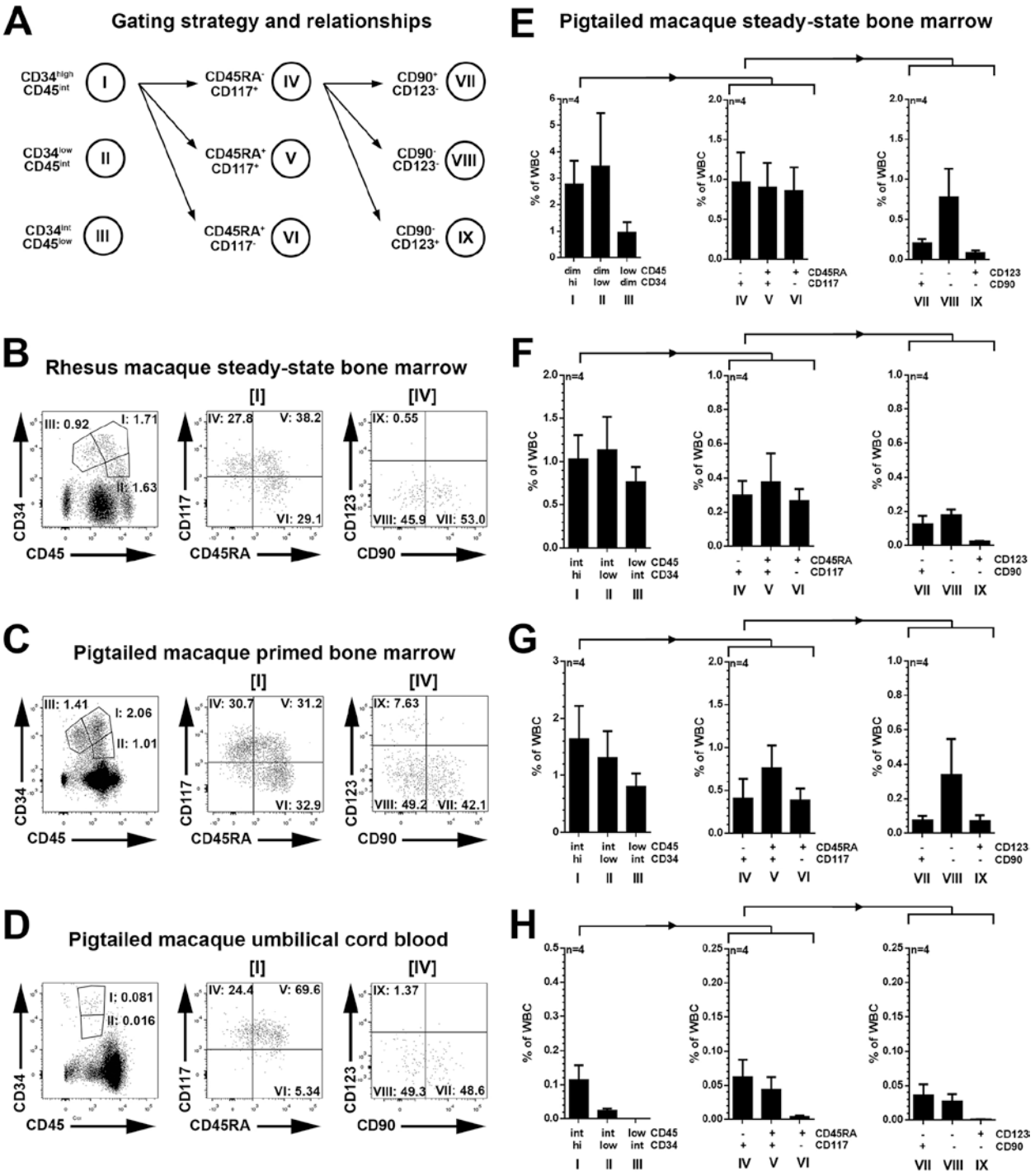


Fig S2. ***(Related to Figure 2.)*** **HSPC composition in nonhuman primate stem cell populations from different sources.**

(A) Gating strategy for nonhuman primate HSPC subpopulations. Cells were first subdivided into three populations showing low, intermediate (int), or high expression of CD45 and CD34: $CD34^{high}CD45^{int}$ (I), $CD34^{low}CD45^{int}$ (II) and $CD34^{int}CD45^{low}$ (III). From population I we identified three subdivisions: $CD45RA^{-}CD117^{+}$ (IV), $CD45RA^{+}CD117^{+}$ (V) and $CD45RA^{+}CD117^{-}$ (VI) cells. Population IV was further subdivided into $CD90^{+}CD123^{-}$ (VII),

$CD90^{-}CD123^{-}$ (VII) and $CD90^{-}CD123^{+}$ (IX) cells. **(B-D)** Cell surface expression of CD45 vs. CD34, CD45RA vs. CD117 and CD90 vs. CD123 on **(B)** rhesus macaque steady-state bone marrow, **(C)** pigtailed macaque primed bone marrow and **(D)** pigtailed macaque umbilical cord blood white blood cells. Within shown stem cell sources, CD45RA/CD117 expression is gated on $CD34^{high}CD45^{int}$ (I) populations; CD90/CD123 analysis is shown for $CD34^{high}CD45^{int}CD45RA^{-}CD117^{+}$ (IV) cells. **(E-H)** Average frequency of CD34-subpopulations I-IX (defined in Figure 1) in **(E)** pigtailed macaque steady-state bone marrow white blood cells, **(F)** rhesus macaque steady-state bone marrow white blood cells, **(G)** pigtailed macaque primed bone marrow and **(H)** pigtailed macaque umbilical cord blood. (All values given as mean ± SEM)

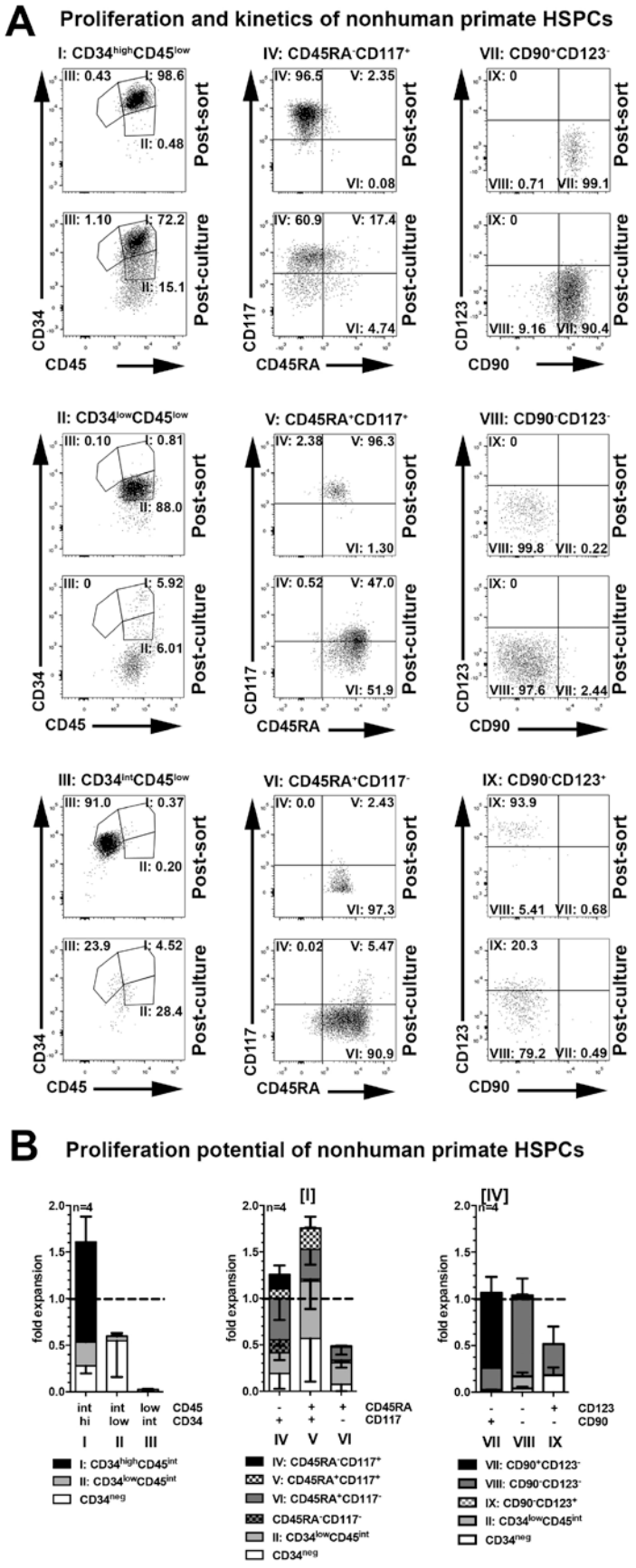


Fig S3. ***(Related to Figure 2.) Ex vivo*** **culture of nonhuman primate HSPCs.** **(A)** $CD34^{high}CD45^{low}$ (I), $CD45RA^{-}CD117^{+}$ (IV) and $CD90^{+}CD123^{-}$(VII), $CD34^{low}CD45^{low}$ (II), $CD34^{int}CD45^{low}$ (III), $CD45RA^{+}CD117^{+}$ (V), $CD45RA^{+}CD117^{-}$ (VI), $CD90^{-}CD123^{-}$ (VIII) and $CD90^{-}CD123^{-}$ (IX) populations from pigtailed macaque steady-state bone marrow were sort-purified (upper row, post sort) and cultured in StemSpan supplemented with either SCF, TPO and FLT3-L (II, III, VIII and IX) or SCF and IL-3 (V and VI). Arising progeny were analyzed on day 3 (lower row, post culture). **(B)** Fold expansion (total height of bars) and phenotypical composition of progeny (color-coded, stacked) derived from sort-purified pigtailed macaque steady-state bone marrow CD34-subpopulations cultured in StemSpan supplemented with either SCF, TPO and FLT3-L (I, II, III, VII, VIII and IX) or SCF and IL-3 (IV, V and VI) for 3 days. (mean ± SEM)

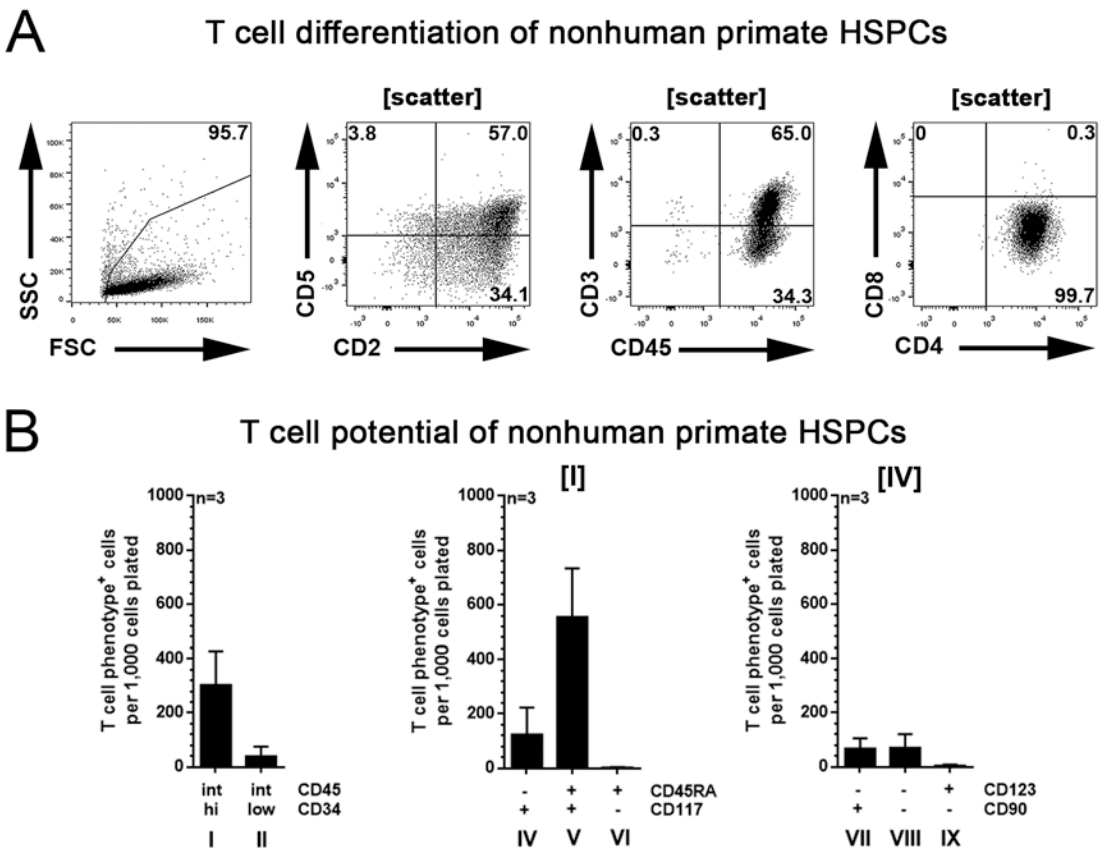


Fig S4. ***(Related to Figure 2.)*** **T cell potential of nonhuman primate HSPCs.**
(A) Detailed gating-strategy for the flow cytometric analysis and quantification of *in vitro* differentiated mature and immature T cells. Within T cell assays hematopoietic cells expressing the T cell marker CD3, CD4 or CD8 were gated and quantified separately. **(B)** Quantification of *in vitro* T cells differentiation potential within sort-purified pigtailed macaque steady-state bone marrow CD34-subpopulations. (mean ± SEM)

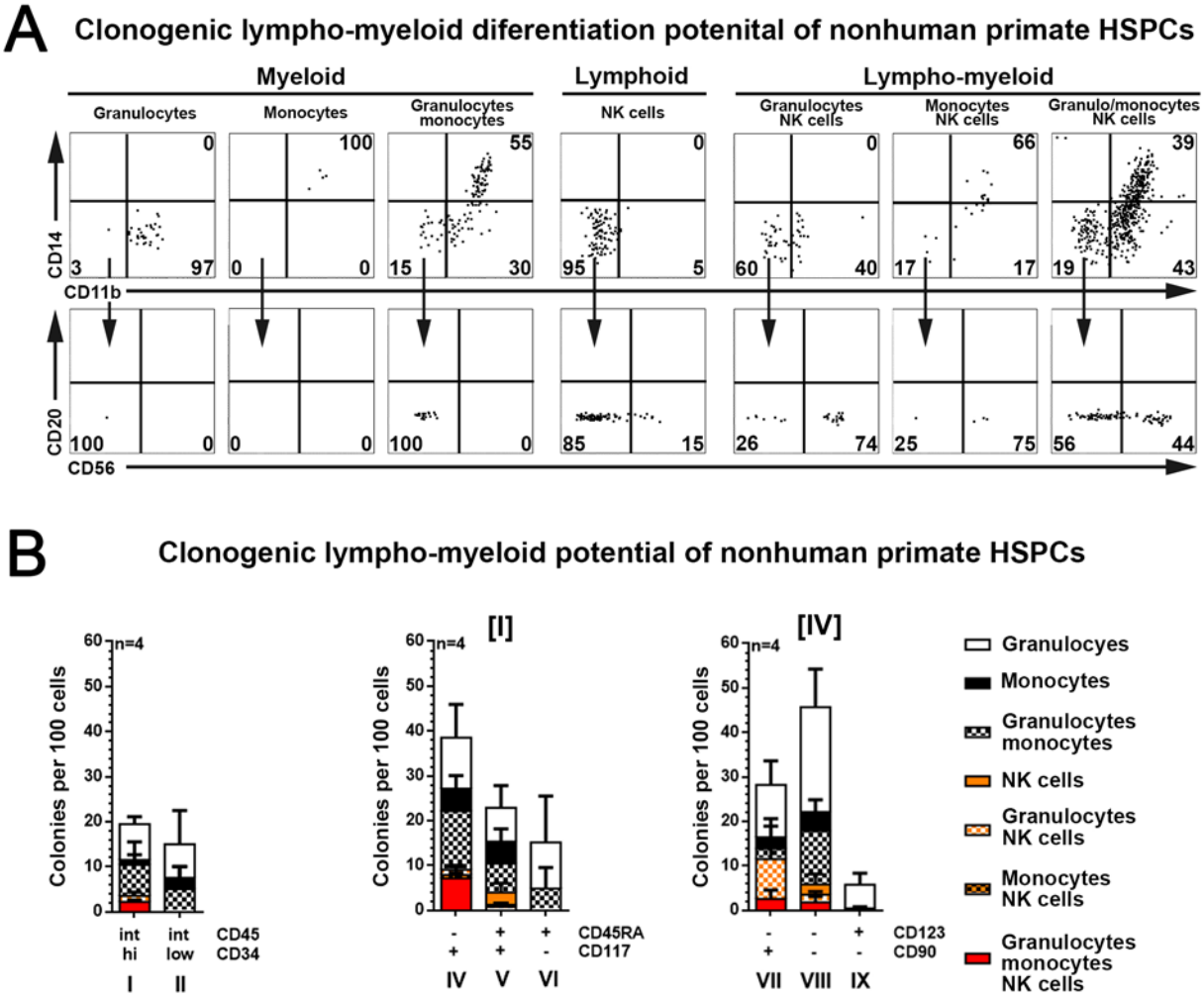


Fig S5. ***(Related to Figure 2.)*** **Lympho-myeloid potential of nonhuman primate HSPCs.** **(A)** Representative flow-cytometric analysis of colonies with myeloid, lymphoid or myeloid/lymphoid blood cells in the MS-5 assay. Upper row (CD11b vs. CD14) is gated on $CD45^+$ nonhuman primate cells to exclude MS-5 stroma cells. The second row (CD20 vs. CD56) is gated on $CD11b^-CD14^-$ cells from the upper panel. The bottom row (CD20 vs. CD34) is gated on $CD11b^-CD14^-CD20^-CD56^-$ hematopoietic cells. Numbers in flow-plots indicate percentage of marker positive cells. 96-wells with a minimum of five $CD45^+$ nonhuman primate cells were counted as positive. **(B)** Frequency of progenitor cells with clonogenic myeloid (granulocyte, macrophage, granulocyte-macrophage), lymphoid (NK cell, B cell), and lympho-myeloid (granulocyte-NK, monocyte-NK, granulocyte-monocyte-NK) differentiation potential. (mean ± SEM)

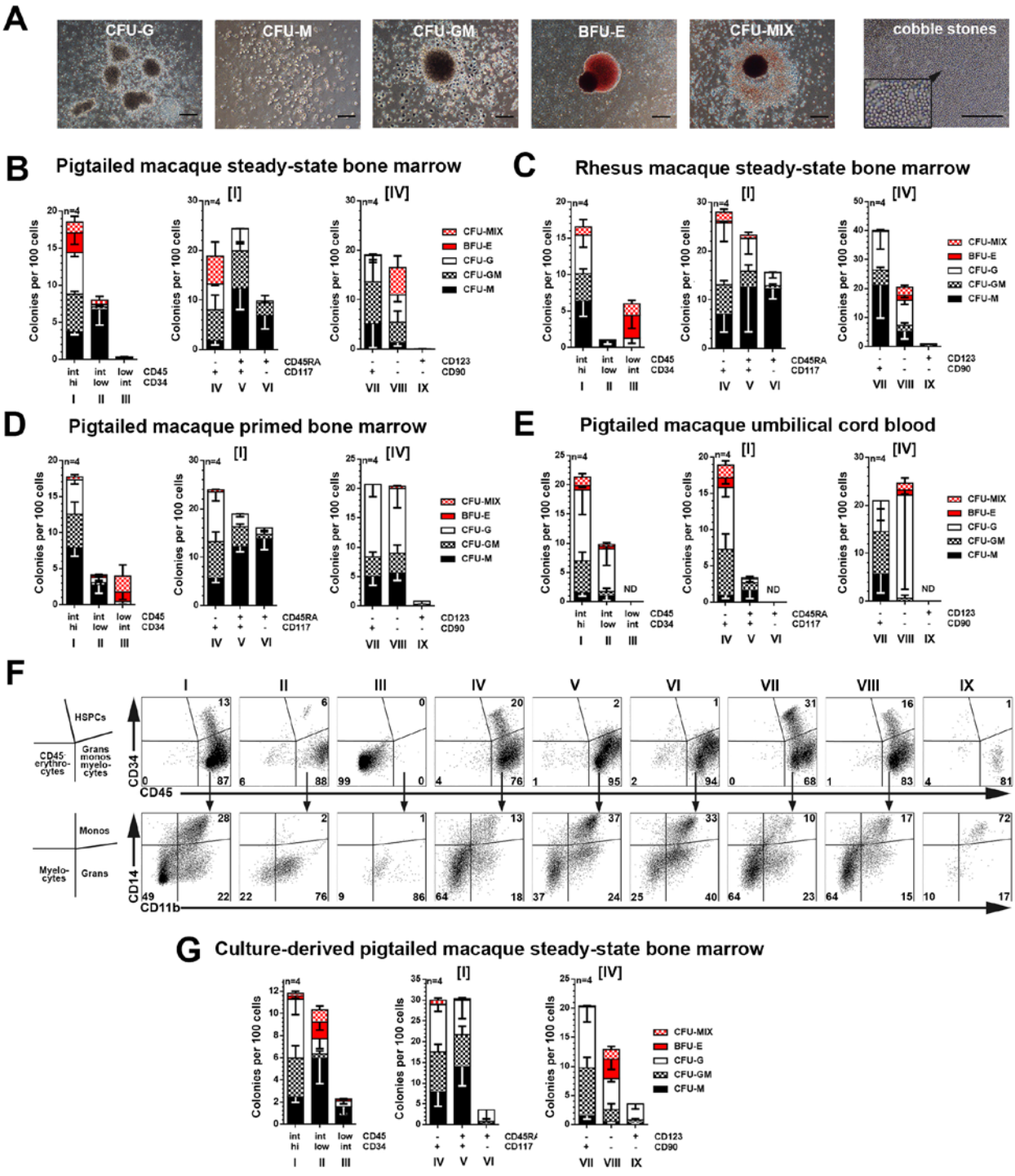


Fig S6. ***(Related to Figure 2.)*** **Colony-forming cell potential of nonhuman primate HSPCs.** **(A)** Representative images of different colony-subtypes of sort-purified pigtailed macaque steady-state bone marrow-derived CD34^{+} cells. Arising colonies were identified as colony forming unit-(CFU) granulocyte (CFU-G), macrophage (CFU-M), granulocyte-macrophage (CFU-GM) and burst forming unit-erythrocyte (BFU-E). Colonies consisting of erythroid and myeloid cells were scored as CFU-MIX (scale bar = 100 µm). **(B-E)** Colony-forming frequency of **(B)** pigtailed

macaque steady-state bone marrow, **(C)** rhesus macaque steady-state bone marrow and **(D)** pigtailed macaque primed bone marrow, **(E)** pigtailed macaque umbilical cord blood-derived CD34-subpopulations I-IX. The different colony-subtypes are color-coded as indicated. **(F)** To confirm counting of primary colony-forming cell assays from freshly isolated nonhuman primate-HSPCs, total colonies were harvested, washed and stained for flow cytometry. Colony-forming cell assay contained erythrocytes ($CD45^-$), HSPCs ($CD34^+CD45^+$), monocytes ($CD11b^+CD14^+CD34^-CD45^+$) and granulocytes ($CD11b^+CD14^-CD34^-CD45^+$). **(G)** Colony-forming frequency of CD34-subpopulations from pigtailed macaque steady-state bone marrow cultured in StemSpan supplemented with either SCF, TPO and FLT3-L (I, II, III, VII, VIII and IX) or SCF and IL-3 (IV, V and VI) for 3 days.

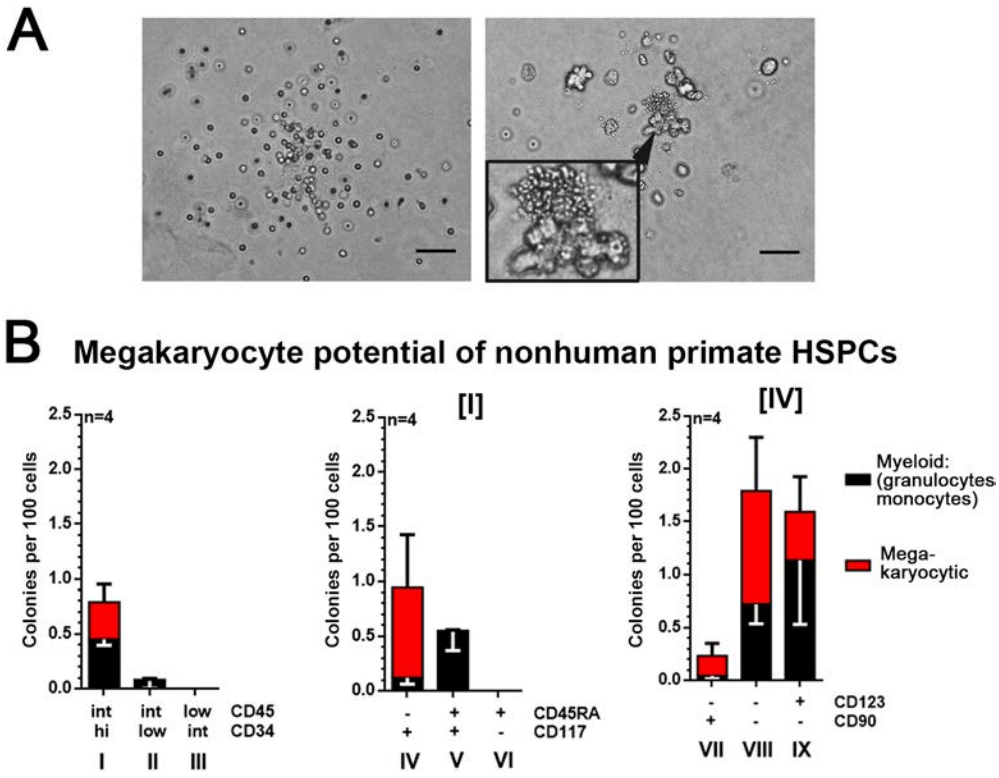


Fig S7. ***(Related to Figure 2.)*** **Megakaryocyte potential of nonhuman primate HSPCs.**
(A) Representative images of myeloid (upper picture, scale bar 100µm) and megakaryocytic (lower picture, scale bar 100µm, inlet 40x magnification) colonies generated in MegaCult assays. **(B)** Frequency of progenitor cells with clonogenic myeloid (black bars) and megakaryocyte (red bars) colony-forming potential. (mean ± SEM)

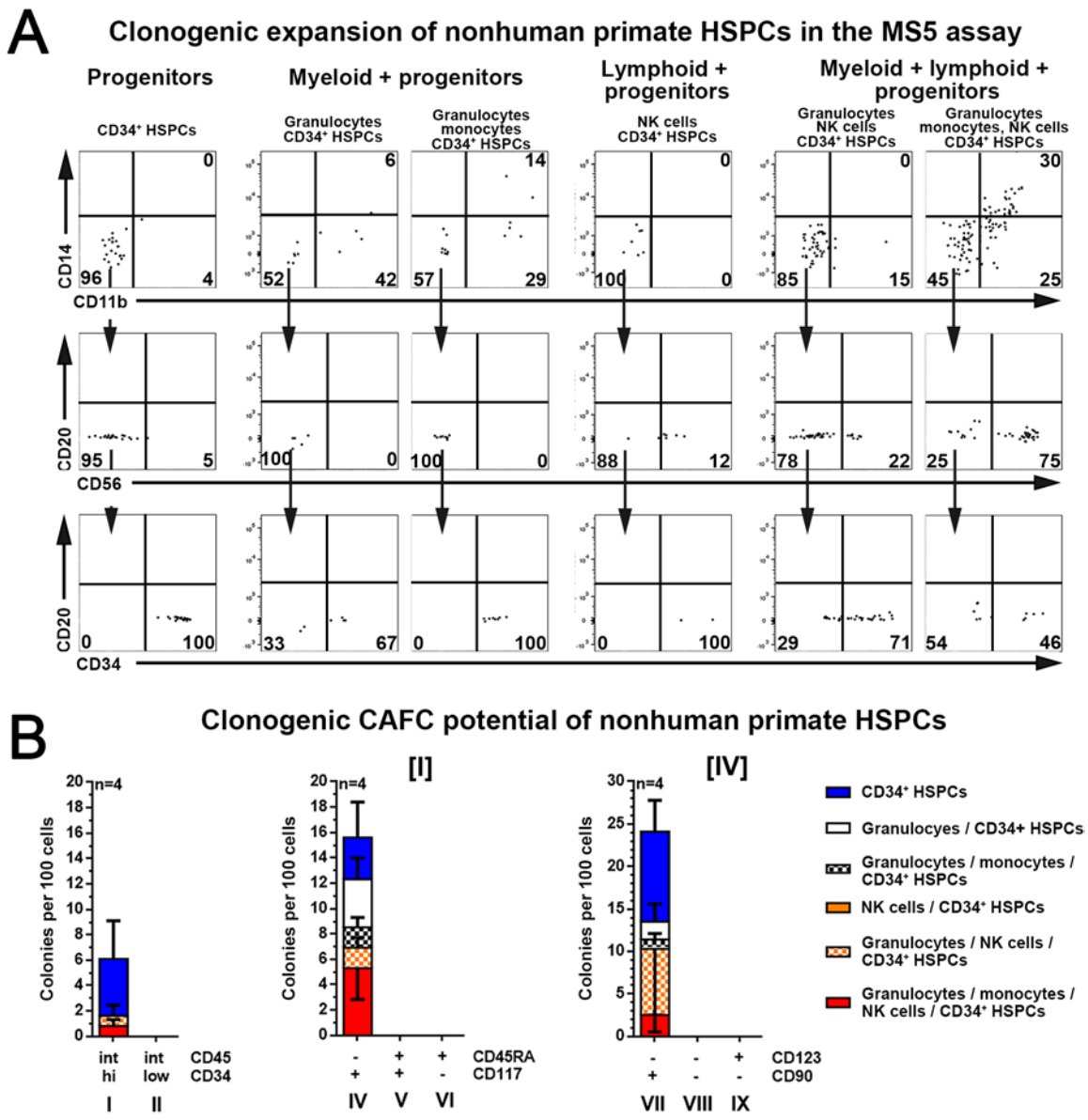


Figure S8. ***(Related to Figure 2.)*** **Cobblestone area-forming cell potential of nonhuman primate HSPCs.**

(A) Representative flow-cytometric analysis of colonies with myeloid, lymphoid or myeloid/lymphoid blood cells in the MS-5 assay. Upper row (CD11b vs. CD14) is gated on $CD45^+$ nonhuman primate cells to exclude MS-5 stroma cells. The second row (CD20 vs. CD56) is gated on $CD11b^-CD14^-$ cells from the upper panel. Additional acquisition of colonies containing $CD34^+$ progenitor cells. The bottom row (CD20 vs. CD34) is gated on $CD11b^-CD14^-CD20^-CD56^-$ hematopoietic cells. Numbers in flow-plots indicate percentage of marker positive cells. 96-wells with a minimum of five $CD45^+$ nonhuman primate cells were counted as positive (**Table S3**, number in parentheses). **(B)** Frequency of progenitor cells in pigtailed macaque steady-state bone marrow-derived CD34-subpopulations I-IX with myeloid, lymphoid or lympho-myeloid differentiation potential also containing undifferentiated $CD34^+$ HSPCs. (mean ± SEM)

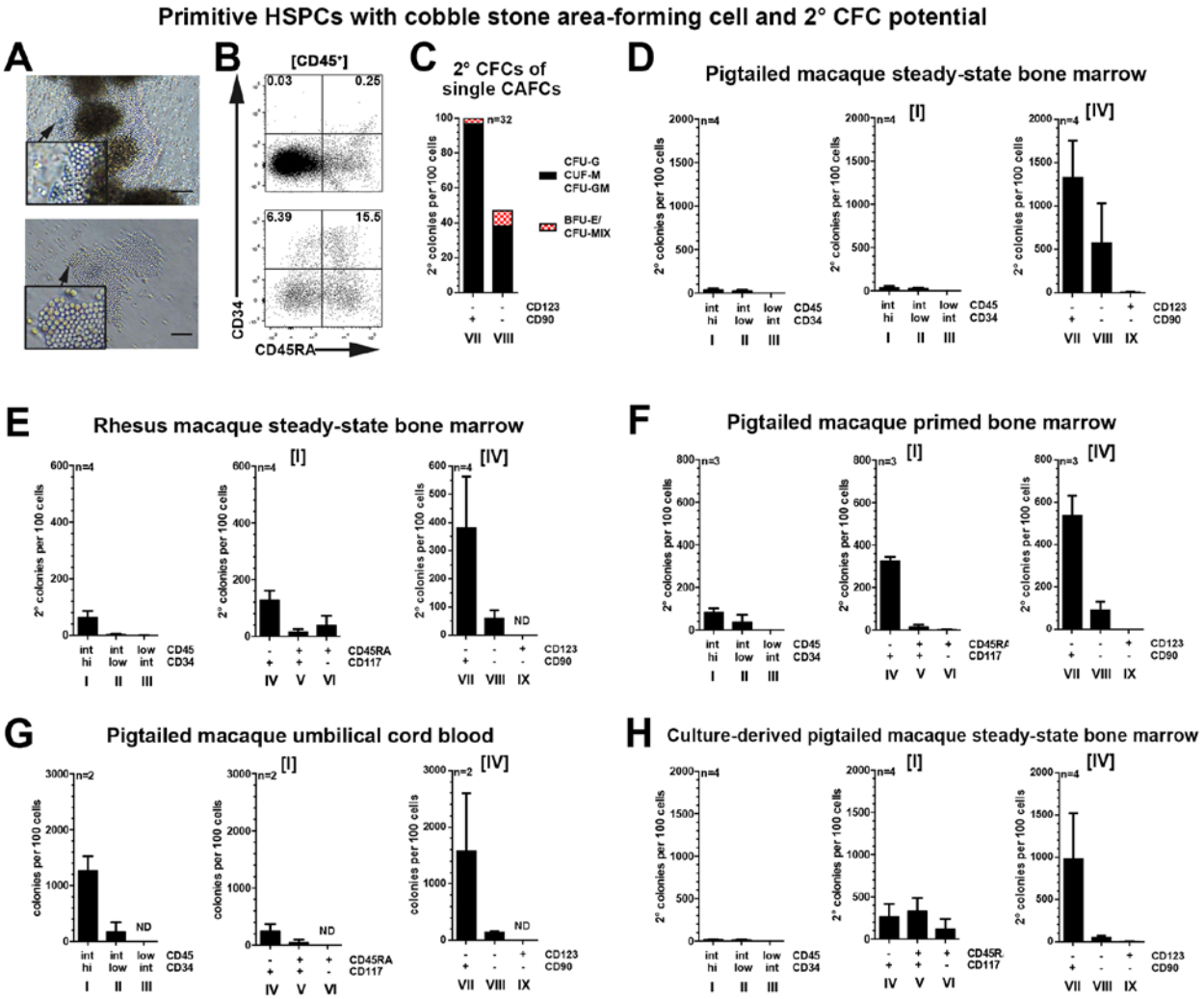


Fig S9. ***(Related to Figure 2.)*** **Secondary colony-forming cell potential of nonhuman primate HSPCs.**

(A) Representative images of cobblestone area-forming cells. Cobblestones are either associated with differentiated colonies (upper picture, scale bar 100µm) or autonomous (lower picture, scale bar 20µm, inlay 100x magnification). **(B)** Cobblestones associated with colonies (upper plot) and autonomous (lower plot) were harvested for analysis. After exclusion of $CD45^-$ erythrocytes, expression of the progenitor markers CD34 and CD45RA were analyzed. **(C)** Single cobblestones from primary colony-forming cell assays of populations VII and VIII were harvested, washed, and re-plated to evaluate secondary colony-forming potential evaluating myeloid (G/M/GM) and erythroid/erythro-myeloid (BFU-E/CFU-MIX) differentiation potential. (**D-H)** Total primary colony-forming cell assays were harvested, washed, and cells re-plated to evaluate secondary colony-forming potential. Within secondary colony-forming cell assays, only total colony-forming potential was evaluated without discrimination of colony-subtypes. Secondary colony-forming cell assays of **(D)** pigtailed macaque steady-state bone marrow, **(E)** rhesus macaque steady-state bone marrow, **(F)** pigtailed macaque primed bone marrow, **(G)** pigtailed macaque umbilical cord blood, and **(H)** culture-derived pigtailed macaque steady-state bone marrow CD34-subpopulations were harvested, washed and cells re-plated to evaluate secondary colony-forming potential. Within

secondary colony-forming cell assays, total colony-forming potential was evaluated without discrimination of colony-subtypes. (mean ± SEM)

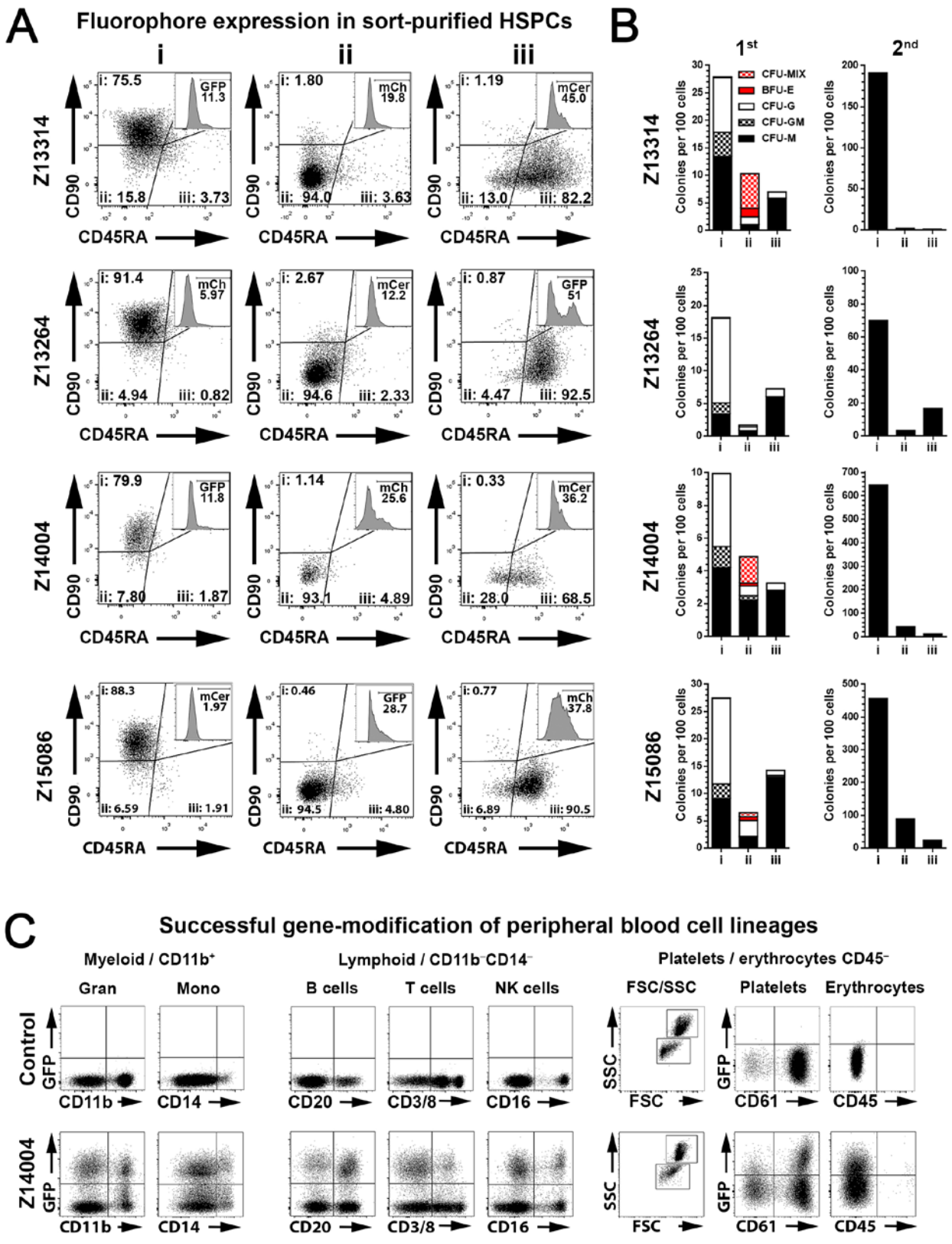


Fig S10. ***(Related to Figure 3.)*** **Quality control for competitive repopulation experiments.** **(A)** Representative flow-cytometry of sorted and lentivirus transduced fractions *i*, *ii* and *iii* of the

infusion product immediately before pooling and autologous infusion. The transduction efficiency (% cells expressing respective fluorochrome) of fraction *i*, *ii* and *iii* 24 hours after transduction is indicated in the histogram inlay in the top-right corner of each flow plot. **(B)** Fraction *i*, *ii* and *iii* cells were introduced into colony-forming cell assays after transduction culture for 14 days and the frequency of primary colony-forming cell assays (1^{st}) was determined by counting arising colonies. The different colony-subtypes are color-coded as indicated in the figure legend. Total primary colony-forming cell assays were harvested, washed, and re-plated to evaluate secondary colony-forming cell potential (2^{nd}). Within secondary colony-forming cell assays, total colony-forming cell potential was evaluated without discrimination of colony-subtypes. **(C)** Representative flow-cytometric analysis of gene-marked granulocytes ($CD11b^+CD14^-$), monocytes ($CD11b^+CD14^+$), B cells ($CD20^+$), T cells ($CD3^+$), NK cells ($CD16^+$), platelets ($CD45^-CD61^+$) and erythrocytes ($CD45^-$) in peripheral white blood cells of animal Z14004 within first 6 months after transplant.

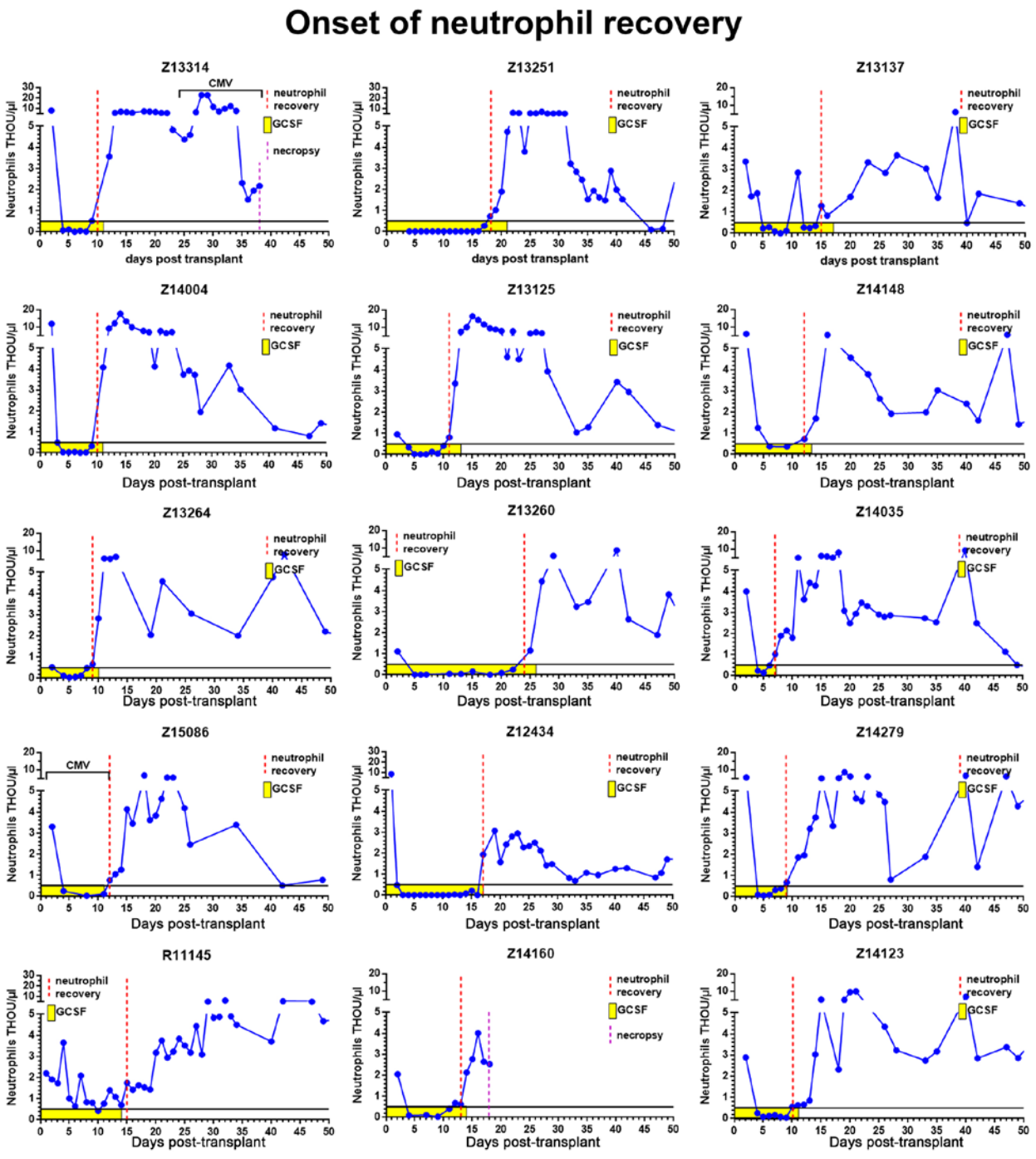


Fig S11. ***(Related to Figures 3. and 5.)*** **Neutrophil recovery in transplanted nonhuman primates.**

Tracking of peripheral blood neutrophil counts in 15 nonhuman primates following myeloablative total body irradiation and autologous hematopoietic cell transplantation (more details in **Table S4**, **S5** and **S6**). Dashed red line: day of neutrophil engraftment; lower horizontal black bar – LLE: defined minimum number of neutrophils (500/µl) for the onset of recovery; upper horizontal black bar – LLN; yellow bar: duration of G-CSF administration.

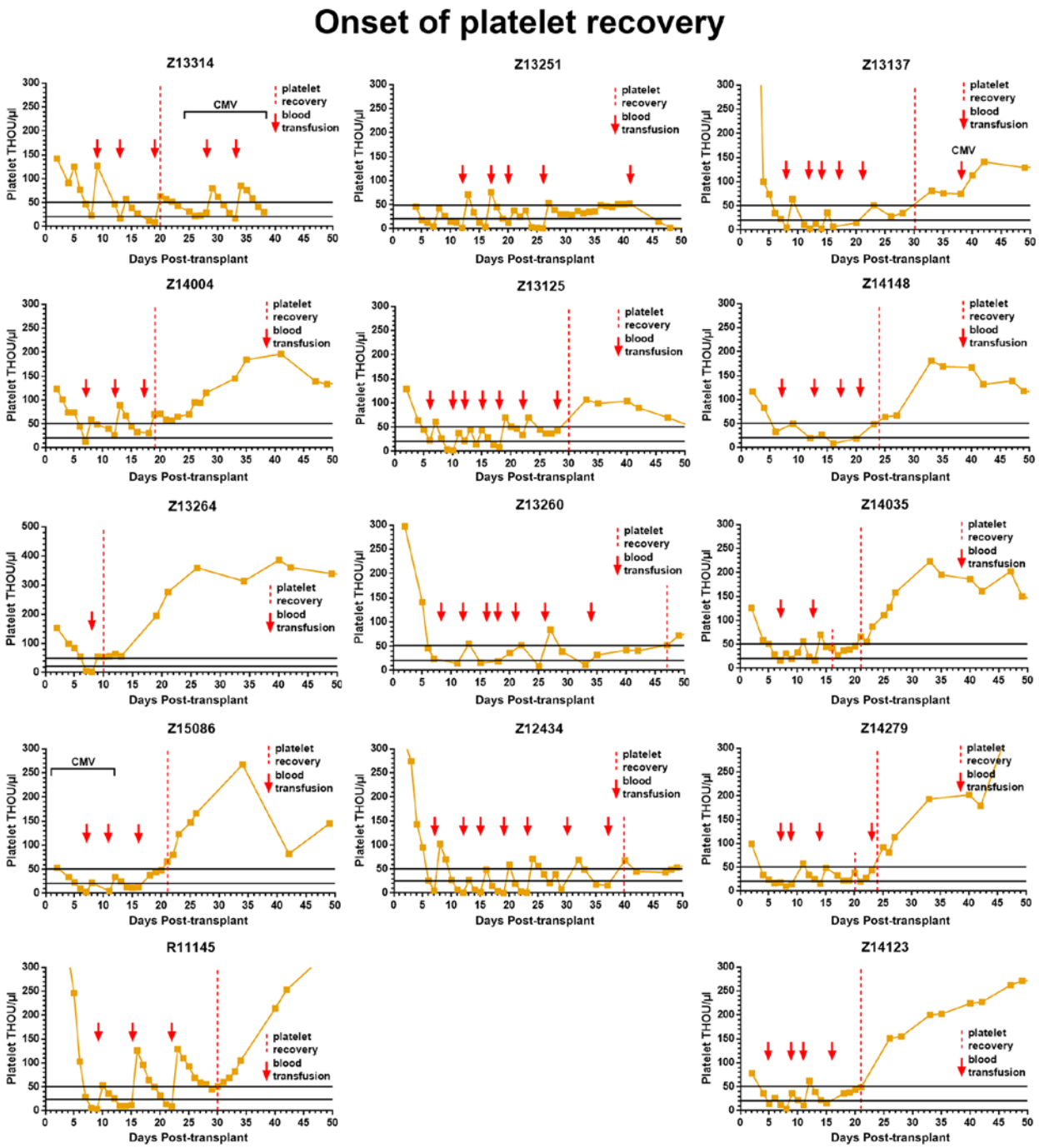


Fig S12. ***(Related to Figures 3. and 5.)*** **Platelet recovery in transplanted nonhuman primates.** Tracking of peripheral blood platelet counts in 14 nonhuman primates following myeloablative total body irradiation and autologous hematopoietic cell transplantation (more details in **Table S4**, **S5** and **S6**). Dashed red line: day of platelet engraftment; lower horizontal black bar – LLE: defined minimum number of platelets (20,000-50,000/µl) for the onset of recovery; upper horizontal black bar – LLN; red arrows: platelet transfusions.

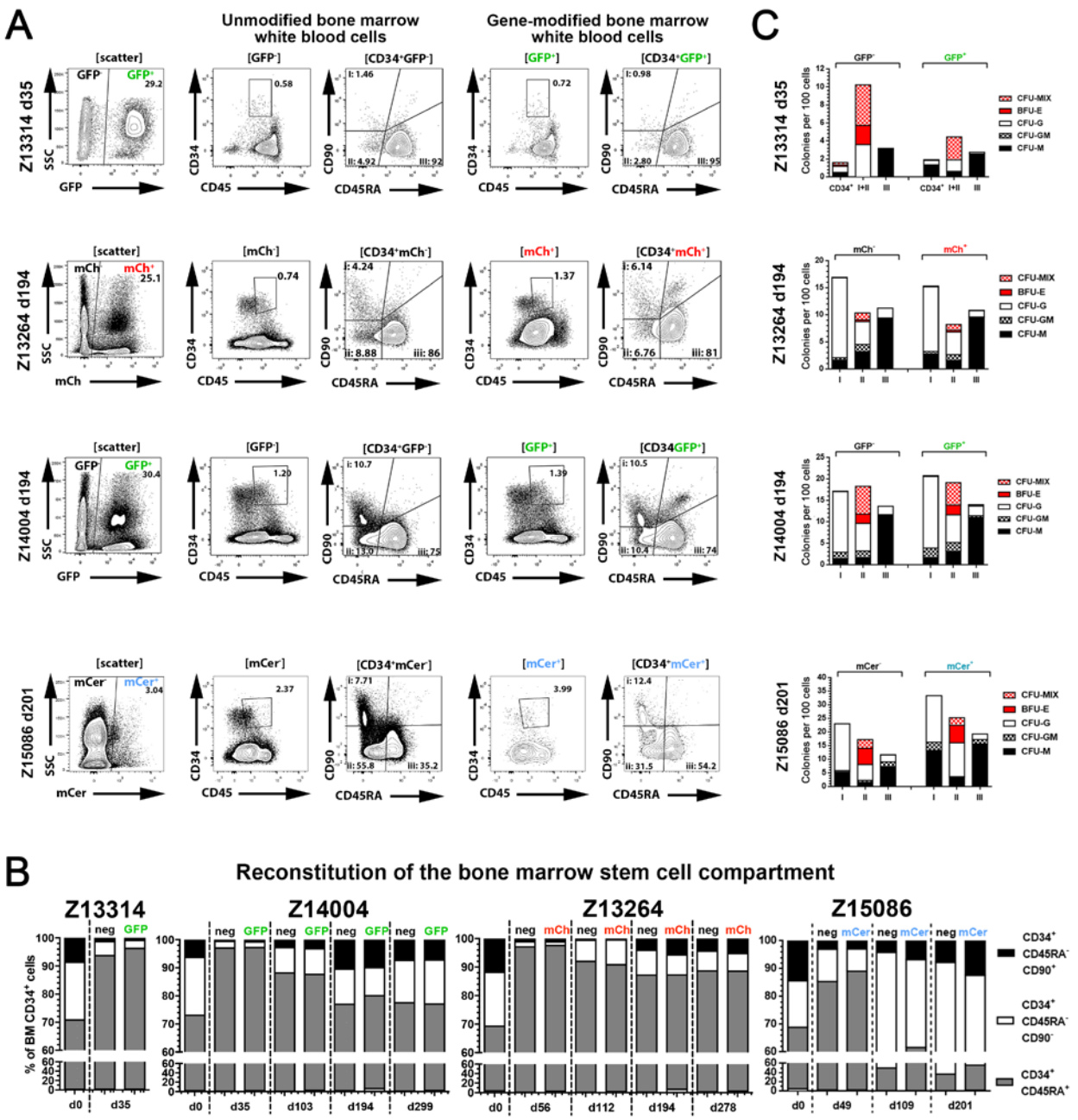


Fig S13. ***(Related to Figure 3.)*** **Bone marrow engraftment of gene-modified HSPCs.**
(A) Flow-cytometric analysis of engrafted unmodified and gene-modified white blood cells (scatter vs. GFP/mCherry/mCerulean) and $CD34^+$ HSPC subpopulations in the bone marrow of Z13314, Z14004, Z13264 and Z15086 approximately 6 months post infusion. **(B)** Unmodified and gene-modified bone marrow-derived HSPC fractions *i*, *ii* and *iii* from Z13314, Z14004, Z13264 and Z15086 were sort-purified and introduced into colony-forming cell assays. After 14 days, the

frequency of primary colony-forming cell assays was determined by counting arising colonies. The different colony-subtypes are color-coded as indicated in the figure legend. **(C)** Frequency of unmodified and gene-modified HSPC fractions *i*, *ii* and *iii* in the bone marrow of all four animals over time.

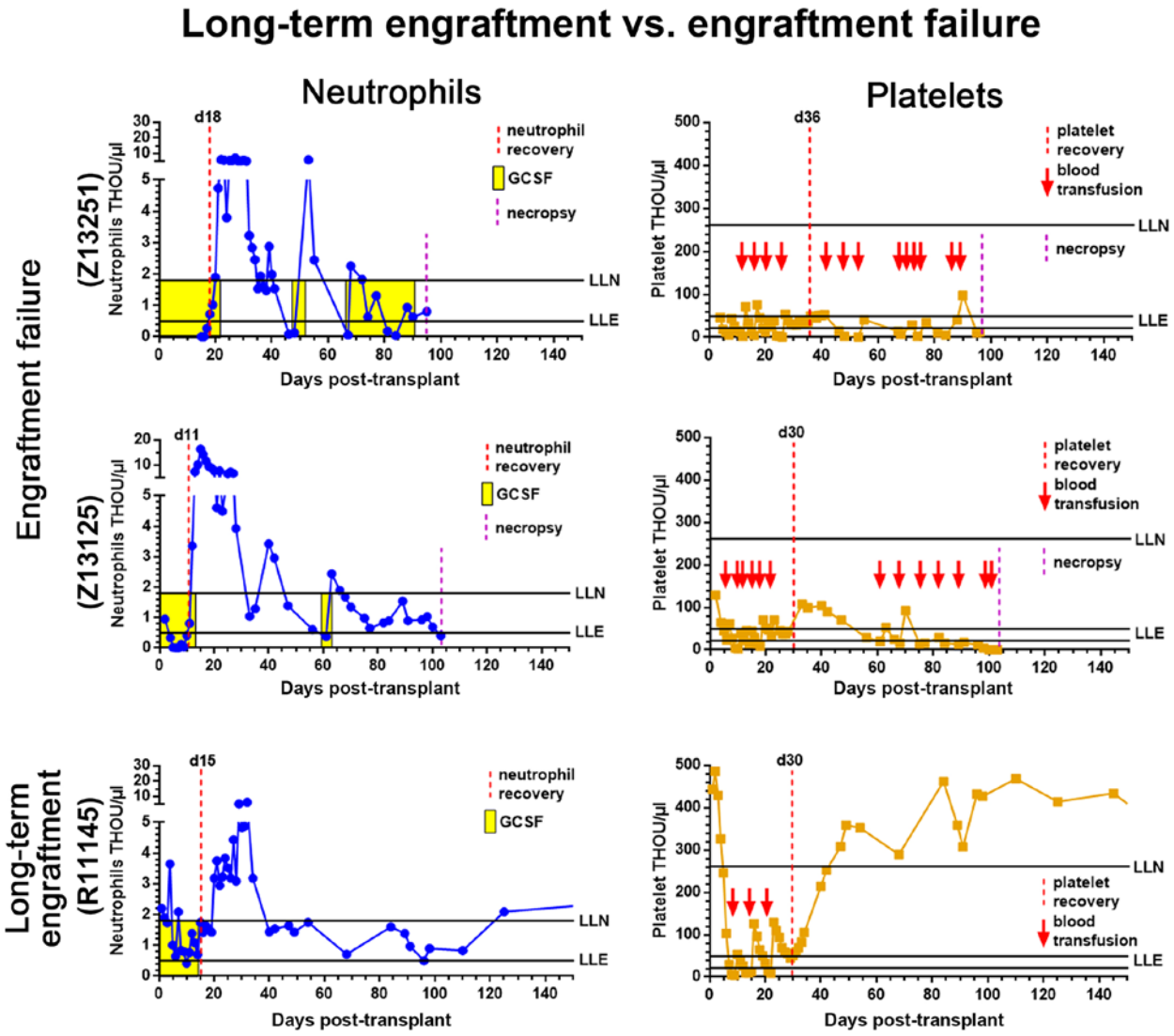


Fig S14. ***(Related to Figure 5.)*** **Long-term engraftment vs. failure.**
Representative tracking of neutrophil and platelet counts following myeloablative total body irradiation and lentivirus gene modified autologous HSC transplant from two individual nonhuman primates showing engraftment failure (animal Z13251 and Z13125) or successful long-term engraftment (animal R11145). Dashed red line: day of neutrophil/platelet engraftment; dashed purple line: day of necropsy; lower horizontal black bar – LLE: defined minimum number of neutrophils (500/µl) and platelets (20,000-50,000/µl) for the onset of recovery; upper horizontal black bar – LLN. Yellow bar: duration of G-CSF administration; red arrows: platelet transfusions.

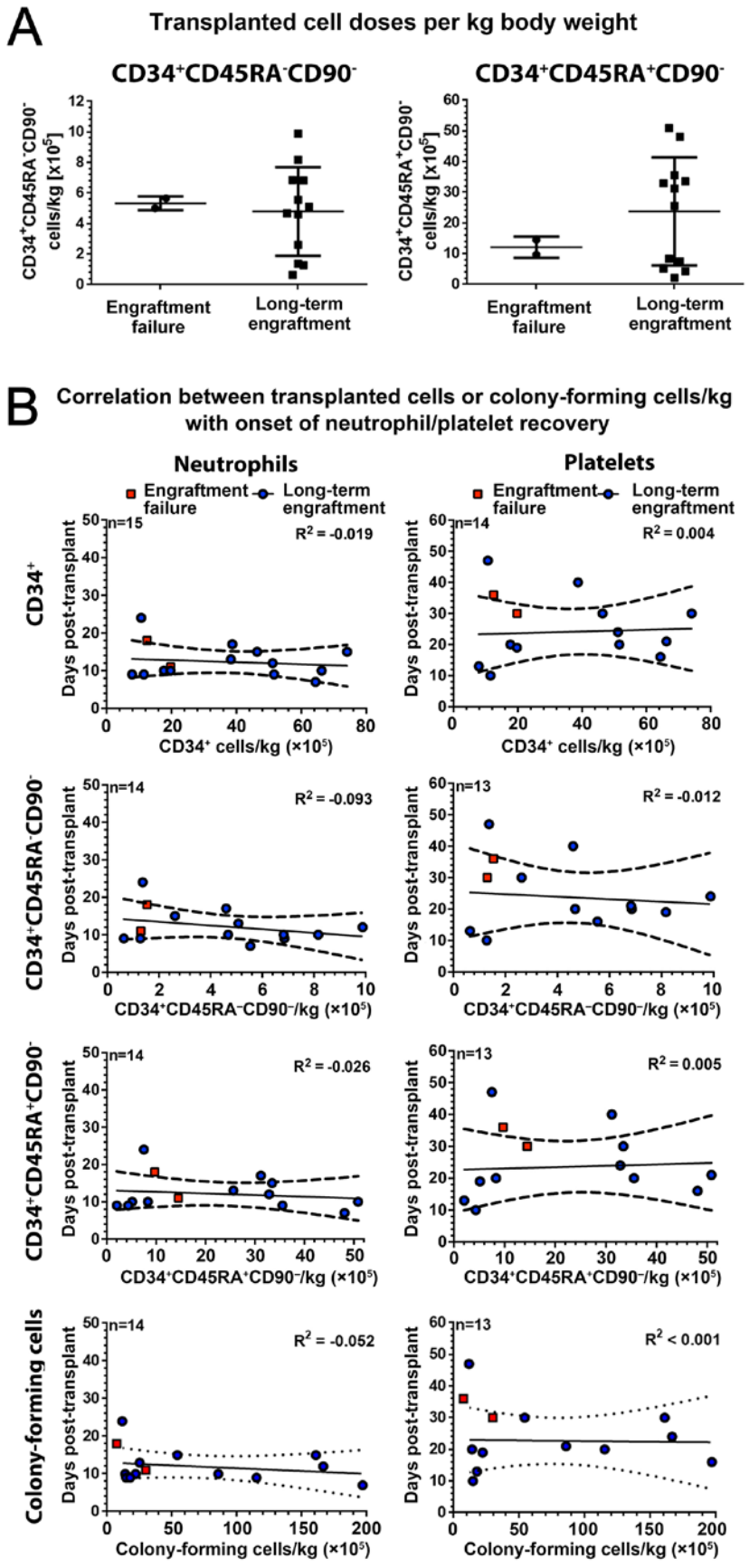


Fig S15. ***(Related to Figure 5.)*** **Correlation of recovery with infused HSPCs and colony-forming cells.**
(A) Statistical comparison of transplanted $CD34^+CD45RA^-CD90^-$ and $CD34^+CD45RA^+CD90^-$ cells/kg body weight in animals with engraftment failure and long-term engraftment. **(B)** Linear correlation of transplanted $CD34^+$ HSPCs and total colony-forming cells/kg body weight with the day of neutrophil and platelet engraftment was calculated using Spearman´s rank correlation coefficient. Only nonhuman primates demonstrating long-term engraftment were included in this analysis. The linear regression and the 95% confidence interval for each correlation are indicated with the solid and the dotted line, respectively.

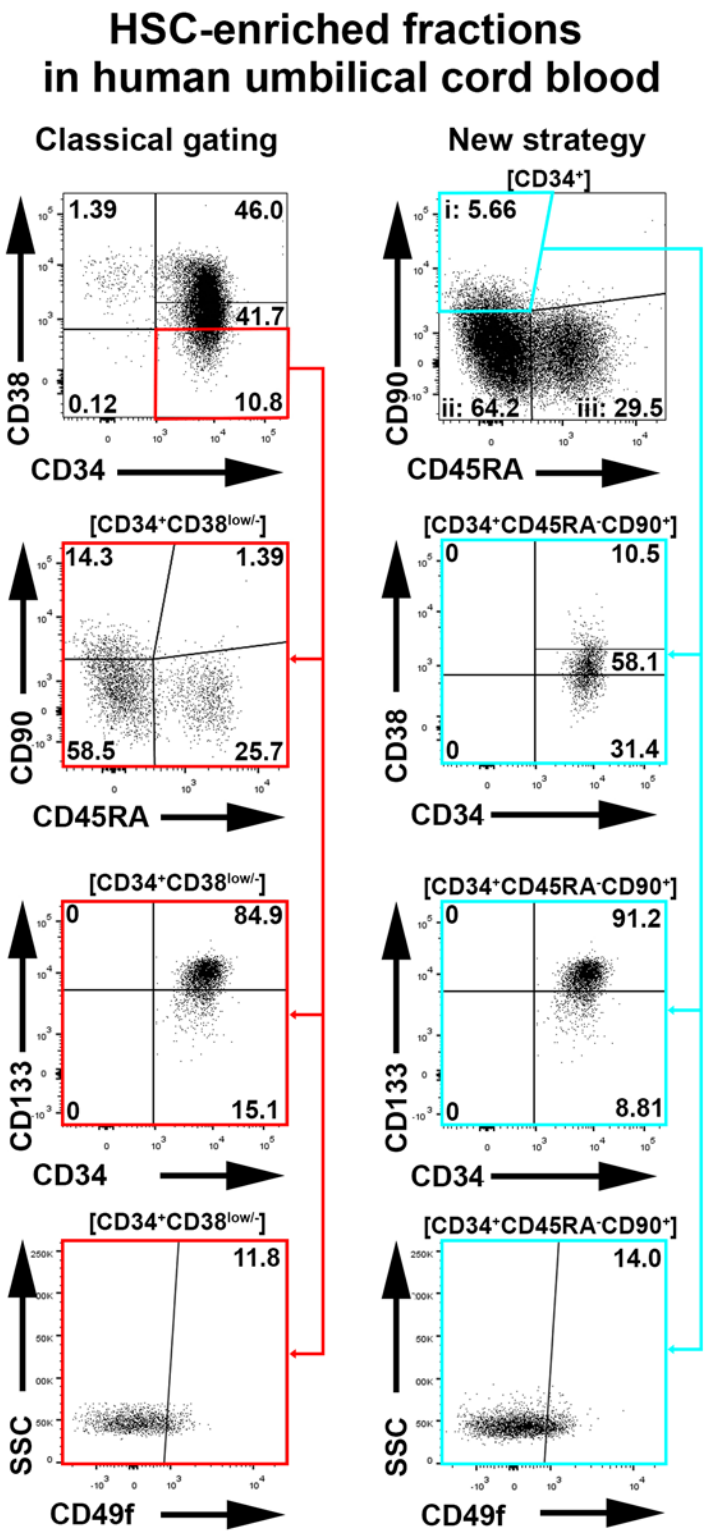


Fig S16. ***(Related to Figure 6.)*** **Gating of HSC-enriched cell fractions in human umbilical cord blood.**
Gating of an HSC-enriched cell fraction in human umbilical cord blood stem cells using the classical (left) and our new (right) gating strategy for a $CD34^{+}CD45RA^{-}CD90^{+}$ subfraction. Additional gating of $CD34^{+}CD38^{low/-}$ and $CD34^{+}CD45RA^{-}CD90^{+}$ cell fraction for expression of CD133 and CD49f (bottom two plots).

SUPPLEMENTAL TABLES

Table S1. *(Related to Figure 1.)* Identification of clones displaying HSC biology *in vivo*.

Animal	Total No. Unique Clones Identified	Sorted Hematopoietic Cell Subsets Available	No. of Subset Sorts Performed	Follow-up time Point at Which Sort(s) Performed (days after transplant)	No. of HSC Clones Identified	No. HSC Clones Not Found in PB	Mean Max. Freq. of HSC Clones in PB	No. of High Abundance HSC Clones
J02370	32,076	T + Mono	1	2482	291	0 (0.00%)	0.57%	52
M05189	6,830	NA	NA	NA	UD	UD	UD	UD
T04228	16,534	NA	NA	NA	UD	UD	UD	UD
Z08103	35,530	Gran + B	4	308, 602, 644, 747	740	40 (5.41%)	0.37%	54
Z09132	43,558	Gran + B	2	579, 616	175	16 (9.14%)	0.39%	51

Abbreviations: NA (none available); UD (undetermined due to insufficient data); B ($CD20^+$ B cells); Mono ($CD14^+$ monocytes); Gran (granulocytes); T ($CD3^+$/$CD4^+$/$CD8^+$ T cells); PB (peripheral blood).

Table S2 *(Related to Figure 2.)* Cross-species reactive antibodies in nonhuman primates.

Epitope	Clone	Company	Pigtailed macaque	Rhesus macaque
CD2	RPA-2.10	Biolegend	Yes	-
CD3	SP34-2	BD Biosciences	Yes	-
CD4	L200	BD Bioscience	Yes	-
CD5	UCHT2	Biolegend	Yes	-
CD8a	RPA-T8	BD Bioscience	Yes	-
	RPA-T8	eBioscience	Yes	-
CD11b	M1/70	Abcam	Yes	-
	ICRF44	BioLegend	Yes	-
	ICRF44	Fisher Scientific	Yes	-
CD14	M5E2	BD Biosciences	Yes	-
	63D3	eBioscience	Yes	-
CD15	HI98	BD Biosciences	NO	-
	W6D3	BioLegend	NO	-
	MC-480	BioLegend	NO	-
	VIMC6	Miltenyi Biotec	NO	-
CD16	3G8	BD Biosciences	Yes	-
	3G8	BioLegend	Yes	-
CD19	4G7	BD Biosciences	Yes	-
CD20	2H7	BD Biosciences	Yes	-
	2H7	BioLegend	Yes	-
CD34	563	BD Biosciences	Yes	Yes
CD38	AT-1	Stemcell Techn.	Yes	Yes
	OKT10	Nonhuman primate-Reagents	NO	-
CD45	D058-1283	BD Biosciences	Yes	Yes
CD45RA	5H9	BD Biosciences	Yes	Yes
	T6D11	Miltenyi Biotec	Yes	-
CD49f	GoH3	BD Biosciences	NO	-
CD56	B159	BD Biosciences	Yes	-
CD90	5E10	BD Biosciences	Yes	Yes
	DG3	Miltenyi Biotec	No	-
	F15-42-1-5	Beckman Coulter	No	-
CD117	104D2	BD Biosciences	Yes	Yes
	YB5.B8	BD Biosciences	NO	-
CD123	9F5	BD Biosciences	Yes	Yes
	7G3	BD Biosciences	Yes	-
CD133	AC133	Miltenyi Biotec	NO	NO
	HC7	Hybridoma Bank	NO	NO
CD135	4G8	BD Biosciences	NO	-
	BV10A4H2	BioLegend	NO	-
CD235a	HIR2	BioLegend	NO	-
	HI264	BioLegend	NO	-
	YTH89.1	Thermo Scientific	NO	-
	REA175	Miltenyi Biotec	NO	-
	CLB-ery-1	Life Technologies	NO	-
CD235ab	HIR2	BioLegend	NO	-

Table S3. *(Related to Figure S8.)* MS-5 assay colonies.

Population:	I	II	III	IV	V	VI	VII	VIII	IX
Gran	17	18	-	26 (5)	19	25	29 (5)	56	13
Mono	2	6	-	12	10	0	6	10	1
Gran/Mono	16	12	-	26 (2)	12	11	4 (2)	27	0
NK	0	0	-	0	1	0	0	5	0
NK/Gran	3 (2)	0	-	2 (2)	0	0	20 (19)	2	0
NK/Mono	0	0	-	2 (2)	0	0	0	0	0
NK/Gran/Mono	5 (2)	0	-	16 (9)	0	0	4 (4)	4	0
CD34$^+$	10	0	-	3	0	0	26	0	0
Positive Total	53	36	-	87	42	36	88	104	14
Negative Total	187	204	-	153	198	204	152	130	226
∑ Tested (n=4)	240	240	-	240	240	240	240	234	240

Numbers indicate total count of positive wells based on optical counting of colonies combined with flow-cytometric analysis of hematopoietic cell types. Numbers in parentheses indicate colonies containing cobblestone area-forming cells expressing CD34 by flow cytometry.

Table S4. *(Related to Figure 3.)* Characteristics of nonhuman primate transplants.

	Z13314	Z14004	Z13264	Z15086
Stem cell source	pBM	pBM	pBM	pBM
Experimental design	fraction *i* ($CD34^+CD45RA^-CD90^+$) vs. fraction *ii* ($CD34^+CD45RA^-CD90^-$) vs. fraction *iii* ($CD34^+CD45RA^+CD90^-$)			
Condition 1	fraction *i* GFP	fraction *i* GFP	fraction *i* mCh	fraction *i* mCer
Condition 2	fraction *ii* mCh	fraction *ii* mCh	fraction *ii* mCer	fraction *ii* GFP
Condition 3	fraction *iii* mCer	fraction *iii* mCer	fraction *iii* GFP	fraction *iii* mCh
Days of culture	2	2	2	2
Weight [kg]	3.56	4.29	3.83	2.80
GCSF admin. [d]	0-11	0-11	0-11	0-11
Neu. engft. [d]	10	10	9	9
Plt. engrf. [d]	20	19	10	13
Adverse events	CMV, EA	CMV	Giardia	CMV
TNC cond. 1	1,520,000	1,900,000	2,400,000	1,360,000
TNC cond. 2	1,880,000	8,400,000	1,200,000	381,000
TNC cond. 3.	12,400,000	13,000,000	7,000,000	2,140,000
TNC SUM	15,800,000	23,300,000	10,600,000	3,881,000
Total $CD34^+$	6,302,520	8,483,700	4,415,000	2,198,659
$CD34^+$/kg	1,770,371	1,977,552	1,152,742	785,235
Total fraction *i* cells	1,339,300	1,723,800	2,165,940	1,450,619
fraction *i* cells/kg	376,208	401,818	565,520	518,078
Total fraction *ii* cells	1,666,824	3,505,700	490,600	180,554
fraction *ii* cells/kg	468,209	817,179	128,094	64,484
Total fraction *iii* cells	2,963,272	2,210,600	1,649,820	567,486
fraction *iii* cells/kg	823,380	515,291	430,762	202,674
Total CFCs	491,174	941,483	558,574	492,031
CFCs/kg	137,970	219,459	145,841	175,725

Abbreviations: pBM: G-CSF/SCF primed bone marrow (BM); GFP: green fluorescent protein; mCh: mCherry; mCer: mCerulean; d: days; admin.: administration; Neu.: neutrophils; Plt.: platelets; engft.: engraftment; CMV: cytomegalovirus; EA: early anesthesia; fraction i: $CD34^+CD45RA^-CD90^+$; fraction ii: $CD34^+CD45RA^-CD90^-$; fraction iii: $CD34^+CD45RA^+CD90^-$; CFC: colony-forming cells; HSPC: hematopoietic stem and progenitor cells.

Table S5. *(Related to Figure 5.)* Characteristics of nonhuman primate transplants II.

	Z13251	Z13125	Z13260	Z12434	R11145
Stem cell source	pBM	pBM	ssBM	ssBM	ssBM
Experimental design	$CD45RA^+$ vs. $CD45RA^-$	VSVG vs. Cocal envelope	VSVG vs. Cocal envelope	VSVG vs. Cocal envelope	EC co-culture vs. SP-STF
Condition 1	$CD45RA^+$ GFP	VSVG GFP	VSVG GFP	Cocal GFP	EC co-culture GFP
Condition 2	$CD45RA^-$ mCh	Cocal mCh	Cocal mCh	VSVG mCh	SP-STF mCh
Condition 3	-	-	-	-	-
Days of culture	2	2	9	9	9
Weight [kg]	3.26	4.33	4.28	2.98	4.10
GCSF admin. [d]	0-21	0-13	0-26	0-16	0-14
Neu. engft. [d]	18	11	24	17	15
Plt. engrf. [d]	-	30	47	40	30
Adverse events	EF, EA	EF, EA	-	-	-
TNC cond. 1	42,000,000	15,600,000	6,500,000	21,700,000	123,000,000
TNC cond. 2	55,000,000	12,900,000	5,100,000	34,000,000	12,000,000
TNC cond. 3.	-	-	-	-	-
TNC SUM	97,000,000	28,000,000	11,600,000	55,700,000	135,000,000
Total $CD34^+$	4,061,000	8,567,400	4,557,000	11,549,400	30,381,000
$CD34^+$/kg	1,245,706	1,978,614	1,064,720	3,875,638	7,410,000
Total fraction *i* cells	106,000	444,780	523,400	501,542	1,116,840
fraction *i* cells/kg	32,515	102,721	122,290	168,303	272,400
Total fraction *ii* cells	501,000	565,032	587,900	1,370,809	-
fraction *ii* cells/kg	153,681	130,492	137,360	460,003	-
Total fraction *iii* cells	3,159,100	6,285,816	3,210,900	9,284,380	-
fraction *iii* cells/kg	969,049	1,451,690	750,210	3,115,564	-
Total CFCs	236,690	1,287,678	495,228	-	2,220,642
CFCs/kg	72,604	297,385	115,707	-	541,620

Abbreviations: pBM: G-CSF/SCF primed bone marrow (BM); ssBM: steady state bone marrow; EC: endothelial cells; SP-STF: StemSpan serum-free expansion media (Stemcell Technologies) supplemented with SCF (stem cell factor), TPO (Thrombopoietin) and FLT3-L (Fms-like transcription factor 3 ligand); GFP: green fluorescent protein; mCh: mCherry; d: days; admin.: administration; Neu.: neutrophils; Plt.: platelets; engft.: engraftment; EF: engraftment failure; EA: early anesthesia; fraction i: $CD34^+CD45RA^-CD90^+$; fraction ii: $CD34^+CD45RA^-CD90^-$; fraction iii: $CD34^+CD45RA^+CD90^-$; CFC: colony-forming cells; HSPC: hematopoietic stem and progenitor cells.

Table S5. continued *(Related to Figure 5.)* Characteristics of nonhuman primate transplants II.

	Z13137	Z14148	Z14035	Z14279	Z14123	Z14160
Stem cell source	pBM	pBM	pBM	pBM	pBM	pBM
Experimental design	Lentivirus - Transduction	Lentivirus - Transduction	Lentivirus - Transduction	Lentivirus - Transduction	Lentivirus - Transduction	Lentivirus - Transduction
Days of culture	2	2	2	2	2	2
Weight [kg]	4.94	3.77	3.6	3.74	4.3	4.5
GCSF admin. [d]	0-17	0-13	0-7	0-9	0-11	0-14
Neu. engft. [d]	15	12	7	9	10	13
Plt. engrf. [d]	30	24	15	20	21	-
Adverse events	CMV	-	Amobea, CMV	-	-	Kidney failure, EA
TNC SUM	80,000,000	66,000,000	68,000,000	55,000,000	97,000,000	55,000,000
Total CD34$^+$	22,960,000	19,338,000	23,188,000	19,205,000	28,518,000	17,215,000
CD34$^+$/kg	4,647,773	5,129,443	6,441,111	5,161,765	6,632,093	3,825,556
Total fraction *i* cells	1,249,024	1,345,924	1,790,113	1,671,813	1,579,897	986,42
fraction *i* cells/kg	252,839	357,009	497,254	447,009	367,418	219,204
Total fraction *ii* cells	1,294,944	3,732,234	1,996,486	2,567,565	2,937,354	2,289,595
fraction *ii* cells/kg	262,134	989,982	554,579	686,515	683,106	508,799
Total fraction *iii* cells	16,531,200	12,395,658	17,321,436	13,301,145	21,873,306	11,534,050
fraction *iii* cells/kg	3,346,397	3,287,973	4,811,510	3,556,456	5,086,815	2,563,122
Total CFCs	7,956,600	6,285,714	7,095,652	4,310,806	2,684,525	1,122,418
CFCs/kg	1,610,648	1,667,298	1,971,014	1,152,622	856,866	249,426

Abbreviations: pBM: GCSF/SCF primed bone marrow (BM); d: days; admin.: administration; Neu.: neutrophils; Plt.: platelets; engft.: engraftment; CMV: cytomegalovirus; EA: early anesthesia fraction i: CD34$^+$CD45RA$^-$CD90$^+$; fraction ii: CD34$^+$CD45RA$^-$CD90$^-$; fraction iii: CD34$^+$CD45RA$^+$CD90$^-$; CFC: colony-forming cells; HSPC: hematopoietic stem and progenitor cells.

Table S6. *(Related to Figure 6.)* Differentially expressed genes in human and nonhuman primate HSCs.

Upregulated in nonhuman primate AND human	Only nonhuman primate	Downregulated in nonhuman primate AND Human	Only human
ARHGEF12, ITSN2, MEIS1, SVIL, STARD9, ZNF154, FAM30A, ZBTB20, MYEF2, SCAI, LRBA, CAPN2, PDZD2, PRKCH, ZBTB21, HEG1, C7orf49, ABLIM1, ENSG00000189089, NFAT5, CCPG1, C1orf21, ZFHX3, MAML2, ARMCX4, MYCT1, ZNF286A, LPP, ENSG00000281195, CRIM1, DSG2, TMEM136, SCN9A, ABCA5, HOXB3, SNTB1, CLU, GPRASP1, CHN2, ENSG00000260244, FZD6, NR1D2, TCEANC2, PARP11, DST, FHL1, ZKSCAN1, RBPMS, NRIP1, ZMYM1, MLLT3, PLAG1, LOC100505501, TFPI, HMGA2, SOCS2, BTB37, SETBP1, GUCY1A3, RAB29, MAPK11, ZNF462, TMEM163, PLK2, EVA1C, BMPR2, DNAH6, ALDH1A1, INPP4B, TSPAN2, TOX, HOPX, ARHGEF40, C11orf63, IDS, SAMD12, STYK1, FAT4, MIPOL1, SLC22A17, MAGI2-AS3, DLK1, PRDM16, GATAD1, MMRN1, AR, AKAP6, ARMCX2, MECOM, COL6A2, ENSG00000215208, PARD3B, PDGFD, ZNF532, ADGRG6, WASF3, GATA3, ENSG00000263345, HLA-DQA1, ENSG00000259591, ZNF165, ANK3, PCDHGB7, SKIDA1, GPRASP2, ENSG00000242795, LIMCH1, ZNF483, HLF, TCEAL2, AFDN, LOC101927577, ST8SIA1, PREX2, MEG3	MME, CR2, SEMA3C, PTGER3, CD84, PIK3AP1, CYP2U1, ABCA1	MS4A2, CPA3, ENSG00000256663, F13A1, MYBL2, TREM1, CYBB, AFF2, CLEC12A, RAB44, KCNK5, HES6, AZU1, DNTT, IRF8, MPO, TIMELESS, ENSG00000229989, ESPL1, MS4A4E, ADAMTS14, ENSG00000253525, FAM107B, ZWINT, E2F1, NOTCH3, CDC45, PLD1, CDCA2, HCK, SLC40A1, PLIN2, MCM2, MCM10, UHRF1, CTNNBL1, ZEB2-AS1, RRM2, CP, NKG7, TSPOAP1, HGF, PLEK, IL7R, CEBPA, LMNB1, CD180, COTL1, PCNA, ELANE, BIRC5, GTSE1, WDHD1, BRI3BP, MYC, MITF, ANKLE1, CHAF1A, CENPU, MCM4, ENSG00000263846, DPP3, AAAS, RAB31, OLQ, TIMM8B, CXCR4, ENSG00000236480, SLC29A1, BAHCC1, DTL, MCM3, ABCC4, MTA2, NT5DC3, ISOC1, UNG, SRM, TYMS, CCNE1, ENSG00000231799, TLR2, PPM1G, MLEC, C1QBP, GINS2, HELLS, RAD54L, CTPS1, HNRNPAB, SHMT2, CHEK1, DNMT1, TMSB4X, DUSP10, ZEB2, HN1, FNDC3B, SND1, SNRPF, HSPA8, NOP56, MAP4K4, SRPRB, RRM1, PPP3R1, ATAD5, SEC61A1, SNHG17, PRKDC, ATP5B	LOC153684, DRAM1, LDLR, ID2, PNP, DGKH, ENSG00000271204, POLE4, KCNC3, ENSG00000236484, ENSG00000225745, ENSG00000228509, LAMTOR2

Genes sorted by p-values, highest to lowest ($p<0.05$)