Materials and Methods.

Human CD34$^+$ cells

Human CD34$^+$ mobilized peripheral blood (mPB) cells were donated by Hôpital Maisonneuve-Rosemont in Montreal (Canada).mPB was collected from consenting donors according to an ethically approved protocol at Hôpital Maisonneuve-Rosemont in Montreal (Canada). Pre-selected human CD34$^+$ mPB cells were purchased from StemExpress. Human CD34$^+$ bone marrow cells were purchased from Lonza and fresh whole samples from iSpecimen that were CD34$^+$ purified by ExcellThera (Canada) using a cliniMACS Separator by Miltenyi Biotec. No information concerning gender identity of cell donors (for BM-, mPB- and UCB-derived CD34$^+$ cells) is available for ethical reasons.

Analysis of bulk transcriptomes

Around 3-5 x 10^5 UCB cells sorted from day 7 expanded CD34$^+$ HSPCs based on EPCR or ITGA3 expression were used for bulk sequencing. Both EPCR[12] and ITGA3[16] RNA-sequencing dataset have been previously reported. The analysis for these datasets were done as previously described[16,19] using the Kallisto/Slueth pipeline based on the GRCh38 v92 annotation. From this a list of differentially expressed genes (DEGs) from the ITGA3 and EPCR dataset were analysed. The intersection of these two datasets was also published recently to give the upregulated and downregulated DEGs[19].

Flow cytometry and sorting

Cells were stained using the following antibodies: anti-CD34 (BV421, BD Biosciences, Catalog No. 562577), anti-CD45RA (FITC, TONBO Biosciences, Catalog No. 35-0458-T100), anti-CD90 (PE-Cy7, BioLegend, Catalog No. 328124), anti-EPCR (APC-Cy7, Miltenyi, Catalog No. 130-105-259), anti-ITGA3 (APC, eBioscience, Catalog No. 17-0494-42), anti-CD133 (PE, Miltenyi, Catalog No.130-080-801), anti-GPR56 (PE, BioLegend, Catalog No.358206), and CD318 (PEE, BioLegend, Catalog No.324006), anti-CD38(PerCP-eFlour 710, eBioscience, Catalog No.46038841). Cultured cells were analyzed using FACSCanto II (BD Biosciences) and sorted on a FACSAria II cell sorter (BD Biosciences). Flow cytometry data were analyzed using FlowJo software.

Transplantation Assay

All experiments with animals were conducted using protocols approved by University of Montreal Animal Care Committee. UCB cells were transplanted by tail vail injection into sublethally irradiated (250 cGy) 8- to 16-week-old female NOD-Scid IL2-Rγ null (NSG, the Jackson Laboratory) mice. The progeny of 5000 CD34$^+$ cells from fresh UCB samples were transplanted per mouse. The engraftment of human UCB cells in the bone marrow of NSG mice were measured by flow cytometry at 3, 12- and 24-week post transplantation. BM cells were collected by femoral aspirations (week 3 and 12) or by flushing 2 femurs, tibias, and hips when the animal was sacrificed at week 24. The criteria used for successful engraftment and evaluation of HSC activity were the same as previously reported by our group. We used an engraftment criterion of 0.1% human CD45 cells in the BM as assessed by flow cytometry to establish a biologically significant cut-off. For secondary transplantations, 80% of total BM cell from primary NSG recipients whose human reconstitution was >5% (24-week post-transplant) were injected into secondary sublethally irradiated NSG mice for 20 weeks. To perform the limiting dilution assay (LDA), cells were transplanted at 3 different dilutions (outcome of 5000, 1000, 100 CD34+ cells from fresh UCB sample per mouse). The results of the LDA experiments were analyzed using the ELDA software from the Walter and Eliza Hall Institute of Medical Research. Flow cytometry was

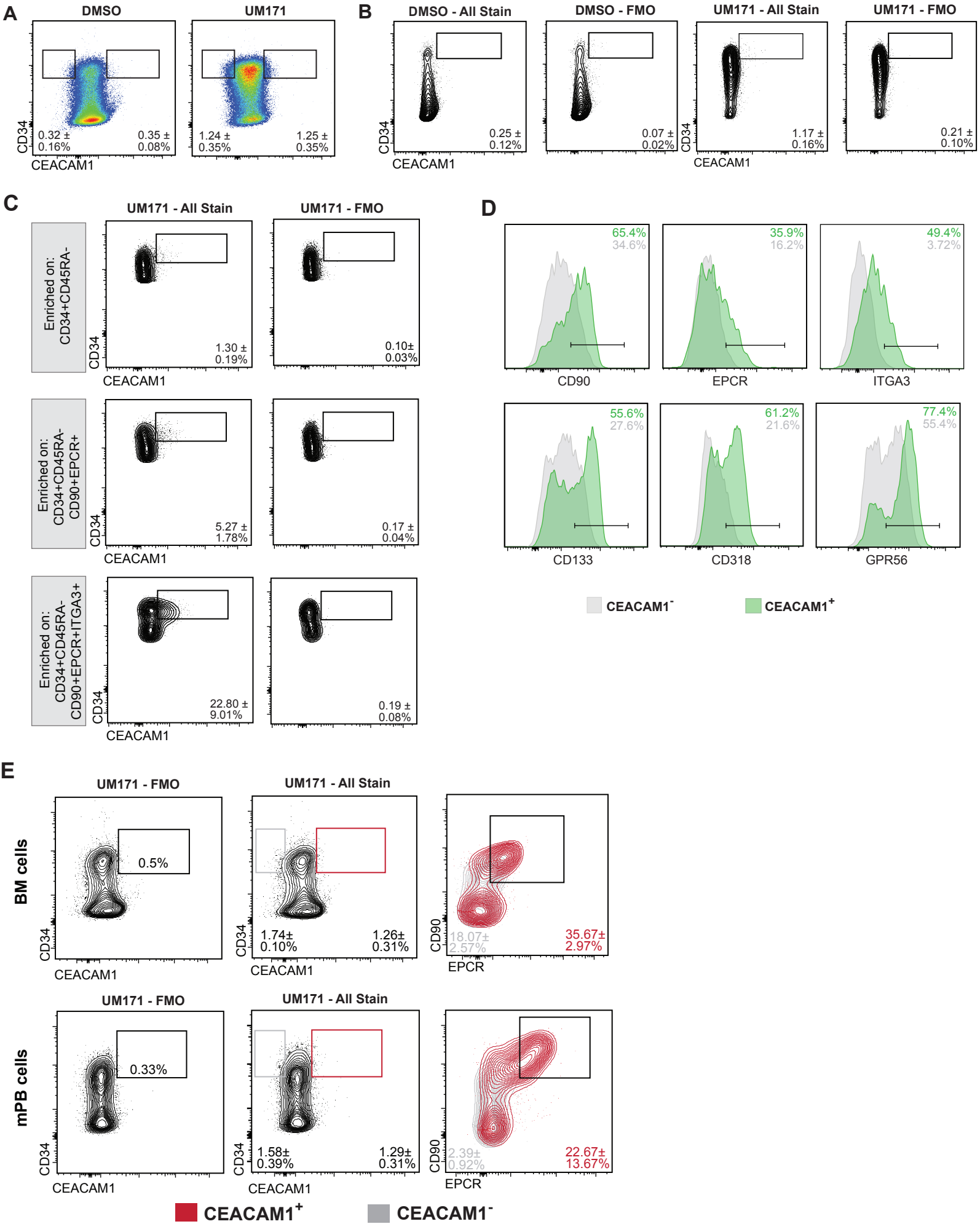
performed on freshly collected BM cells. Cells were treated with red blood cell lysis buffer (STEMCELL Technologies, Catalog No. 20110) washed and stained with anti-mouse CD45.1 (APC-eFluor780, eBiosciences, Catalog No. 47-0453-82), anti-CD45 (Pacific Blue, BioLegend, Catalog No. 304029), anti-CD33 (PE, BD Biosciences, Catalog No. 555450), anti-CD19 (PE-Cy7, BD Biosciences, Catalog No. 557835), anti-CD3 (FITC, BD Biosciences, Catalog No. 555332), anti-CD34 (APC, BD Biosciences, Catalog No. 555824).

Clonal Assay on MS-5 Stroma

MS-5 was seeded in 96-well plates (Nunc) coated with 0.1% (wt/vol) gelatin at a density of 5 x 10^3 cells per well in α-MEM medium (Gibco, Catalog No. 2003802) supplemented with 12.5% heat inactivated fetal bovine serum (Winsent, Catalog No. 098-150), 12.5% horse serum (Gibco, Catalog No. 16050-122), 100ng/ml SCF (Shenandoah Biotechnology, Catalog No. 100-04), 20ng/ml IL-7 (Shenandoah Biotechnology, Catalog No. 100-11), 50ng/ml TPO (R&D Systems, Catalog No. 288-TP) and 10ng/ml IL-2 (Shenandoah Biotechnology, Catalog No. 100-12) and B-mercaptoethanol. Cells were exposed to 15 cGy γ-irradiation to mitotically inactivate the cells 2 days after plating. After 48h, single sorted progenitor cells were sorted onto stromal monolayers. MS-5 cultures were maintained for 4 weeks with weekly changes of half the media. Cells were stained with anti-CD45 (FITC, BioLegend, Catalog No. 304006), anti-CD33 (PE, BD Biosciences, Catalog No. 555450), anti-CD19 (PE-Cy7, BD Biosciences, Catalog No. 557835), anti-CD56 (APC, BD Biosciences, Catalog No. 555518), anti-CD14 (BV, BioLegend, Catalog No. 301830) and then analyzed by high-through flow cytometry using FACSCelesta (BD Biosciences). Human-$CD45^+$ lymphoid colonies were identified as $CD19^+$ (B cells) and $CD56^+$ (NK cells). Myeloid colonies were identified as $CD33^+$ and/or $CD14^+$ in the $CD19^-CD56^-$ cells. ELDA software from Walter and Eliza Hall Institute of Medical Research was used to calculate the frequencies.

Statistical Analysis.

Statistical details of each experiment are denoted in the legend of each figure, where required. Statistical analyses were performed using GraphPad Prism 8 software. The Mann-Whitney t test and unpaired t test was used to compare the fold change between the conditions tested. Results are usually present as mean±SD unless specified. Bars and error bars represent means or medians and standard deviations as specified. Extreme limiting dilution analysis software (http://bioinf.wehi.edu.au/software/elda/) was used to estimate the frequencies with 95% confidence intervals. Differences were considered significant at *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001



Supplementary Figure 1

Fig S1. CEACAM1 expressing cells help identify naïve HSCs in culture (Related to Fig 1)

(A) Flow cytometry analysis showing the proportion of $CD34^+CEACAM1^+$ and $CD34^+CEACAM1^-$ cells in UCB samples expanded for 7 days with UM171 (35nM) or DMSO. Data presented as mean ± SD for 4 independent biological replicates.

(B) CD34+ UCB cells expanded with DMSO or UM171 for 7 days were stained with CD34, CD45RA, CD90, EPCR, ITGA3 and CEACAM1 (All Stain) antibodies as well as Fluorescence Minus One (FMO) for CEACAM1. Plots illustrated for d7 DMSO and UM171 cells either having all the antibody staining or the FMO for CEACAM1. Percentages represent the relative size of each population in depicted gates. Data presented as mean ± SD of 3 independent biological replicates.

(C) FMO controls for CEACAM1 antibody compared to complete staining mixture illustrate the expression of CEACAM1 enriched in $CD34^+CD45RA^-$, $CD34^+CD45RA^-CD90^+EPCR^+$, and $CD34^+CD45RA^-CD90^+EPCR^+ITGA3^+$ population in UM171-expanded UCB cells.

(D) FACS analysis of $CD34^+CEACAM1^+$ and $CD34^+CEACAM1^-$ cells expanded with UM171 for 7days showing their distribution with respect to other markers like CD90, EPCR, ITGA3, CD133, CD328 and GPR56. Representative of 3 independent experiments

(E) Expression of $CD34^+CEACAM1^+$ and $CD34^+CEACAM1^-$ cells in UM171-expanded mobilized peripheral blood (mPB) cells and bone marrow (BM) cells. $CD34^+CEACAM1^+$ cells enrich at a better proportion in $CD34^+CD90^+EPCR^{+\ +}$ fraction compared to $CD34^+CEACAM1^-$ cells. Data presented as mean ± SD for 3 independent biological replicates.

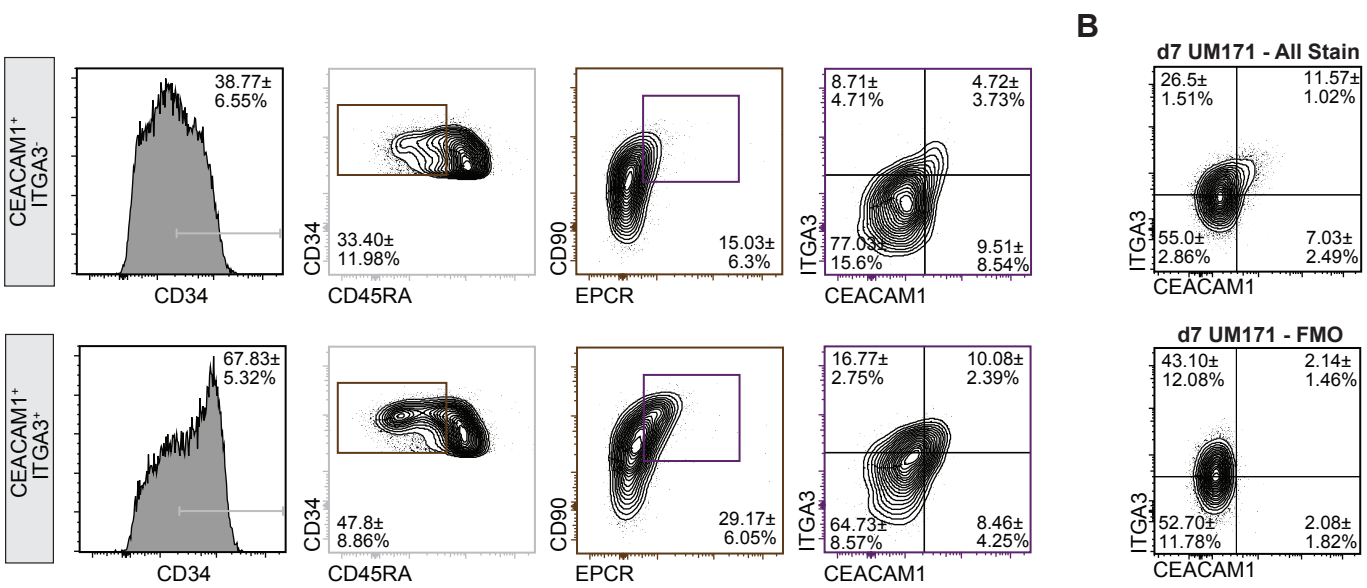
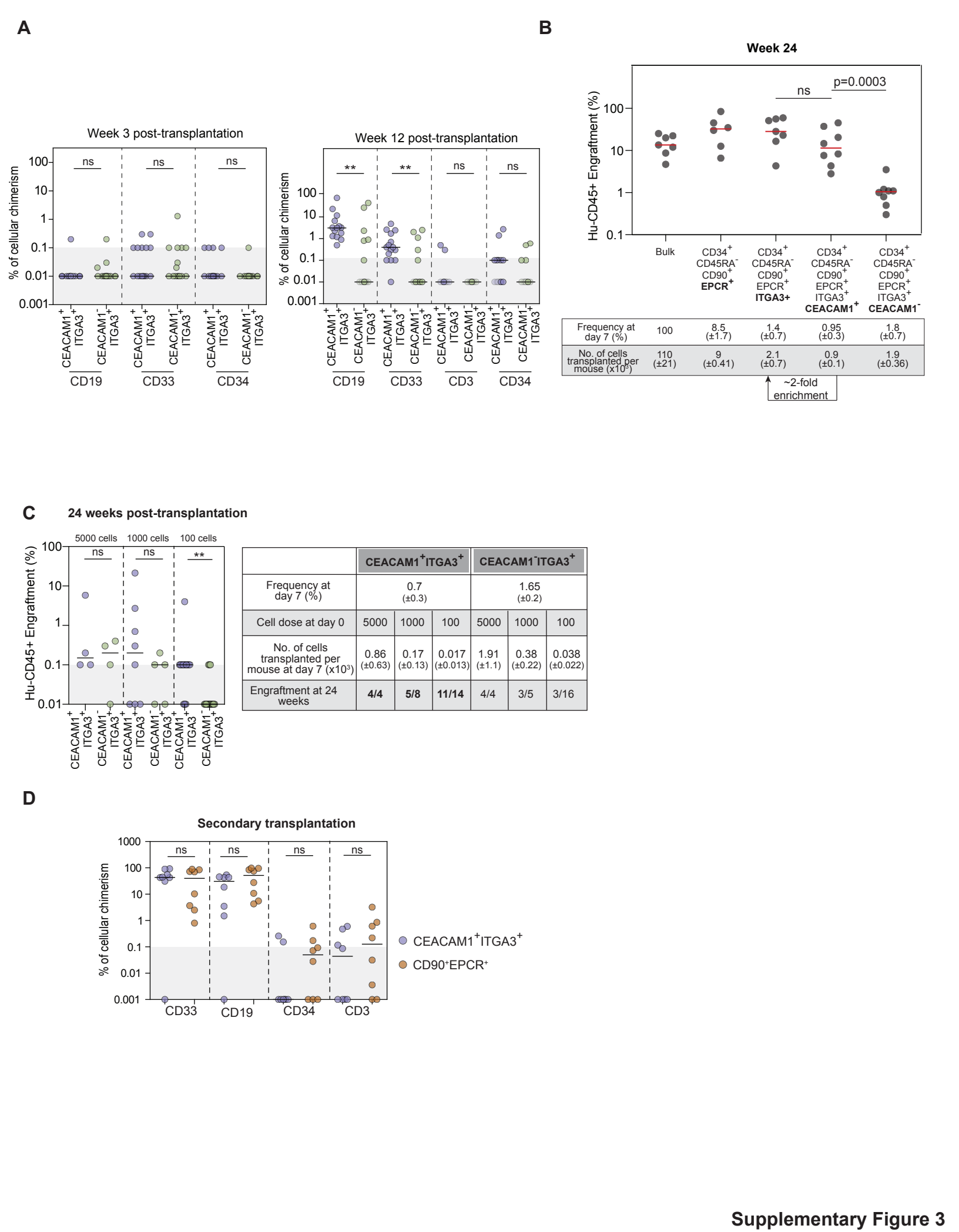


Fig S2. CEACAM1 expression identified a primitive population with clonogenicity via in vitro co-culture system. (Related to Fig 1)

(A) Contour plot analysis of CD34, CD90, EPCR, ITGA3 and CEACAM1 expression in UCB cells. The cells were expanded with UM171 for 7 days and sorted based on the expression of CEACAM1 and ITGA3 from CD34+CD45RA-CD90 +EPCR+ cells and put back into culture for an additional 7 days. Percentages indicate the size of each population of different groups. Data represented as mean SD for 3 independent biological replicates.

(B) Fluorescence minus one (FMO) staining of CEACAM1 done for day 7 UM171-expanded UCB cells to illustrate the gating scheme for CD34+CD45RA-CD90+EPCR+ITGA3+CEACAM1+population. Data represented as mean SD for 3 independent biological replicates.

Supplementary Figure 2



Supplementary Figure 3

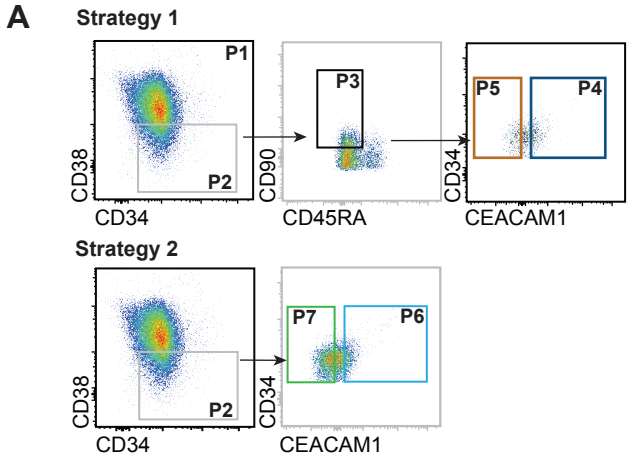
Fig S3. Expression of CEACAM1 can isolate functional LT-HSCs *in vivo* (Related Fig 2)

(A) Lineage potential in NSG mice for week 3 and week 12 post-transplant for $CEACAM1^{+}ITGA3^{+}$ and $CEACAM1^{-}ITGA3^{+}$ cells sorted from day 7 UM171-expanded UCB cells. **p=0.0032, **p = 0.0063. (Mann-Whitney U test, two-tailed)

(B) Engraftment of human CD45 potential of $CEACAM1^{+}ITGA3^{+}$ and $CEACAM1^{-}ITGA3^{+}$ cells when compared to $ITGA3^{+}EPCR+$ cells to detect the fold enrichment. Based on the frequency and number of cells a fold enrichment for $CD34^{+}CD45RA^{-}CD90^{+}EPCR^{+}ITGA3^{+}CEACAM1+$ cells can be calculated for an ~2-fold enrichment over $CD34^{+}CD45RA^{-}CD90^{+}EPCR^{+}ITGA3^{+}$ cells.

(C) Bone marrow engraftment of sorted $CEACAM1^{+}ITGA3^{+}$ and $CEACAM1^{-}ITGA3^{+}$ cells at week 24 post-transplantation in a limited diluted assay (LDA). **p = 0.0019. (Mann-Whitney U test, two-tailed). Depiction of the LDA experiment of $CEACAM1^{+}ITGA3^{+}$ and $CEACAM1^{-}ITGA3^{+}$ sorted from expanded UCB cells. Cells were transplanted as an outcome of 5000, 1000 and 100 cells based on occurring frequency at day 7. Engraftment at week 24 is shown as ratio a/b where a = engrafted mouse and b = transplanted mouse.

(D) Lineage potential of secondary mice at 18 weeks post transplantations in NSG mice. The distinct pool of cells originating from primary mice of which about 80% of collected BM grafts were transplanted. n = 2 biological replicate.



B

	% cells at d0	Number of cells transplanted/mouse
Bulk (P1)	100	5000
$CD34^+CD38^-CD45RA^-$ $CD90^+CEACAM1^+$ **(P4)**	1.5 (±0.7)	75 (±35)
$CD34^+CD38^-CD45RA^-$ $CD90^+CEACAM1^-$ **(P5)**	1.2 (±0.2)	60 (±14)
$CD34^+CD38^-$ $CEACAM1^+$ **(P6)**	5.3 (±1.8)	265 (±91)
$CD34^+CD38^-$ $CEACAM1^-$ **(P7)**	5 (±1.4)	250 (±70)

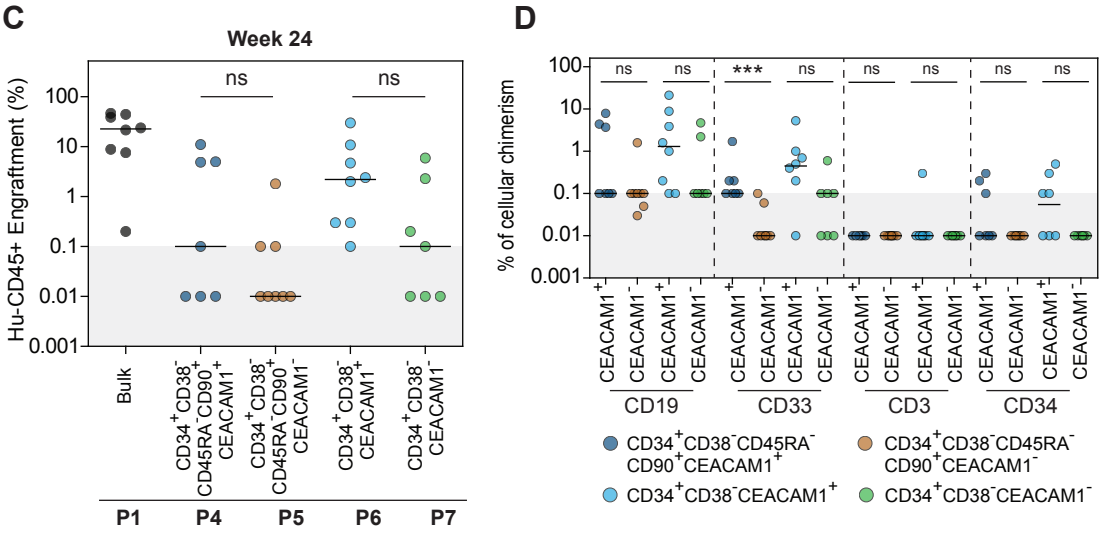


Fig S4. CEACAM1 expression in unexpanded UCB cells (Related to Fig 2)

(A) Sorting strategies used for analysis of CEACAM1+/- cells from fresh day 0 UCB cells. Strategy 1 employs enrichment of CEACAM1+ (P4) and CEACAM1- (P5) cells from CD34+CD38-CD90+CD45RA-. Strategy 2 enriches CEACAM1+ (P6) and CEACAM1- (P7) cells from CD34+CD38-.

(B) Percentage of cells in each population and the number of cells transplanted for each population for P4, P5, P6 and P7 from fresh UCB cells.

(C) Reconstitution of human CD45 cells in NSG mice of fresh unexpanded cells sorted in populations – P4, P5, P6 and P7 to assess engraftment potential.

(D) Lineage potential of fresh CEACAM1 expressing cells for each engrafted mouse. *** p = 0.009 (Mann-Whitney U test, two-tailed)