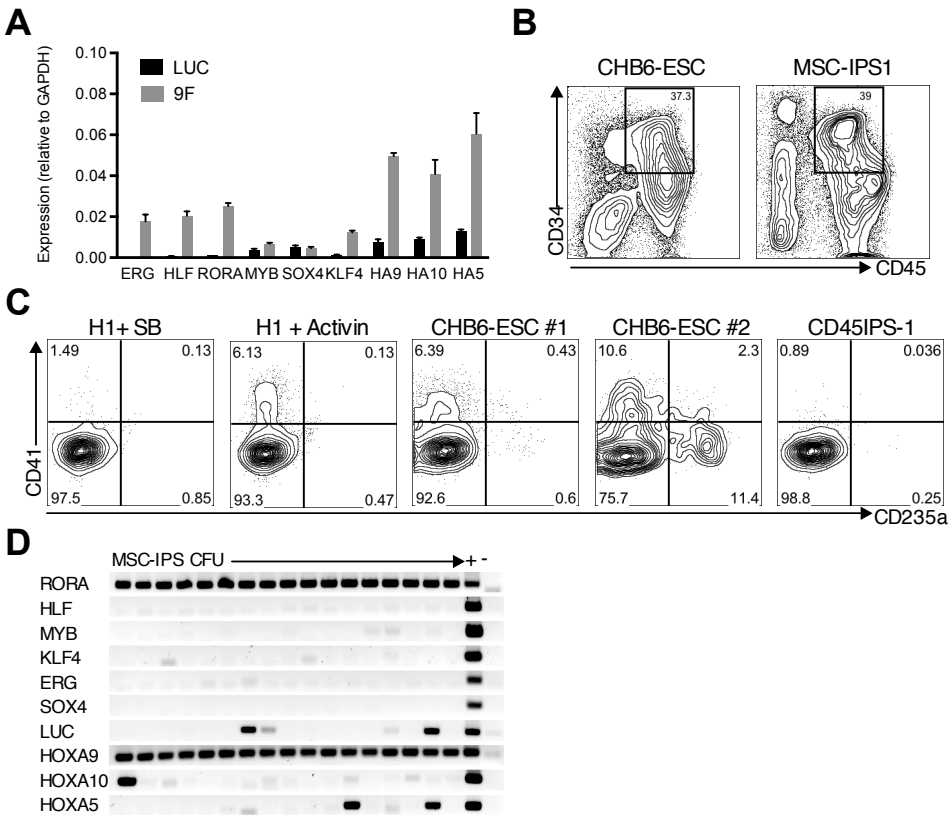


Figure S1, related to **Figure 1**. (A) Quantitative PCR measurement of expression of the 9 HSC transcription factors in 293T cells transduced with the 9F library or LUC control. (B) Highly efficient differentiation of CHB6 ESCs and MSC IPSCs into day 14 $CD34^{+}CD45^{+}$ EB progenitors. (C) Expression of primitive hematopoietic markers CD41a and CD235a on $CD45^{+}$ day 14 EB cells across multiple hPSC lines and replicates. As a control, H1 ESCs were cultured in serum-free condition with Activin A (permits primitive hematopoiesis) or SB (inhibits primitive hematopoiesis). (D) Transgene insertions in individual colonies from MSC-IPS1 EB progenitors transduced with 9F library and cultured for 7 days. Data in (A) are shown as mean ± SEM of 3 replicates.

Figure S2

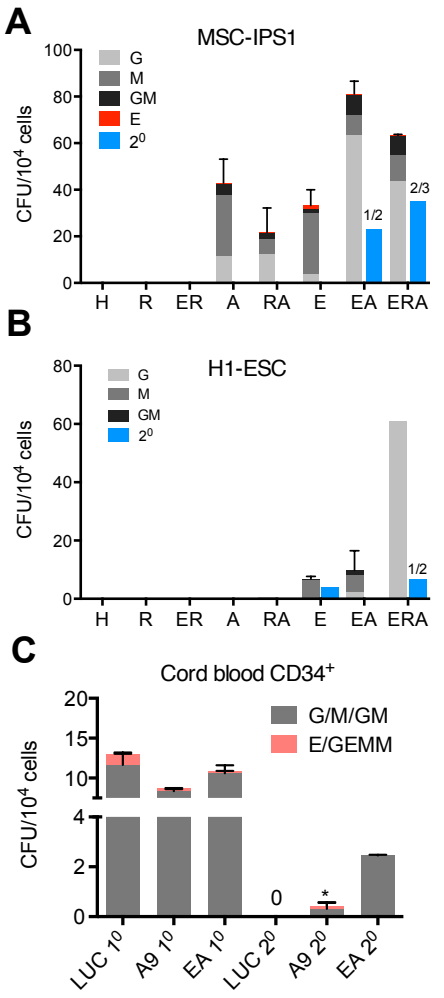


Figure S2, related to **Figure 2**. Prospective analysis of HOXA9 (A), ERG (E), HLF (H), and RORA (R) combinations in MSC-IPS (A) and H1 (B) progenitors using the in vitro colony assay. Primary and secondary (blue bars) colony-forming efficiency is shown, and numbers above the bars indicate the proportion of replicates that gave rise to secondary colonies. (C) Serial re-plating of CB progenitors. CB CD34$^+$ progenitors were transduced with ERG-GFP and HOXA9-TagBFP lentiviruses, or LUC control. GFP$^+$BFP$^+$ (EA) or BFP$^+$GFP$^-$ (A9) cells were sorted and cultured for 14 days before plating primary colonies (1^0) and secondary (2^0) colonies. *A9-alone 2^0 colonies were GFP$^+$ indicating they were derived from EA double-transduced cells. Data in (A) and (C) are shown as mean ± SEM of 3 independent replicates. Data in (B) are shown as mean ± SEM of 2 independent replicates.

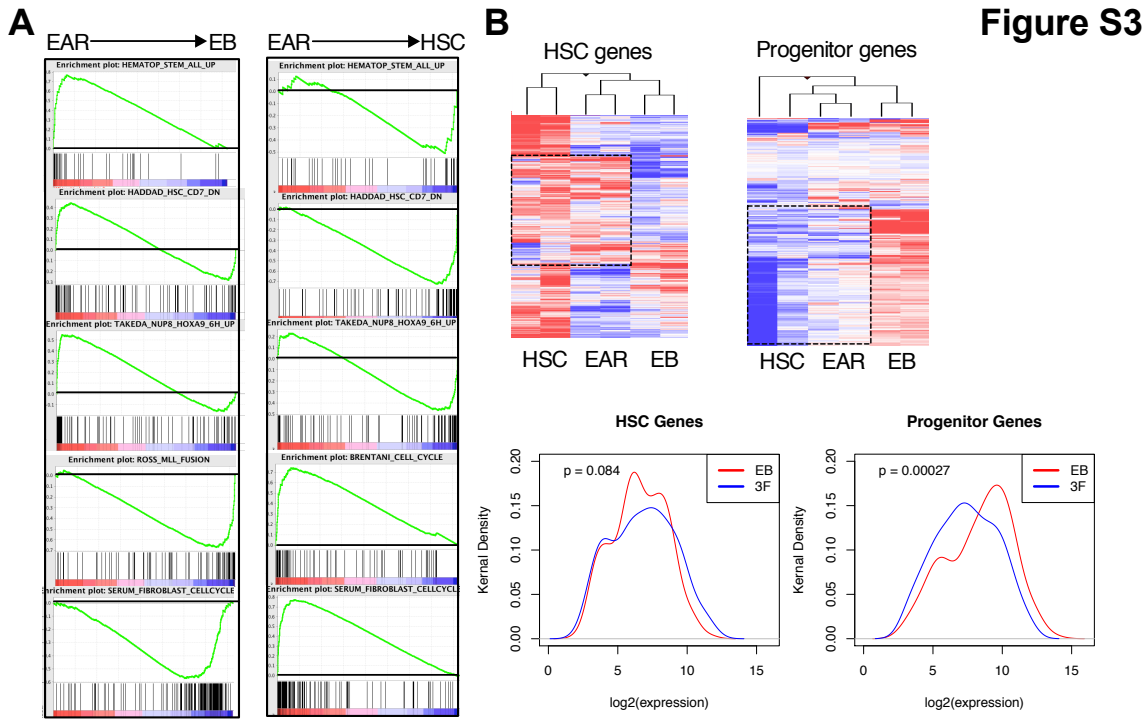


Figure S3

Figure S3, related to **Figure 4**. (A) Gene set enrichment analysis of annotated gene sets in MSigDB comparing *EAR* versus EB progenitors (left column), and *EAR* versus primary HSCs (right column). Significant gene sets that appear in both comparisons are shown (FDR $<10^{-4}$). (B) (Top) Supervised clustering of the differentially expressed genes between $CD34^{+}CD38^{-}$ CB/FL HSCs and $CD34^{+}CD38^{+}$ progenitors. Dotted boxes show groups of HSC genes upregulated by *EAR*, or progenitor genes repressed by *EAR*. (Bottom) Distribution of expression levels of HSC and progenitor genes in EB and *EAR* (3F) progenitors. Kolmogorov-Smirnov test was performed to assess the significance of differences between the two distributions.

Figure S4

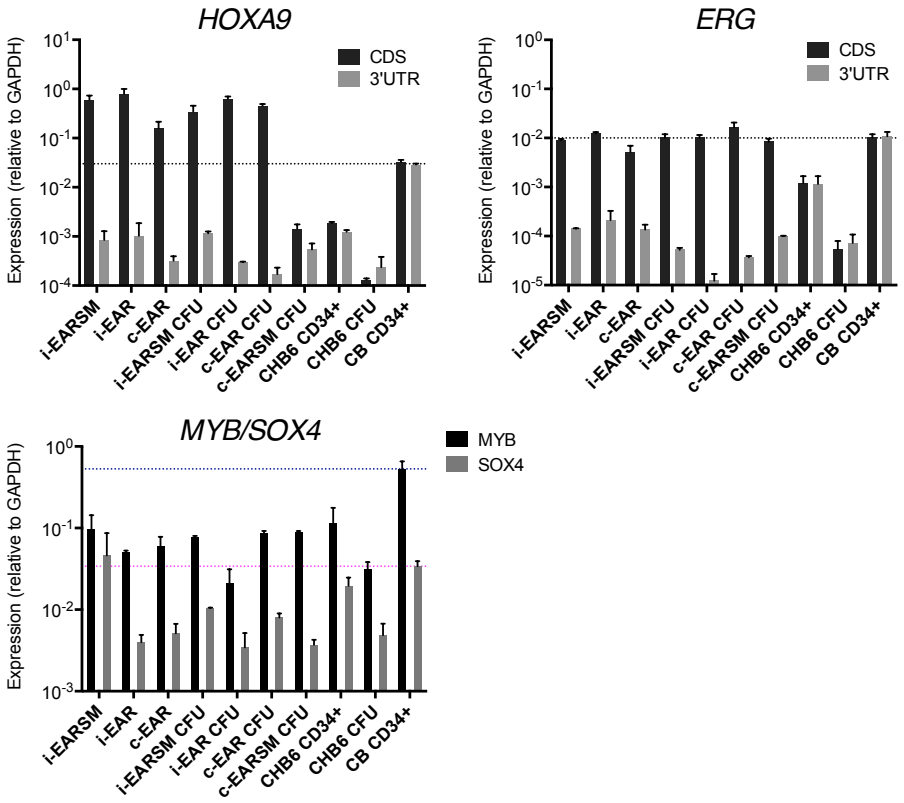


Figure S4, related to **Figure 5**. Quantitative PCR for endogenous and ectopic *HOXA9* and *ERG* expression with constitutive (labeled "c") and inducible (labeled "i") lentiviral vectors. The levels of *HOXA9* and *ERG* were measured using two sets of primers: coding region primers (CDS) which detect both endogenous and ectopic (lentiviral) transcripts, and 3'UTR primers which only detect endogenous transcripts. Both primary culture and dissociated colonies (CFU) are shown; for inducible constructs only +Dox condition (in which factors are ON) is shown. Controls include CHB6-ESC $CD34^+CD45^+$ progenitors, CHB6 CFU, and $CD34^+$ CB cells. All data are shown as mean ± SEM of 3 replicates.

Figure S5

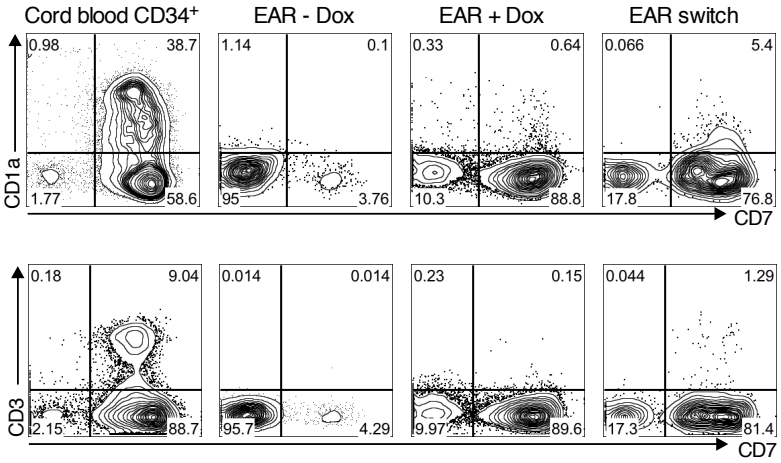


Figure S5, related to **Figure 5G**. T cell potential of re-specified progenitors in OP9-DL1 stromal co-culture. *EAR*-infected MSC-IPS1 progenitors were cultured for 2 weeks with Dox, and plated on OP9-DL1 without Dox (EAR - Dox), with Dox (EAR + Dox), or with Dox for the first 2 weeks only followed by Dox removal (EAR switch). CB $CD34^+$ cells served as a positive control. T cell development was assessed after 35 days using CD7, CD1a, and CD3. CD4/8 staining is shown in Figure 5G. All plots are gated on $CD45^+$ human cells.

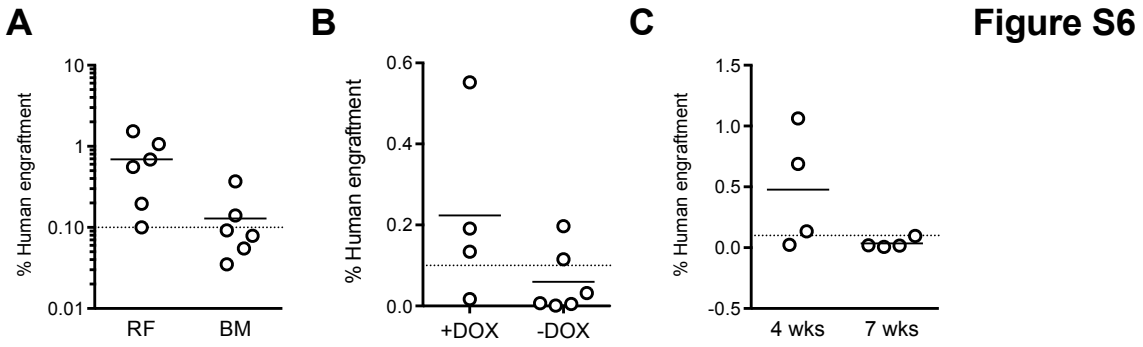
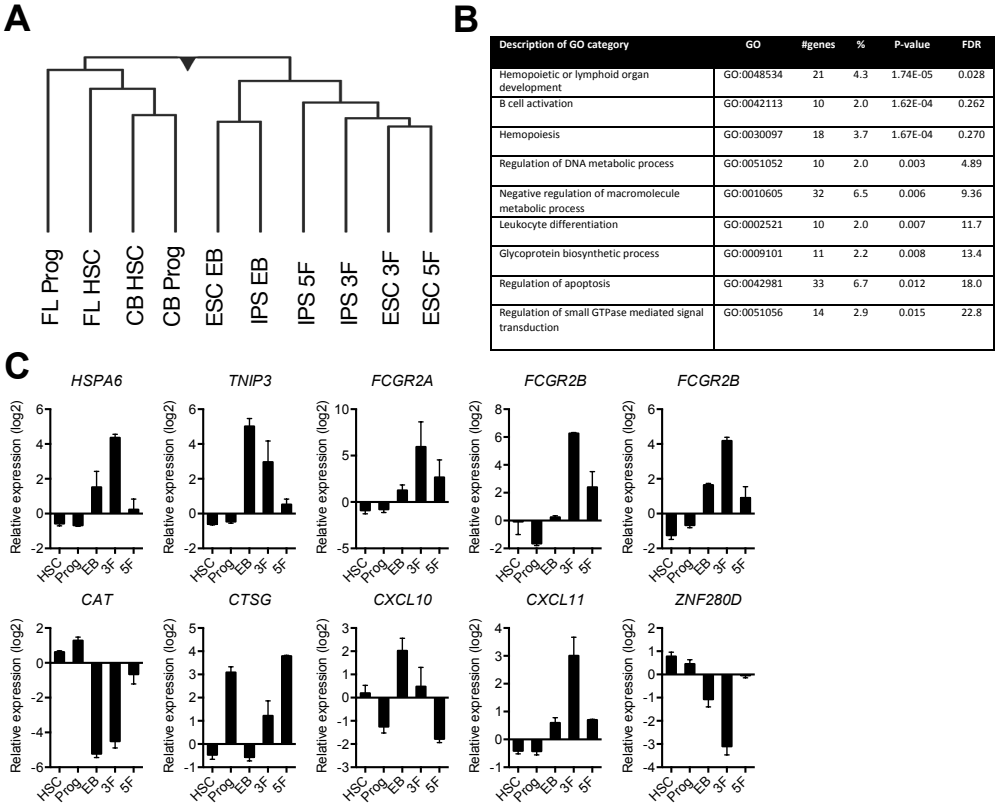


Figure S6, related to **Figure 6**. **(A)** Comparison of the engraftment in the injected right femur (RF) and the rest of the bone marrow (BM) at 4 weeks. Only mice engrafted in the right femur are shown. Dotted line marks the engraftment threshold of 0.1%. **(B)** The effect of continued transgene expression on engraftment of *EARSM* progenitors. Following transplant, mice were given Dox in the drinking water (+Dox) or Dox was removed (-Dox). Engraftment was analyzed after 4 weeks. **(C)** Extended transplantation extinguishes engraftment of *EARSM* progenitors. Engraftment was analyzed after 4 or 7 weeks.

Figure S7



Description of GO category	GO	#genes	%	P-value	FDR
Hemopoietic or lymphoid organ development	GO:0048534	21	4.3	1.74E-05	0.028
B cell activation	GO:0042113	10	2.0	1.62E-04	0.262
Hemopoiesis	GO:0030097	18	3.7	1.67E-04	0.270
Regulation of DNA metabolic process	GO:0051052	10	2.0	0.003	4.89
Negative regulation of macromolecule metabolic process	GO:0010605	32	6.5	0.006	9.36
Leukocyte differentiation	GO:0002521	10	2.0	0.007	11.7
Glycoprotein biosynthetic process	GO:0009101	11	2.2	0.008	13.4
Regulation of apoptosis	GO:0042981	33	6.7	0.012	18.0
Regulation of small GTPase mediated signal transduction	GO:0051056	14	2.9	0.015	22.8

Figure S7. Comparison of global gene expression between 3- (*EAR*) and 5-factor (*EARSM*) $CD34^{+}CD38^{-}$ progenitors. **(A)** Unsupervised clustering (Euclidean, average linkage) based on all probes. 3F = *EAR*, 5F = *EARSM*. **(B)** Top GO categories differentially enriched between 3F and 5F. (C) Expression of top 10 differentially expressed genes (ranked by fold change) between 5F and 3F (ttest, FDR<0.05; >2-fold) in primary HSCs, progenitors (Prog), and EB progenitors. All but two (*CTSG*, *CXCL10*) show aberrant expression in 3F-transduced cells relative to HSCs, which is corrected by 5F. Data are shown mean ± SEM of 6 biological samples for CB/FL HSCs; 5 samples for progenitors, EBs, and 3F; and 2 samples for 5F.

Supplemental Table legends

Table S1 – related to Figure 4.
List of differentially expressed genes between EB and *EAR* progenitors sorted based on $CD34^{+}CD38^{-}$. EBs derived from ESCs and IPSCs (fractions E and K) were compared with their respective re-specified fractions A and F. Ttest with multiple testing correction (FDR <0.05) and fold change cut-off (>2-fold) were applied. Genes are ranked by fold change.

Table S2 – related to Figure 4.
List of differentially expressed genes between primary HSCs and progenitors from CB and FL. CB and FL $CD34^{+}CD38^{-}$ HSCs (fractions C and G) were compared with CB and FL $CD34^{+}CD38^{+}$ progenitors (fractions D and H) respectively. Ttest with multiple testing correction (FDR <0.05) and fold change cut-off (>2-fold) were applied. Genes are ranked by fold change.

Table S3 – related to Figure 4.
List of differentially expressed genes between EB and primary HSCs sorted based on $CD34^{+}CD38^{-}$. EBs derived from ESC and IPSC (fractions E and K) were compared with primary HSCs from CB and FL (fractions C and G), respectively. Ttest with multiple testing correction (FDR <0.05) and fold change cut-off (>2-fold) were applied. Genes are ranked by fold change.

Table S4 – related to Figure 4.
List of differentially expressed genes between *EAR* progenitors and primary HSCs sorted based on $CD34^{+}CD38^{-}$. *EAR* cells re-specified from ESC and IPSC progenitors (fractions A and F) were compared with primary HSCs from CB and FL (fractions C and G), respectively. Ttest with multiple testing correction (FDR <0.05) and fold change cut-off (>2-fold) were applied. Genes are ranked by fold change.

Supplemental Experimental Procedures

Colony assays. At indicated times during the culture, $5x10^4$ cells were plated into 3 ml of complete methylcellulose (H3434; StemCell Technologies). Additional cytokines added were: 10 ng/ml FLT3, 10 ng/ml IL6, and 50 ng/ml TPO. The mixture was distributed into two 60 mm dishes and maintained in a humidified chamber for 14 days. Colonies were scored manually or using the BD Pathway 855 fluorescent imager.

Flow cytometry. Cells grown in progenitor culture or dissociated colonies were stained with the following antibody panels. HSC panel: CD90 PE (5E10; BD), CD38 PE-Cy5 (LS198-4-3; Clontech), CD49f Alexa647 (GoH3; BD), CD34 PE-Cy7 (8G12; BD), CD11b APC-Cy7 (ICRF44; BD), CD14 Alexa700 (M5E2; BD), and CD15 V450 (HI98; BD). Lineage panel: CD71 PE (M-A712; BD), CD34 APC (8G12), CD235a/Glycophorin A PE-Cy7 (11E4B-7-6; Coulter), CD11b APC-Cy7, CD14 Alexa700, and CD15 V450. All stains were performed with $<1x10^6$ cells per 100 μl staining buffer (PBS + 2% FBS) with 1:100 dilution of each antibody, 20 min at RT in dark. Compensation was performed by automated compensation with anti-mouse Ig^k and negative beads (BD). All acquisition was performed on BD Fortessa cytometer.

Microarray profiling. All cells were obtained viably frozen, thawed as described above, and stained with the 'HSC panel' of cell surface markers (see previous section). $CD34^+$ CB and $CD34^+$ FL cells were purchased from AllCells. Dissociated CHB6 and MSC-IPS1 EBs were prepared as above. Progenitors re-specified with either 3- (*EAR*) or 5-factors (*EARSM*) were cultured for 14 days, plated into colony assays, and dissociated into single-cell suspension. All cells were sorted for $CD34^+CD38^-$ HSC phenotype; in addition, CB/FL HSCs were also sorted for $CD34^+CD38^+$ progenitor phenotype. Between 10,000 - 50,000 cells were sorted for each cell type with 2 or 3 biological replicates. RNA was extracted using Trizol reagent (Invitrogen). Microarray data were analyzed per standard protocol using R/Bioconductor. Briefly, raw microarray signal intensities were RMA-summarized and quantile normalized and corrected for batch effects using ComBat batch correction (Bolstad et al., 2003; Irizarry et al., 2003; Johnson et al., 2007). We used

hierarchical clustering with the 'average' linkage method and correlation distance for the dendrogram. Stability of clusters were checked via boostrapping 500 iterations with pvclust package (Suzuki and Shimodaira, 2006). For a cluster with AU p-value > 0.95, the hypothesis that "the cluster does not exist" is rejected with significance level 0.05. GSEA was performed using the GSEA software with default parameters. Microarray data from this study has been deposited to GEO database under accession number GSE49938

Mouse transplantation. NOD/LtSz-scidIL2Rg^{null} (NSG) (Jackson Labs) mice were bred and housed at the Boston Children's Hospital animal care facility. Animal experiments were performed in accordance to institutional guidelines approved by BCH animal care committee. Intra-femoral transplants have been previously described. Briefly, 6 - 10 week old mice were irradiated (250 rads) 24 hrs before transplant. To ensure consistency between experiments, only female mice were used. Prior to transplantation, mice were temporarily sedated with isoflurane. A 27g needle was used to drill the right femur (injected femur), and 0.8-1.2 x10^6 cells were transplanted in a 25 μL volume using a 28.5g insulin needle. Baytril was administered in drinking water to prevent infections after irradiation. Doxycycline (1.0 mg/ml) was added to the drinking water to maintain transgene expression in vivo.

Assessment of human cell engraftment. Mice were sacrificed at indicated time points, and injected femur, uninjected femur and tibiae, and spleen were collected. Single cell suspension was prepared using standard flushing and cell dissociation techniques in 1 ml of IMDM. From that suspension, 100 μl of injected femur, 50 μl uninjected marrow and spleen (~1x10^6 cells) were stained in a total volume of 200 μl staining buffer. None of the samples were lysed with red blood cell lysis buffer as not to lyse human erythrocytes. Samples were stained with the following panel of human markers containing two non-competing CD45 clones: CD19 PE (4G7; BD), CD45 PE-Cy5 (Immu19.2; Coulter), CD33 APC (P67.6; BD), CD235a/Glycophorin A (11E4B-7-6; Coulter), CD45 APC-Cy7 (2D1; BD), CD5 Alexa700 (UCHT2; BD), and DAPI. To accurately detect and isolate human erythroid cells, the following panel of markers was used: CD71 PE (M-A712; BD), mTer119 PE-Cy5 (Ter-119; eBioscience), CD235a/Glycophorin A (11E4B-7-6;

Coulter), CD45 APC-Cy7 (BD), and DAPI. Staining was performed with 1:75 dilution of each antibody. Uninjected mouse marrow was used as a control for non-specific staining; CB mononuclear cells (AllCells) were used as a positive control for antibody staining and proper compensation. Compensation was performed using automated compensation with anti-mouse Ig^k and negative beads (BD). All acquisition was performed on BD Fortessa cytometer. Sorting was performed on a BD FACS Aria II cell sorter.

Quantitative and single cell PCR. RNA extraction was performed using Trizol reagent (Invitrogen). Reverse transcription was performed using Superscript III (>5,000 cells) or VILO reagent (<5,000 cells) (Invitrogen). Quantitative PCR was carried out in triplicate with SYBR Green (Applied Biosystems). Transcript abundance was calculated using the standard curve method. Single cells were sorted directly into PCR plates containing 2 ml 2% guanidine thiocyanate lysis buffer. As a control, 50 cells and no cells were sorted into some wells. Reverse transcription was performed using the VILO reagent, and split into 3 globin PCR reactions. Wells with Ct values below a cut-off based on 'no cell' controls were discarded, and the average of all single cell data was compared to '50 cell' controls to ensure absence of skewing.

Supplemental References

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118-127.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics *22*, 1540-1542.