Hematopoietic stem cell heterogeneity and age-associated platelet bias are evolutionarily conserved

Merve Aksöz, Grigore-Aristide Gafencu, Bilyana Stoilova, Mario Buono, Ying Zhang, Sven Turkalj, Yiran Meng, Niels Asger Jakobsen, Marlen Metzner, Sally-Ann Clark, Ryan Beveridge, Supat Thongjuea, Paresh Vyas[#] and Claus Nerlov[#]*.

MRC Molecular Hematology Unit, MRC Weatherall Institute of Molecular Medicine, Oxford NIHR BRC Haematology Theme, University of Oxford, Oxford, United Kingdom.

[#]Equal contribution

*Corresponding author. Email: claus.nerlov@imm.ox.ac.uk.

Abstract: Hematopoietic stem cells (HSC) reconstitute multi-lineage human hematopoiesis after clinical bone marrow transplantation and are the cells-of-origin of some hematological malignancies. Though HSC provide multi-lineage engraftment, individual murine HSCs are lineage-biased and contribute unequally to blood cell lineages. Here, we performed high-throughput single cell RNA sequencing in mice following xenograft with molecularly barcoded adult human bone marrow (BM) HSCs. We demonstrate that human individual BM HSCs are also functionally and transcriptionally lineage biased. Specifically, we identify platelet-biased and multi-lineage human HSCs. Quantitative comparison of transcriptomes from single HSCs from young, and aged, BM show that both the proportion of platelet-biased HSCs, and their level of transcriptional platelet priming, increases with age. Therefore, platelet-biased HSCs, and their increased prevalence and transcriptional platelet priming during ageing, are conserved features of mammalian evolution.

One-Sentence Summary: *In vivo* barcoding and single cell RNA sequencing identifies platelet-biased human bone marrow HSCs.

INTRODUCTION.

Hematopoietic stem cells (HSCs) sustain multi-lineage blood cell production throughout mammalian life. HSCs also reconstitute life-long blood cell production after hematopoietic cell transplantation (HCT), the most well-established and common form of clinical stem cell therapy (*1*). However, functional characterization of individual murine HSCs using single cell transplantation (*2-5*) and *in vivo* molecular barcoding (*6, 7*) has shown that in addition to HSCs with multi-lineage (ML) reconstitution capacity, a significant proportion of adult BM HSCs preferentially or exclusively generate a subset of hematopoietic lineages, and therefore are lineage-biased or fate-restricted (*1*).

A key characteristic of murine fate-restricted HSCs is higher output of platelets compared to other lineages (*5*), and such HSCs are therefore frequently referred to as platelet-biased (P-) HSCs. Single cell transplantation has also demonstrated that ML-HSCs have higher overall cellular output compared to fate-restricted HSCs (*5*), a finding recapitulated in cellular barcoding experiments. These studies also identified transcriptional signatures associated with high vs. low cellular output and multi-lineage vs. platelet-biased lineage reconstitution (*8*).

In the mouse, hematopoietic ageing is associated with significant expansion of P-HSCs (*9*). In contrast, the proportion of ML-HSCs, the only HSCs capable of lymphopoiesis, decreases with age, contributing to decreased adaptive immunity in aged individuals. Importantly, decreased lymphoid output has also been observed in aged human HSCs after xenografting into immune-deficient mice (*10*). Increased HSC transcriptional platelet priming may contribute directly to decreased lymphopoiesis during ageing, as inhibition of TGFβ signalling in aged HSCs leads to both decreased HSC platelet bias and increased lymphoid cell output (*11*). Finally, P-HSCs have been implicated as the cells-of-origin in a mouse model of essential thrombocythemia (ET) (*12*), and in human ET (*13*). In contrast to the mouse, characterization of the heterogeneity of human HSCs (hHSCs) is more limited. By xenografting, human cord blood (CB) hHSCs with distinct long-term reconstitution capacity (*14, 15*) and repopulation patterns (*16, 17*) have been identified. These studies did not address hHSC contributions to the platelet and erythroid lineages, or include adult BM hHSCs. However, tracking of viral insertions in patients treated with HCT of gene corrected cells has identified single cell-derived reconstitution patterns consistent with the presence of fate-restricted hHSCs in adult BM (*18, 19*), and RNA sequencing revealed increased transcriptional platelet-lineage priming of aged compared to young BM hHSCs (*20*).

To identify platelet-biased hHSC population we here use large-scale single cell RNA sequencing to show that molecular signatures of murine platelet-biased and multi-lineage HSCs identify corresponding, molecularly biased human HSC populations, and xenograft-based *in vivo* barcoding and classification of hHSCs based on their lineage output identified platelet-biased and multi-lineage hHSC populations. That were molecularly congruent with those identified by single cell RNA sequencing. Comparison of hHSCs from young and old donors showed an overall increase in molecular platelet bias of the hHSC population with age, accompanied by an increased

proportion of platelet-biased hHSCs. Therefore, both HSC functional heterogeneity and age-associated HSC platelet bias are evolutionarily conserved between human and mouse.

*

RESULTS

Identification of platelet-biased hHSCs by single cell transcriptome clustering.

To identify putative lineage-biased hHSCs we performed high-throughput single cell RNA sequencing (RNAseq) on phenotypic bone marrow hHSCs (defined as LIN–CD34+CD38–CD45RA–CD90+; Fig. S1) from 3 independent young adult donors (Table S1) using the 10X Chromium platform. A total of 13633 transcriptomes were assessed using area-under-the-curve (AUC) scores of transcriptional signatures associated with murine HSC lineage bias and output in individual hHSCs (*8*). This showed that multilineage- and high cellular output signatures were heterogeneously expressed and positively correlated at the single cell level, as were platelet-bias and low-cellular-output signatures, with individual hHSC generating a continuum along both these axes (Fig. 1A). Furthermore, at the transcriptional level, platelet-bias was correlated to stemness and self-renewal signatures, whereas the multi-lineage signature was correlated to high-output and proliferation signatures (Fig. 1B), analogous to what has been observed in the mouse (*8*). To identify P-hHSCs within the hHSC compartment we performed unsupervised clustering of single hHSCs (Fig. 1C). We found that cells from the 3 independent donors were evenly distributed across the Uniform Manifold Approximation and Projection (UMAP) (Fig. S2A), showing that donor effects were successfully regressed out of the dataset. We next calculated AUCell scores for the lineage bias signatures for each cluster (Fig. S2B), and grouped the clusters based on their median platelet-bias and multi-lineage AUCell scores using hierarchical clustering (Fig. 1D), identifying putative P-hHSC and ML-hHSC containing cluster groups (Fig. 1E). While the ML-hHSC populations appeared discontinuous in the two-dimensional UMAP representation, a 3-dimensional representation showed that it is a coherent population (Fig. S3). To determine if the putative P-hHSC populations had the expected molecular properties we compared their expression of lineage-bias and stemness/proliferation signatures to that of the putative ML-hHSC population. As expected, platelet-bias and low-output signatures were enriched in the putative P-hHSC population, as were signatures associated with stemness and quiescence; conversely, multi-lineage/lymphoid bias and proliferation signatures were depleted (Fig. 1F). The identification of putative P- and ML-hHSC populations also allowed the identification of genes differentially expressed between the two cell populations, representing the basis for the human signatures of these two hHSC subsets (Data S1). Here, a potential issue is that phenotypically defined hHSCs may not be a pure population. To assess this, we identified the cells within the Chromium data that co-expressed the known hHSC molecular markers HLF (*21*) and CRHBP (*22*), and performed similar analysis (Fig. S4A). This identified the same molecularly defined P- and ML-hHSC clusters (Fig. S4B-E), and the P- and ML-hHSC signatures generated from this clustering showed >90% overlap with those identified from the total hHSC population (Fig. S4F). The analysis is therefore robust against impurity of phenotypic hHSCs. This analysis identifies hHSC subsets with

the molecular properties of platelet-biased and multi-lineage HSCs, and shows that, as in the mouse, these are correlated to low- and high-output molecular signatures, respectively.

Identification of platelet-biased hHSCs by molecular barcoding.

To determine if hHSCs with functional lineage bias could be identified in adult BM, and whether the transcriptional profiles of computationally and functionally defined lineage-biased hHSCs were correlated, we combined barcoding and xenografting of BM hHSCs to simultaneously measure hHSC lineage output and characterize individual hHSC transcriptomes (Fig 2A). This involves transducing purified hHSCs (defined as above, Fig. S1) with a lentiviral barcode library (*23*) followed by xenografting into non-irradiated c-Kit–deficient immune-deficient NSG mice (NSGW41; (*24*)). The barcoded lentiviral vector expresses EGFP, allowing transduced cells to be identified and purified. Furthermore, as the barcode is transcribed as part of the proviral mRNA they can be identified by RNAseq of FACS-purified cells. Engraftment of hHSCs was verified by analysis of BM aspirates at 12- and 16-weeks post-transplant. We identified 7 mice, transplanted with hHSCs from 3 different donors, with high levels of human BM engraftment (30-80%) (Fig. S5A) and a significant proportion (10-40%) of transduced EGFP+ cells in the human BM fraction for subsequent analysis (Fig. S5B). The engrafted mice were analyzed at the 16-week time point in order to focus on long-term repopulating hHSCs. BM cells were harvested and divided into CD34+ and CD34– fractions by magnetic cell sorting. From the CD34– fraction, EGFP+ cells were sorted for bulk RNAseq from the following populations: B-lymphoid cells (Bly; hCD45+CD34–CD33–CD19+), myeloid cells (Myl: hCD45+CD34–CD33+CD19–), erythroid progenitors (EryP; CD235+CD71+) and megakaryocytes (Mk; hCD45+CD34–CD33–CD19–CD41+SSChi) (Fig. 2B). The sorted populations were validated using AUCell scores of lineage-specific signatures (Fig. S5C). From the CD34+ BM fraction individual EGFP+ hHSCs were sorted (Fig. 2C) for single cell RNAseq. The number of barcodes retrieved from bulk RNAseq in each xenograft experiment is shown in Table S4. Similarly, the barcodes retrieved from hHSC single cell transcriptomes in each xenografted mouse is shown in Table S5. In total, 57 barcodes, representing 377 single hHSCs, were detected in both bulk and single cell RNAseq (Fig. 2D). In contrast, 29 of the identified hHSC clones did not contribute detectably to hematopoietic cell output, consistent with previous reports using viral integration sites to detect clonal hHSC output after xenografting (*25*), or autologous transplantation (*18*), where a high proportion of non-contributing hHSCs was also detected. The 57 hHSC clones that contributed to hematopoiesis were hierarchically clustered according to their lineage output, as determined by the distribution of the corresponding barcode in the bulk RNAseq of the 4 lineage-restricted populations. This identified 6 clusters with distinct output patterns, the 4 major clusters representing P- (cluster E), platelet-erythroid- (PE – cluster C), platelet-erythroid-myeloid-biased (PEM – cluster A) and ML (cluster B) patterns (Fig 2E), output patterns similar to those observed after transplantation of single murine HSCs (*5*), indicating conservation of HSC fate-restriction patterns between human and mouse. We here focused on the P- and ML-reconstitution patterns where molecular correlates have been identified in the mouse, as this provides a means of cross-validation. Some of the observed reconstitution patterns (erythroid-biased, B lymphoid-biased) have not been observed in single cell

transplantation of murine HSCs (*5*). However, erythroid-biased clonal output has been observed after *ex vivo* barcoding and transplantation of murine HSPCs (*26*). The validity and significance of these output patterns is therefore not clear and need further investigation. This analysis identified functionally lineage-biased hHSCs and their associated molecular signatures, and in particular showed the existence of platelet-biased hHSCs.

Platelet-biased hHSCs defined by clustering and barcoding are molecularly correlated

We next used differential gene expression analysis to identify marker genes for each cluster (Data S6), and the significance of overlap of each cluster signature with the computationally generated P- and ML-hHSC signatures from the 10X Chromium data calculated (Fig. 3A). This showed selective overlap of marker genes from Cluster E, which contained hHSCs with P-biased lineage output, with the P-hHSC signature. Similarly, marker genes for Cluster B, which contains hHSCs with multi-lineage output, showed significant overlap with the ML-hHSC signature. Consistent with this, comparison of barcoded P- and ML-hHSC transcriptomes showed significant overlap between genes differentially expressed between barcoded and computationally generated P- and ML-HSCs (Fig. 3B; Data S2). In addition, AUCell scores of computationally and barcode-derived P-hHSC signatures showed very high correlation across the P-biased, intermediate and ML-hHSCs for all 3 donors analysed (Fig. 3C). Finally, Gene Set Enrichment Analysis (GSEA)-based comparison of P- and ML-hHSCs defined by barcoding and clustering, respectively, showed a similar enrichment pattern, with stemness and quiescence signatures enriched in P-hHSCs and proliferation signatures in ML-hHSCs (Fig. 3D). This showed that the computationally and functionally identified P-biased and ML hHSC populations were molecularly congruent, consistent with the two methodologies identifying highly similar hHSC sub-populations.

Platelet-biased hHSC abundance increases during physiological ageing

To address the effect of ageing on hHSC platelet-bias we next performed single cell RNAseq of aged hHSCs. For this purpose, LIN–CD34+CD38–CD45RA–CD90+ hHSCs were sorted from 3 aged donors (68-75 years old – Table S1). We found that the frequency of hHSCs was increased in old compared to young BM (Fig. 4A), as previously reported (*10*). Single cell transcriptomes were generated from each of the 3 old donors. After batch correction, all donors showed similar distribution across the UMAP projection (Fig. S6A), and we observed the same correlation of HSC output and lineage signatures in old hHSCs as was observed in young hHSCs (Fig. S6B). Clusters were identified (Fig. 4B), and cluster groups generated by hierarchical clustering using platelet-bias and multi-lineage AUCell scores (Fig. S6C), mirroring the analysis of the young hHSCs (Fig. 4C), allowing old ML- and P-hHSC populations to be identified (Fig. 4D). Analogous to the young donor hHSC dataset analysis, we also identified genes differentially expressed between the putative old P- and ML-hHSCs (Data S1). To determine if the overall hHSC molecular lineage bias changed with age we first identified P- and ML-hHSC signatures in both young and old bone marrow and used the union of these as age-agnostic human P- and ML-hHSC signatures (Fig. S7A, Data S3). AUCell scores of these signatures showed up-regulation of the P-hHSC signature, and down-regulation of the ML-hHSC signature during ageing (Fig. 4E). Extending this analysis to

include fetal liver (FL)-hHSCs (Fig. S8A) showed that these had lower expression of P-hHSC (Fig. S8B) and higher expression of ML-hHSC signatures (Fig. S8C) compared to young hHSCs, indicating that molecular platelet bias increases across the entire hHSC developmental trajectory. In addition, we observed a consistent and significant increase in P-hHSC frequency, and a commensurate decrease in ML-hHSC abundance (Fig. 4F) in old compared to young hHSCs, Therefore, the overall increase in molecular platelet bias during human ageing was due to both increased prevalence of P-hHSCs and increased transcriptional platelet priming within this population, consistent with previous observations in the mouse (*9*). As for young hHSCs, restricting the analysis to old CHRBP+HLF+ hHSCs (Fig. S9A) generated the same clustering (Fig. S9B-E), showed the same age-related increase in P-hHSCs and decrease in ML-hHSCs (Fig. S9F), and identified P- and ML-hHSC signatures with >90% overlap with those generated from the total phenotypic hHSC population (Fig. S9G), showing robustness of the analysis against potential impurity of the hHSC population. The enrichment pattern of stemness and proliferation signatures in old P- and ML-hHSC clusters was conserved relative to young hHSCs (Fig. S7B, compared to Fig. 1F), and similar analysis using canonical Hallmark gene signatures (MSigDB) also showed a high degree of conservation between young and old hHSCs. In particular, metabolic (oxidative phosphorylation, glycolysis) and proliferation-associated signatures (including Myc targets, E2F targets, G2M checkpoint) were enriched in both young and old ML-hHSC, whereas inflammatory pathways (IL-6 and TNFα signalling) were enriched in both young and old P-hHSCs (Fig. S7C). Finally, to determine if the increased molecular platelet bias at the hHSC level observed during ageing resulted in bias towards platelet production, we performed limiting dilution analysis of myeloid and Mk colony forming units in young and old bone marrow, using the CD34+ progenitor fraction. We observed a higher frequency of CFU-MK (Fig. 4H), and a lower frequency of CFU-GM (Fig. 4I), consistent with increased prevalence of hHSCs biased towards or restricted to platelet production. These results show that hHSC platelet bias increases with age, with a corresponding increase in the prevalence of platelet-biased hHSCs and megakaryocyte bias of the progenitor compartment.

DISCUSSION

Studies using comprehensive readout of hematopoietic lineage outputs from individual murine bone marrow HSCs have consistently identified P- and platelet-myeloid biased HSC subtypes (*3, 5, 7*). However, parallel human studies have generally been performed using cord blood hHSCs, and not included measurement of the erythroid- and platelet-lineages. By barcoding of adult hHSCs and comprehensive lineage readout we here find that hHSCs with similar fate-restrictions to those identified in the mouse (*5*), and P-hHSCs in particular, could be identified in human bone marrow. In addition, comparing the single cell transcriptomes of P- and ML-hHSCs, the differences in expression of stemness- and proliferation signatures were found to be conserved in hHSCs. P- and ML-hHSCs could also be identified using the previously defined murine HSC lineage-bias signatures (*8*), and P-hHSCs defined in this manner were transcriptionally similar to

those identified by barcoding, with correlated expression of platelet-bias signatures and similar pathway enrichment.

Increased molecular and functional platelet bias is a key feature of aged mHSCs (*9, 20*), and this involves both increased prevalence of P-mHSCs and their increased platelet-lineage priming at the single cell level (*9*). We here find that both of these features of murine hematopoietic ageing are conserved in humans. Similarly, we observed an increased functionally defined megakaryocyte progenitors, and a decrease in myeloid progenitors, consistent with hHSC platelet bias being translated into progenitor bias, also as previously seen in the mouse. Furthermore, using GSEA we identify a number of molecular pathways that are consistently enriched in young and old P-hHSC such as inflammatory signalling pathways. Increased interferon signalling, as observed in the young P-hHSCs, has been observed in platelet-biased malignant hHSCs in essential thrombocythemia (ET), and our observations indicate that this is an inherent property of P-hHSCs that is conserved upon their transformation by *JAK2* mutation (*12*). P-hHSCs in both young and old donors are characterized by enrichment in gene sets associated with high self-renewal capacity and quiescence (*8, 27, 28*). Conversely ML-hHSCs showed enrichment in cell cycle progression gene sets (*29*) and MTORC1 signalling (*30*) indicate a more activated state, consistent with their higher cellular output (*8*). P-mHSCs have previously been observed to occupy specific niches enriched in megakaryocytes (*31*), which produce Cxcl4 and TGFβ1 critical for maintenance of HSC quiescence (*32*). The distinct inflammatory and metabolic signatures of P- and ML-hHSCs are consistent with P-hHSCs inhabiting similar niches within human bone marrow.

While our results show the existence of molecularly and functionally hHSC heterogeneity in an experimental setting, it remains important to validate these findings in native human hematopoiesis. Similarly, the physiological contributions of human P- and ML-HSCs, and how these change during ageing remain to be understood. In the mouse P-HSCs and ML-HSCs were found to utilize a separate cellular pathway of megakaryopoiesis (*33*), and to generate megakaryocytes with distinct functions in immunity and the HSC niche (*34*). Furthermore,the P-mHSC pathway was found to be exacerbated by and to promote thrombosis during ageing (*35*). Given the observed conservation of P-HSCs and their age-dependent increase during ageing it will now be highly relevant to investigate physiological roles of HSC heterogeneity in general, and P-hHSC in particular, in the human system, including the existence of distinct cellular pathways of human megakaryopoiesis.

Our results show that functionally platelet-biased hHSCs and their molecular programming are conserved features of mouse and human hematopoiesis, indicating a high degree of conservation of HSC heterogeneity and the molecular properties of lineage biased HSCs between mammalian species. This will inform future studies across multiple areas, including hematopoietic regeneration, ageing, and in the initiation of hematological malignancies where HSC heterogeneity plays a key role.

MATERIALS AND METHODS

Study design.

The aim of this study was to identify platelet-biased human HSCs, and to determine if their prevalence increases with age. For this purpose, we performed single cell RNA sequencing of purified human HSCs, and used gene expression signatures from lineage-biased murine HSCs to identify equivalent lineage-biased human HSC populations. In parallel, xenograft-based *in vivo* barcoding was used to identify and molecularly profile functionally platelet-biased human HSCs. To cross-validate these data sets, we compared the computationally and functionally identified platelet-biased human HSCs, which showed that their gene expression signatures were highly correlated. Finally, to assess how the prevalence of platelet-biased human HSCs changes with age, we performed integrated analysis of single cell RNA sequencing data of HSCs obtained from young and old human donors, which showed increased molecular platelet bias of human HSCs during ageing, as well as increased prevalence of platelet-biased human HSCs.

Animals. The NSGW41 (NOD.Cg-Kit$^{W\text{-}41J}$Tyr^{+}Prkdc^{scid}Il2rg^{tm1Wjl}/ThomJ) mouse strain (*24*) was obtained from the Jackson Laboratory (Stock #026622) and bred in the animal facility at the University of Oxford (Biomedical Services). Female mice aged 8-11 weeks were used for xenografting. All experimental procedures, mouse breeding, and maintenance were performed according to UK Home Office regulations and with the ethical approval from the University of Oxford Medical Sciences Division Animal Welfare and Ethical Review Board.

Human bone marrow. Fresh BM from healthy young volunteers (21 to 34 years old males) were purchased from AllCells (Berkeley, CA) or Lonza (Lonza Bioscience, Basel, Switzerland). Bone marrow from older donors was obtained from orthoplastic surgery after informed consent under the Mechanisms of Age-Related Clonal Haematopoiesis (MARCH) Study (REC Ref: 17/YH/0382). BM cells were filtered through 70 μm cell strainers (Corning), mononuclear cells (MNCs) were isolated using Ficoll (GE Healthcare) density gradient centrifugation and CD34+ cell enrichment was carried out according to the manufacturer's protocol (Miltenyi Biotech).

Cell sorting. For sorting of human HSCs BM, CD34^{+} cells were stained with the following lineage markers (CD2, CD3, CD4, CD7, CD8, CD11b, CD14, CD19, CD20, CD56 and CD235a, all PE-Cy5 conjugated), CD34-APC, CD38-PE-Texas Red, CD90-PE, CD45RA- FITC, CD10-BV650, and CD123-PE-Cy7. 7-amino-actinomycin D (7AAD, Sigma Aldrich) viability dye was added to the samples prior to sorting to exclude non-viable cells. The clones and dilutions of all antibodies are listed in Table S2. Cell sorting and flow cytometry was performed using a BD FACSAria Fusion Cell Sorter with FACSDIVA software v.8.0.1 (Becton Dickinson and Company, Franklin Lakes, US-NJ) and BD LSRFortessa or BD LSR II analysers, respectively. Data analysis was performed using FlowJo software (FlowJo LLC, v10.5.3).

Lentiviral barcoding and xenotransplantation. HSCs were transduced in SFEM serum-free media (Stem Cell Technologies) supplemented with human Flt3 ligand (100 ng/ml), KitL (100 ng/ml), Thpo (50 ng/ml), and IL-6 (50 ng/ml) (all from Peprotech). The peGZ2-linkerBC322 barcoding library (*23*) consisting of 725 barcodes was prepared as described (*36*). Purified HSCs were transduced at an MOI of 100-125 in StemSpan serum-free media supplemented with human cytokines Flt3 ligand, KitL, Thpo, and IL-6 as above for 18 hours at 37ºC and 5% CO_2, followed by intraosseous injection into 8-11 weeks old female NSGW41 murine recipients. The number of HSC transplanted per recipient and transduction efficiency (as measured by EGFP positivity) are shown in Table S4.

Purification of xenografted cells. BM cells were harvested from femurs, tibias, hips, sternum, and spine 16-weeks post-transplantations. Human CD34 magnetic separation then was performed to isolate $CD34^+$ and $CD34^-$ cell populations. $CD34^+$ cells were stained as above with the addition of human CD45-BV650 and mouse CD45-BV605. Single cells were index-sorted into 96-well PCR plates (Thermo Scientific) containing 4µl of lysis buffer using a 85µm nozzle (*37*) for single cell RNAseq. $CD34^-$ cells were stained with the following antibody panel: CD71-APC-Cy7, CD235a-PE-Cy5, CD19-PE-Cy7, CD33-PE, CD41-AF700, CD34-APC, human CD45-BV650 and mouse CD45 BV605. Hoechst 33342 was used as a viability marker. Barcoded human ($mCD45^-$) erythroid cells (EryP: $CD235^+CD71^+EGFP^+$), B-cells (Bly: $hCD45^+CD19^+CD33^-EGFP^+$), myeloid cells (Myl: $hCD45^+CD19^-CD33^+EGFP^+$), and megakaryocytes (Mk: $hCD45^+CD41^+CD19^-CD33^-EGFP^+$) were sorted directly into PCR tubes containing 4µl of lysis buffer (50 or 100 cells/biological replicate) for bulk RNAseq using a 130µm nozzle. The total number of cells used for RNA sequencing per population for each recipient is shown in Table S4.

Single cell and bulk RNA sequencing library preparation. Single-cell and bulk libraries were generated using Smart-seq+ (single cell RNAseq) (*38*) and targeted Smart-seq2 (bulk RNAseq) (*39*), respectively. Custom primers for barcoding amplification were included during the initial cDNA preparation and pre-amplification steps. The sequences of the primers used in single-cell and bulk RNA-seq is provided in Table S3. 22 cycles were performed for PCR amplification. After pre-amplification, cDNA libraries were purified using Ampure XP (Beckman Coulter) magnetic beads according to the manufacturer's instructions, in a ratio of 1.2 to 1 with cDNA. Following tagmentation and 12-cycles of index PCR using Nextera adapters (Illumina), libraries were bead-cleaned up using a 1:1 ratio of Ampure XP beads (Beckman Coulter). All cDNA and indexed libraries were quality checked for their fragment sizes using Agilent high-sensitivity chips on an Agilent 2100 Bioanalyzer (Agilent Technologies) or an Agilent fragment analyzer high-sensitivity kit with an Agilent fragment analyzer (Agilent Technologies). Library concentrations were assessed using a Qubit dsDNA HS kit (Invitrogen) and a Qubit fluorometer (Invitrogen). Dual-indexed libraries were pooled to a final concentration of 2 nM and sequenced on an Illumina NextSeq500 instrument (75-bp single-end reads) using a NextSeq 500/550 High Output kit v2.5 (Illumina).

Single cell 10X Chromium library preparation. HSCs were purified from young and aged human BM (N=3 for both ages; Table S1). Libraries were prepared using the Chromium Single Cell 3' Reagent Kits v2 (10X Genomics, USA). cDNA concentrations were assessed using a Qubit dsDNA HS kit (Invitrogen) and a Qubit fluorometer (Invitrogen). Fragment sizes were checked using an Agilent high-sensitivity chip (Agilent Technologies) on an Agilent 2100 Bioanalyzer (Agilent Technologies) according to the manufacturer's instructions. Libraries were sequenced on a NovaSeq6000 S4 flow cell (NovoGene).

Single cell and bulk RNAseq analysis. Nextera adapter sequences and low quality reads were removed from demultiplexed FASTQ files using Trim Galore (v0.5.0) (*40*). Reads were aligned to the hg38 human reference genome using STAR (v2.6.1d) (*41*). Unique read counts were quantified using featureCounts (subread/v1.6.2) (*42*). Quality control (QC) metrics were generated using MultiQC (v0.9) (*43*). Lineage-specific signatures were generated from published data (*44*), comparing myeloid cells (monocytes (GSM609766-70) and neutrophils (GSM609719-22)), megakaryocytes (GSM609746-52), erythroblasts (GSM609691-96) and B-cells (GSM609648-53). For each lineage the signature was defined as the top 50 genes up-regulated in a lineage compared to the 3 other lineages ($P<0.05$, log2 fold change>0.25) using GEO2R and DESeq2. Bulk RNA-seq counts data were ranked using the AUCell *AUCell_buildRankings* function, and signature scores for sorted cell populations were generated using the *AUCell_calcAUC* function with default parameters.

Barcode detection and analysis. The merged FastQ files were used for the barcode detection in single-cells and bulk samples. The barcodes in HSCs and their lineage outputs were identified using a custom R script to quantify the barcode count number. All reads were searched for sequence matching the following barcode pattern: NNNACNNNGTNNNCGNNNTANNNCANNNTGNNN. The biological replicates of Bly, Myl, Mk, and EryP lineages were aggregated and the total barcode reads for each barcode and each lineage was calculated. From 472 single HSC transcriptomes 86 different barcodes were identified, each representing a clonal hHSC population. Of these 86 barcodes, 57 were also present in the 159 barcodes retrieved from lineage-restricted cell populations, and these were used for further analysis. To estimate the number of barcode-reads required for reliable detection of all lineages generated by an hHSC clone a random draw was simulated from a population of 10^6 equally distributed across N subpopulation (N being the number of lineages generated by the clone, ranging from 2-4). 10^5 simulated draws were carried out, and the frequency of detection of 1 to N lineages calculated. The cutoff was set at a 99% chance of detecting all 4 lineages generated by a multi-lineage hHSC (20 reads – see Fig. S10). Barcodes with less than 20 total reads were therefore filtered out. The barcode frequencies were calculated by dividing the barcode sum values by the total aligned reads of their respective lineage. This normalizes the barcode counts to the sequencing depth. Next, for a reliable comparison of the lineages, the barcode detection level was normalized, as barcodes make up different fractions of the transcriptome of different lineages. For this reason, scaling factors were generated using the multilineage clones (clones with barcode reads for all

lineages, where the average contribution of the four lineages was assumed to be even, as observed in mouse multilineage clones) as a reference and used to normalize the clone barcode counts across lineages. To classify HSC clones by their fate, the normalized barcode reads were divided by their sum for each clone, and the resulting scaled values were used for hierarchical clustering of HSC clones using the R dist function to calculate Euclidean distance and the R hclust function to perform wardD2 agglomeration. Clusters were identified by cutting the dendrogram to six clusters using the R cutree function.

Chromium 10X scRNA-seq data processing and quality control. The raw fastQ files were aligned against the *H. sapiens* GRCh38 (Ensembl 93) reference genome (10X Cell Ranger reference GRCh38 v3.1.0) and quantified recommended using the Cell Ranger pipeline (v3.1.0) at default parameters. The *count* functionality of the Cell Ranger pipeline was utilised to generate the filtered cell x gene raw UMI family counts for the data from each sample individually and further processed using Seurat (v4.1.0) (*45*). Quality control was performed at the level of the data from each donor separately by first filtering out cells with < 200 genes detected and then retaining only cells with < 10% mitochondrial reads and gene counts less than double the median gene count detected in the data for that donor. Genes detected in less than 3 cells were removed. Following data quality control data from old donors was processed separately from young donor data, but following the same procedure. Thus, we'll describe the processing procedure for data of young donors, as an example. As to mitigate for batch effects generated during the library preparation procedures, data that passed quality control from each donor individually was considered a batch, and following gene expression normalization, scaling and highly variable gene selection using Seurat's *scTransform* command, 2000 integration anchors across donors were defined using the Seurat's *FindIntegrationAnchors* function. Once the integration anchors were defined, data was batch corrected using the Seurat *IntegrateData* function, setting normalization.method = "SCT" and the rest of the function's parameter at default values. Following the batch correction of the data, principal component analysis (PCA) was conducted using Seurat's *RunPCA* function on the data from the 2000 integration anchor genes. Using an elbow plot describing a ranking of principle components based on the percentage of variance explained by each one, the first 20 PCs were used as the input for the Uniform manifold approximation and projection (UMAP) visualisation, performed using Seurat's *RunUMAP* function , as well as constructing a shared nearest-neighbour (SNN) graph used for cell clustering, with edges drawn between the cells with similar gene expression patterns using Seurat's *FindNeighbors* function at default parameters. Cells were clustered into communities (clusters) of transcriptomic similar cell starting from the previously computed SNN graph and using the Louvain algorithm's implementation in Seurat's *FindClusters* function used at default parameters, except from the resolution parameter which comprised of the following range of values: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. The resolution parameter in the *FindClusters* function used to determine the most robust clustering configuration was determined following a previously described methodology (*46*). Clusters were assigned into clusters groups, reported as P-hHSCs, ML-hHSCs, or an intermediate HSC phenotype based on the hierarchical clustering of the Euclidean distance computed for each cluster's median platelet bias and

multilineage AUCell scores using the Ward agglomeration method as implemented in R (v4.1.1). For the analysis of the fetal HSC dataset, the pre-processed healthy donor CD34+ enriched human cord blood, fetal liver and fetal bone marrow MNC CITE-seq dataset was sourced from the repository that was indicated in publication of origin (*47*) (https://developmental.cellatlas.io/fetal-bone-marrow) as well as the raw UMI by cell matrix single cell RNA-seq UMI by cell matrix matrix deposited at GSE166895 and associated with the previously mentioned CITE-seq dataset. Next, the data labelled by the authors as being HSC/MPPs from fetal bone marrow samples from the previously mentioned fetal dataset were used for further analysis. The HSC rich subset of the fetal HSC/MPP dataset was extracted using the protein expression levels of CD90 as a criteria to discern between human HSCs and MPPs, by first computing a CD90 protein expression scores using the scGate methodology (*48*) (scGate function from the scGate (v1.2.0) (*48*) R package used at default parameters except for the following ones which have been altered: k.param = 20, min.cells = 20, nfeatures = number of proteins asseded in the dataset, maxRank = half of the number of proteins asseded in the dataset) and then using model-based clustering, classification, and density estimation based on finite normal mixture modelling (*49*) on the previously computed CD90 protein expression score distribution (Mclust function from mclust (v6.0.0) (*49*) R package with the modelNames = "E") to select the cell populations from the HSC/MPP dataset with a distinctly high CD90 protein expression. The fetal bone marrow HSC-rich CITE-seq dataset was then processed using a similar procedure as the young donor HSC scRNA-seq dataset, with the amendments that clusters were not grouped in cluster groups, as only three stable clusters were determined in this dataset, and Seurat's FindIntegrationAnchors (dims = 1: cell count of the smallest technical batch -1, k.filter = NA, k.score = cell count of the smallest technical batch -1) and IntegrateData (k.weight = cell count of the smallest technical batch) functions' parameters needed to be altered to accommodate for the size of the experimental batches.

Differential gene expression analysis. Differentially expressed genes between distinct HSC clone types were identified using Seurat's (v4.0.1) *FindMarkers* function using the following parameters changed from the default ones: logfc.threshold = 0, test.use = "MAST", assay = "RNA". The p-values for each gene entry in the output of the Seurat's *FindMarkers* function were corrected for multiple comparisons using the Benjamini & Hochberg correction as implemented in the *p.adjust* function of the stats R package (v4.1.1). Genes were considered differentially expressed between clusters or HSC clone types if they had a corrected p-value < 0.05. Differential gene expression analysis (DGEA) comparing HSCs from different cluster groups by first subsetting the dataset of interest to contain only the cells that were used in the comparison and converting the dataset to a SingleCellExperiment object (*50*) suitable for the MAST DGEA methodology (*51*). Next the SingleCellExperiment was filtered to contain only genes that have non-zero normalised counts. Then genes that are sufficiently highly expressed to afford the convergence of the mixed effects model used for MAST DGEA were identified by using MAST's (v1.20.0 or higher) (*51*) filterLowExpressedGenes function with the parameter threshold = 0.1. Next a zero-inflated regression model was created using MAST for DGEA using MAST's (v1.20.0 or higher) (*51*) zlm function following design formula ~condition + cngeneson + (1|orig.ident), in which condition

represents the vector of cell labels for the conditions the DGEA is carried out on, cngeneson is a vector containing the dataset-scaled number of genes detected per cell as way of informing the model on the detection profile of each gene in the dataset (*51*) and 1|orig.ident representing the random intercept component of the model based on the biological sample identity that aims to mitigate for the pseudoreplication bias inherent to single cell DGEA methods (*52*) as well as the parameters 'glmer', ebayes=FALSE, fitArgsD=list(nAGQ=0) to accommodate for this design. In order to next identify which are the differentially expressed genes between the samples or cluster groups of interest a likelihood ratio test is used by using MAST's (v1.20.0 or higher) (*51*) summary function on the previously computed model with the following parameters doLRT='conditiontest', indicating the element in the design matrix in the condition column that is the group in which genes would be upregulated, and fitArgsD=list(nAGQ=0, tolPwrss=1e-3) to accommodate for the requirements of the mixed-effect nature of the model. The p-value adjusted for multiple hypothesis testing of the likelihood ratio test run for each gene is adjusted using the p.adjust function of the stats R package (v4.1.1) using as the input the p-values of the likelihood ratio test and the Benjamini & Hochberg p-adjusted methodology marked as "fdr" is used for the method parameter of the function.

Gene expression score computation and assessment. Gene expression scores for the gene sets of interest were computed using AUCell (v 1.16.0 or higher) module of the SCENIC suite for scRNA-seq data analysis (*53*). The gene symbols for the gene sets derived from murine data used for computing AUCell scores (*8, 28, 54-58*) were converted into their human counterparts when possible by utilizing the *getLDS* function as implemented in biomaRt (v2.50.3) (*59*) and utilising the *H. sapiens* (GRCh38.p13) and *M. musculus* (GRCm39) Ensembl gene symbol annotation for the gene symbol conversion. The signatures used to compute AUCell scores can be found in Data S3. In order to ensure the generation of robust gene signatures, the P-hHSCs gene expression signature in the context of the young donor human HSC dataset, for example, was done by determining the common genes between the upregulated differentially expressed genes (FDR < 0.05) when comparing the P-hHSCs and the ML-hHSC and the upregulated differentially expressed genes (FDR < 0.05) between the P-hHSCs and the hHSC that were not considered ML-hHSC in that dataset. The ML-hHSC signature was determined in an analogous way to the P-hHSCs signature. The P- and ML- hHSC signatures in the context of the old donor dataset were generated using the same procedure used for the young donor dataset. Following the processing of the gene sets, at the level of each cell in the dataset, the genes were ranked according to their expression from the ones with the highest expression to the one with the lowest one using the AUCell's *AUCell_buildRankings* function at default parameters. Next, the AUCell's *AUCell_calcAUC* function was used at default parameters to determine if the gene sets of interest are enriched in the top ranked genes according to their expression at the level of each of the dataset's cells. The differences in gene expression score distributions across the clusters determined in the datasets as well as the type of donors were evaluated using the Wilcoxon rank-sum test, as implemented in the *pairwise.wilcox.test* and *wilcox.test* function in the stats R package (v4.1.1). When comparing gene expression scores between 3 or more clusters, the significance of

the differences in distributions observed was quantified by using the Wilcoxon rank-sum test in a pairwise fashion and the p-value of the test was corrected for multiple comparisons using the Benjamini & Hochberg correction as implemented in the *p.adjust* function of the stats R package (v4.1.1). The Pearson's R or Kendall's τ correlation coefficients were computed using the *cor.test* function of the stats R package (v4.1.1) to quantify the strength and direction of association between 2 types of AUCell scores; the choice of one type of coefficient over the other was predicated on the normality of the distribution of the AUCell scores investigated in. this study.

Differential cell type abundance analysis. In order to analyse if the frequency of HSCs from a cluster group as recorded at the level of each donor differs substantially between old versus young donors, the t test as implemented in the base R (v4.1.1) *t.test* function was used. When comparing cluster group frequency distributions between 3 or more conditions, the significance of the differences in distributions observed was quantified by using the t test in a pairwise fashion and the p-value of the test was corrected for multiple comparisons using the Benjamini & Hochberg correction as implemented in the base R *p.adjust*. Comparison of the frequencies of CFU-Mk and CFU-GM colonies in young and aged BM was analysed using the *chisq.test* function in base R.

Gene set enrichment analysis. Gene set enrichment analysis (GSEA) was conducted starting from the output of Seurat's *FindMarkers* function and ranking the full list of gene average fold changes per HSC cluster group or clone obtained by comparing the normalised UMI/read counts from different cluster groups or clones by their magnitude and sign. The enrichment of the gene lists of interest, including the majority of murine gene lists whose gene symbols have been converted into human ones for the computation of AUCell scores (*8, 28, 54-58*), a selection of a number of relevant human dataset derived gene sets previously published (*10, 27, 29, 58, 60, 61*) and the MSigDB Hallmark gene signatures (*62*) (Data S4) was evaluated in the ranked gene lists generated following the differential gene expression analysis by using the *fgsea* wrapper function of the fgsea R package (v1.20.0) (*63*) with the following parameters: minSize=5, maxSize=1500, nPermSimple = 100000. Assessing if the overlap between gene signatures was statistically significant in terms of its magnitude was evaluated using the *newGOM* function of the GeneOverlap R package (v1.28.0) (*64*) that implements the Fisher's exact test to perform the hypothesis testing.

CFU assays. Thawing media was prepared with IMDM buffer (Cat# 21056023, Gibco) supplemented with 20% HyClone Defined Fetal Bovine Serum (HyClone-FBS; GE Healthcare) and 11% DNAse I (110 mg/mL). Cryopreserved bone marrow samples were thawed at 37°C in a water bath. 1 mL of warm HyClone-FBS was added to the samples and the suspension was then diluted by dropwise addition of 8 mL thawing media. Cell suspensions were then centrifuged at 350 g for 10 min at room temperature. Next, cells were resuspended in FACS buffer (IMDM with 10% HyClone-FBS and 10 mg/mL DNAse I), filtered, and placed on ice. Cells were stained for flow cytometry with human CD3, CD4, CD8, CD11b, CD14, CD19, CD56, CD235a, CD34, CD38, CD90, CD45RA, CD10, CD123, CD127, and CD41 for 30–40 min on ice. After washing with 1 mL of FACS buffer, cells were centrifuged at 350 g for 5 min at 8°C and resuspended in

500 mL FACS buffer containing Hoechst at 1 in 10^4 dilution for dead cell exclusion. Live Lin-CD34+ cells were sorted using a FACSAria Fusion (BD Biosciences) machine.

To test for megakaryocytic potential, live Lin^-CD34^+ cells were sorted (10 cells /well) into round-bottom 96-well plates containing 100µL of filtered StemSpan (STEMCELL Technologies) medium supplemented with hSCF (20 ng/mL), hTPO (50 ng/mL), and 1% penicillin-streptomycin. Cells were cultured at 37°C and 5% CO_2. Half of the medium was carefully removed and replaced with fresh medium on culture days 3, 5, 8, and 10. At day 14, plates were inspected under the microscope to identify wells with at least 10 cells. Cells from these wells were harvested and stained with antibodies (Supplementary Table S2) against human CD41 and CD42, and FACS analysis was performed using the BD LSRFortessa or BD LSR II (BD Biosciences). Hoechst was used as viability dye, at a final concentration of 1 in 10^4. FlowJo software was used for subsequent data analysis. Wells containing 5 or more cells within the $CD41^+CD42^+$ gate were defined positive for the megakaryocytic output.

To test for myeloid potential, live single Lin^-CD34^+ cells were sorted into Terasaki plates containing 20µL of Iscove's Modified Dulbecco's Medium (IMDM) with L-glutamine (Gibco), 20% HyClone Defined Fetal Bovine Serum (GE Healthcare), 1% penicillin-streptomycin (Invitrogen), and 0.1µM β-mercaptoethanol (Sigma-Aldrich), supplemented with hSCF (20 ng/mL), hFlt3L (20 ng/mL), hIL-3 (20 ng/mL), hIL-6 (20 ng/mL), hGM-CSF (50 ng/mL), and hG-CSF (20 ng/mL). Cells were cultured at 37°C and 5% CO_2 for 3 weeks. Cells from these wells were harvested and stained with antibodies (Supplementary Table S2) against human CD14, CD15, and CD117. Hoechst was used as viability dye, at a final concentration of 1 in 10^4. FACS was performed using BD LSRFortessa and BD LSR II cytometers. FlowJo analysis software was used for subsequent data analysis. Wells containing 5 or more combined CD14+ or CD15+ live cells were defined positive for the myeloid output.

Statistical analysis. The significance of the differences observed while comparing large variable distributions, where variables were not normally distributed, was quantified using the Wilcoxon rank-sum test in a pairwise fashion and the p-value of the test was corrected for multiple comparisons using the Benjamini & Hochberg correction. On the occasion when we used the t test, it was used on comparing the difference in means between small variable distributions and the variables were normally distributed. The strength and direction of association between two variables was assessed using either the Pearson's R, when the variables were normally distributed, or Kendall's τ correlation coefficients, when the variables were not normally distributed &/ contained ties in their rankings. The χ^2 test was utilised to evaluate if the frequency of CFU-Mk and CFU-GM colonies was dependent on the age group of the bone marrow donors, as this test allowed to evaluate this dependency and its strength. Gene sets overlap magnitude was assessed using Fisher's exact test, which is a test employed in widely used over representation analysis tools.

Supplementary materials

Figure S1-S10
Table S1-S5
Supplementary data files S1-S7

References and Notes.

1. S. E. W. Jacobsen, C. Nerlov, Haematopoiesis in the era of advanced single-cell technologies. *Nat Cell Biol* **21**, 2-8 (2019).
2. B. Dykstra, D. Kent, M. Bowie, L. McCaffrey, M. Hamilton, K. Lyons, S. J. Lee, R. Brinkman, C. Eaves, Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell stem cell* **1**, 218-229 (2007).
3. R. Yamamoto, Y. Morita, J. Ooehara, S. Hamanaka, M. Onodera, K. L. Rudolph, H. Ema, H. Nakauchi, Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112-1126 (2013).
4. A. Sanjuan-Pla, I. C. Macaulay, C. T. Jensen, P. S. Woll, T. C. Luis, A. Mead, S. Moore, C. Carella, S. Matsuoka, T. Bouriez Jones, O. Chowdhury, L. Stenson, M. Lutteropp, J. C. Green, R. Facchini, H. Boukarabila, A. Grover, A. Gambardella, S. Thongjuea, J. Carrelha, P. Tarrant, D. Atkinson, S. A. Clark, C. Nerlov, S. E. Jacobsen, Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232-236 (2013).
5. J. Carrelha, Y. Meng, L. M. Kettyle, T. C. Luis, R. Norfo, V. Alcolea, H. Boukarabila, F. Grasso, A. Gambardella, A. Grover, K. Hogstrand, A. M. Lord, A. Sanjuan-Pla, P. S. Woll, C. Nerlov, S. E. W. Jacobsen, Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106-111 (2018).
6. W. Pei, T. B. Feyerabend, J. Rossler, X. Wang, D. Postrach, K. Busch, I. Rode, K. Klapproth, N. Dietlein, C. Quedenau, W. Chen, S. Sauer, S. Wolf, T. Hofer, H. R. Rodewald, Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456-460 (2017).
7. A. E. Rodriguez-Fraticelli, S. L. Wolock, C. S. Weinreb, R. Panero, S. H. Patel, M. Jankovic, J. Sun, R. A. Calogero, A. M. Klein, F. D. Camargo, Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216 (2018).
8. A. E. Rodriguez-Fraticelli, C. Weinreb, S. W. Wang, R. P. Migueles, M. Jankovic, M. Usart, A. M. Klein, S. Lowell, F. D. Camargo, Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* **583**, 585-589 (2020).
9. A. Grover, A. Sanjuan-Pla, S. Thongjuea, J. Carrelha, A. Giustacchini, A. Gambardella, I. Macaulay, E. Mancini, T. C. Luis, A. Mead, S. E. Jacobsen, C. Nerlov, Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun* **7**, 11075 (2016).
10. W. W. Pang, E. A. Price, D. Sahoo, I. Beerman, W. J. Maloney, D. J. Rossi, S. L. Schrier, I. L. Weissman, Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A* **108**, 20012-20017 (2011).
11. S. Valletta, A. Thomas, Y. Meng, X. Ren, R. Drissen, H. Sengul, C. Di Genua, C. Nerlov, Micro-environmental sensing by bone marrow stroma identifies IL-6 and TGFbeta1 as regulators of hematopoietic ageing. *Nat Commun* **11**, 4075 (2020).
12. T. N. Rao, N. Hansen, J. Stetka, D. Luque Paz, M. Kalmer, J. Hilfiker, M. Endele, N. Ahmed, L. Kubovcakova, M. Rybarikova, H. Hao-Shen, F. Geier, C. Beisel, S. Dirnhofer, T. Schroeder, T. H. Brummendorf, D. Wolf, S. Koschmieder, R. C. Skoda, JAK2-V617F and interferon-alpha induce megakaryocyte-biased stem cells characterized by decreased long-term functionality. *Blood* **137**, 2139-2151 (2021).
13. J. Tong, T. Sun, S. Ma, Y. Zhao, M. Ju, Y. Gao, P. Zhu, P. Tan, R. Fu, A. Zhang, D. Wang, D. Wang, Z. Xiao, J. Zhou, R. Yang, S. J. Loughran, J. Li, A. R. Green, E. H.

Bresnick, D. Wang, T. Cheng, L. Zhang, L. Shi, Hematopoietic Stem Cell Heterogeneity Is Linked to the Initiation and Therapeutic Response of Myeloproliferative Neoplasms. *Cell Stem Cell* **28**, 502-513 e506 (2021).

14. F. Notta, S. Doulatov, E. Laurenti, A. Poeppl, I. Jurisica, J. E. Dick, Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218-221 (2011).
15. E. Laurenti, C. Frelin, S. Xie, R. Ferrari, C. F. Dunant, S. Zandi, A. Neumann, I. Plumb, S. Doulatov, J. Chen, C. April, J. B. Fan, N. Iscove, J. E. Dick, CDK6 levels regulate quiescence exit in human hematopoietic stem cells. *Cell Stem Cell* **16**, 302-313 (2015).
16. A. M. Cheung, L. V. Nguyen, A. Carles, P. Beer, P. H. Miller, D. J. Knapp, K. Dhillon, M. Hirst, C. J. Eaves, Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. *Blood* **122**, 3129-3137 (2013).
17. M. E. Belderbos, S. Jacobs, T. K. Koster, A. Ausema, E. Weersing, E. Zwart, G. de Haan, L. V. Bystrykh, Donor-to-Donor Heterogeneity in the Clonal Dynamics of Transplanted HumanCord Blood Stem Cellsin Murine Xenografts. *Biol Blood Marrow Transplant* **26**, 16-25 (2020).
18. L. Biasco, D. Pellin, S. Scala, F. Dionisio, L. Basso-Ricci, L. Leonardelli, S. Scaramuzza, C. Baricordi, F. Ferrua, M. P. Cicalese, S. Giannelli, V. Neduva, D. J. Dow, M. Schmidt, C. Von Kalle, M. G. Roncarolo, F. Ciceri, P. Vicard, E. Wit, C. Di Serio, L. Naldini, A. Aiuti, In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell* **19**, 107-119 (2016).
19. E. Six, A. Guilloux, A. Denis, A. Lecoules, A. Magnani, R. Vilette, F. Male, N. Cagnard, M. Delville, E. Magrin, L. Caccavelli, C. Roudaut, C. Plantier, S. Sobrino, J. Gregg, C. L. Nobles, J. K. Everett, S. Hacein-Bey-Abina, A. Galy, A. Fischer, A. J. Thrasher, I. Andre, M. Cavazzana, F. D. Bushman, Clonal tracking in gene therapy patients reveals a diversity of human hematopoietic differentiation programs. *Blood* **135**, 1219-1231 (2020).
20. A. Rundberg Nilsson, S. Soneji, S. Adolfsson, D. Bryder, C. J. Pronk, Human and Murine Hematopoietic Stem Cell Aging Is Associated with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias. *PLoS One* **11**, e0158369 (2016).
21. B. Lehnertz, J. Chagraoui, T. MacRae, E. Tomellini, S. Corneau, N. Mayotte, I. Boivin, A. Durand, D. Gracias, G. Sauvageau, HLF expression defines the human hematopoietic stem cell state. *Blood* **138**, 2642-2654 (2021).
22. R. A. Voit, L. Tao, F. Yu, L. D. Cato, B. Cohen, T. J. Fleming, M. Antoszewski, X. Liao, C. Fiorini, S. K. Nandakumar, L. Wahlster, K. Teichert, A. Regev, V. G. Sankaran, A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia. *Nat Immunol* **24**, 69-83 (2023).
23. M. E. Belderbos, T. Koster, B. Ausema, S. Jacobs, S. Sowdagar, E. Zwart, E. de Bont, G. de Haan, L. V. Bystrykh, Clonal selection and asymmetric distribution of human leukemia in murine xenografts revealed by cellular barcoding. *Blood* **129**, 3210-3220 (2017).
24. B. E. McIntosh, M. E. Brown, B. M. Duffin, J. P. Maufort, D. T. Vereide, Slukvin, II, J. A. Thomson, Nonirradiated NOD,B6.SCID Il2rgamma-/- Kit(W41/W41) (NBSGW) mice support multilineage engraftment of human hematopoietic cells. *Stem Cell Reports* **4**, 171-180 (2015).
25. J. L. McKenzie, O. I. Gan, M. Doedens, J. C. Wang, J. E. Dick, Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment. *Nat Immunol* **7**, 1225-1233 (2006).

26. C. Weinreb, A. Rodriguez-Fraticelli, F. D. Camargo, A. M. Klein, Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, (2020).
27. S. Belluschi, E. F. Calderbank, V. Ciaurro, B. Pijuan-Sala, A. Santoro, N. Mende, E. Diamanti, K. Y. C. Sham, X. Wang, W. W. Y. Lau, W. Jawaid, B. Gottgens, E. Laurenti, Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat Commun* **9**, 4100 (2018).
28. N. Cabezas-Wallscheid, F. Buettner, P. Sommerkamp, D. Klimmeck, L. Ladel, F. B. Thalheimer, D. Pastor-Flores, L. P. Roma, S. Renders, P. Zeisberger, A. Przybylla, K. Schonberger, R. Scognamiglio, S. Altamura, C. M. Florian, M. Fawaz, D. Vonficht, M. Tesio, P. Collier, D. Pavlinic, H. Geiger, T. Schroeder, V. Benes, T. P. Dick, M. A. Rieger, O. Stegle, A. Trumpp, Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell* **169**, 807-823 e819 (2017).
29. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. K. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jané-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).
30. K. B. Kaufmann, A. G. X. Zeng, E. Coyaud, L. Garcia-Prat, E. Papalexi, A. Murison, E. M. N. Laurent, M. Chan-Seng-Yue, O. I. Gan, K. Pan, J. McLeod, H. Boutzen, S. Zandi, S. I. Takayanagi, R. Satija, B. Raught, S. Z. Xie, J. E. Dick, A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness. *Nat Immunol* **22**, 723-734 (2021).
31. S. Pinho, T. Marchand, E. Yang, Q. Wei, C. Nerlov, P. S. Frenette, Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. *Dev Cell* **44**, 634-641 e634 (2018).
32. M. Zhao, J. M. Perry, H. Marshall, A. Venkatraman, P. Qian, X. C. He, J. Ahamed, L. Li, Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. *Nat Med* **20**, 1321-1326 (2014).
33. Y. Meng, J. Carrelha, R. Drissen, X. Ren, B. Zhang, A. Gambardella, S. Valletta, S. Thongjuea, S. E. Jacobsen, C. Nerlov, Epigenetic programming defines haematopoietic stem cell fate restriction. *Nat Cell Biol* **25**, 812-822 (2023).
34. J. J. Li, J. Liu, Y. E. Li, L. V. Chen, H. Cheng, Y. Li, T. Cheng, Q. F. Wang, B. O. Zhou, Differentiation route determines the functional outputs of adult megakaryopoiesis. *Immunity* **57**, 478-494 e476 (2024).
35. D. M. Poscablo, A. K. Worthington, S. Smith-Berdan, M. G. E. Rommel, B. A. Manso, R. Adili, L. Mok, R. E. Reggiardo, T. Cool, R. Mogharrab, J. Myers, S. Dahmen, P. Medina, A. E. Beaudin, S. W. Boyer, M. Holinstat, V. D. Jonsson, E. C. Forsberg, An age-progressive platelet differentiation path from hematopoietic stem cells causes exacerbated thrombosis. *Cell* **187**, 3090-3107 e3021 (2024).
36. C. Di Genua, S. Valletta, M. Buono, B. Stoilova, C. Sweeney, A. Rodriguez-Meira, A. Grover, R. Drissen, Y. Meng, R. Beveridge, Z. Aboukhalil, D. Karamitros, M. E. Belderbos, L. Bystrykh, S. Thongjuea, P. Vyas, C. Nerlov, C/EBPalpha and GATA-2 Mutations Induce Bilineage Acute Erythroid Leukemia through Transformation of a Neomorphic Neutrophil-Erythroid Progenitor. *Cancer Cell* **37**, 690-704 e698 (2020).

37. S. Picelli, O. R. Faridani, A. K. Bjorklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).
38. A. Rodriguez-Meira, G. Buck, S. A. Clark, B. J. Povinelli, V. Alcolea, E. Louka, S. McGowan, A. Hamblin, N. Sousos, N. Barkas, A. Giustacchini, B. Psaila, S. E. W. Jacobsen, S. Thongjuea, A. J. Mead, Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol Cell* **73**, 1292-1305 e1298 (2019).
39. A. Giustacchini, S. Thongjuea, N. Barkas, P. S. Woll, B. J. Povinelli, C. A. G. Booth, P. Sopp, R. Norfo, A. Rodriguez-Meira, N. Ashley, L. Jamieson, P. Vyas, K. Anderson, A. Segerstolpe, H. Qian, U. Olsson-Stromberg, S. Mustjoki, R. Sandberg, S. E. W. Jacobsen, A. J. Mead, Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* **23**, 692-702 (2017).
40. F. Krueger. (2012).
41. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
42. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).
43. P. Ewels, M. Magnusson, S. Lundin, M. Kaller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016).
44. N. Novershtern, A. Subramanian, L. N. Lawton, R. H. Mak, W. N. Haining, M. E. McConkey, N. Habib, N. Yosef, C. Y. Chang, T. Shay, G. M. Frampton, A. C. Drake, I. Leskov, B. Nilsson, F. Preffer, D. Dombkowski, J. W. Evans, T. Liefeld, J. S. Smutko, J. Chen, N. Friedman, R. A. Young, T. R. Golub, A. Regev, B. L. Ebert, Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309 (2011).
45. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021).
46. R. B. Patterson-Cross, A. J. Levine, V. Menon, Selecting single cell clustering parameter values using subsampling-based robustness metrics. *BMC Bioinformatics* **22**, 1-13 (2021).
47. L. Jardine, S. Webb, I. Goh, M. Quiroga Londono, G. Reynolds, M. Mather, B. Olabi, E. Stephenson, R. A. Botting, D. Horsfall, J. Engelbert, D. Maunder, N. Mende, C. Murnane, E. Dann, J. McGrath, H. King, I. Kucinski, R. Queen, C. D. Carey, C. Shrubsole, E. Poyner, M. Acres, C. Jones, T. Ness, R. Coulthard, N. Elliott, S. O'Byrne, M. L. R. Haltalli, J. E. Lawrence, S. Lisgo, P. Balogh, K. B. Meyer, E. Prigmore, K. Ambridge, M. S. Jain, M. Efremova, K. Pickard, T. Creasey, J. Bacardit, D. Henderson, J. Coxhead, A. Filby, R. Hussain, D. Dixon, D. McDonald, D. M. Popescu, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Regev, S. Behjati, E. Laurenti, N. K. Wilson, A. Roy, B. Gottgens, I. Roberts, S. A. Teichmann, M. Haniffa, Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* **598**, 327-331 (2021).
48. M. Andreatta, A. J. Berenstein, S. J. Carmona, scGate: marker-based purification of cell types from heterogeneous single-cell RNA-seq datasets. *Bioinformatics* **38**, 2642-2644 (2022).

49. L. Scrucca, M. Fop, T. B. Murphy, A. E. Raftery, mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R j* **8**, 289-317 (2016).
50. R. A. Amezquita, A. T. L. Lun, E. Becht, V. J. Carey, L. N. Carpp, L. Geistlinger, F. Marini, K. Rue-Albrecht, D. Risso, C. Soneson, L. Waldron, H. Pages, M. L. Smith, W. Huber, M. Morgan, R. Gottardo, S. C. Hicks, Orchestrating single-cell analysis with Bioconductor. *Nat Methods* **17**, 137-145 (2020).
51. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278 (2015).
52. K. D. Zimmerman, M. A. Espeland, C. D. Langefeld, A practical solution to pseudoreplication bias in single-cell studies. *Nat Commun* **12**, 738 (2021).
53. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**, 1083-1086 (2017).
54. A. Giladi, F. Paul, Y. Herzog, Y. Lubling, A. Weiner, I. Yofe, D. Jaitin, N. Cabezas-Wallscheid, R. Dress, F. Ginhoux, A. Trumpp, A. Tanay, I. Amit, Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nature Cell Biology* **20**, 836-846 (2018).
55. F. K. B. Lauridsen, T. L. Jensen, N. Rapin, D. Aslan, A. S. Wilhelmson, S. Pundhir, M. Rehn, F. Paul, A. Giladi, M. S. Hasemann, P. Serup, I. Amit, B. T. Porse, Differences in Cell Cycle Status Underlie Transcriptional Heterogeneity in the HSC Compartment. *Cell Reports* **24**, 766-780 (2018).
56. N. K. Wilson, D. G. Kent, F. Buettner, M. Shehata, I. C. Macaulay, F. J. Calero-Nieto, M. Sánchez Castillo, C. A. Oedekoven, E. Diamanti, R. Schulte, C. P. Ponting, T. Voet, C. Caldas, J. Stingl, A. R. Green, F. J. Theis, B. Göttgens, Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16**, 712-724 (2015).
57. E. M. Pietras, D. Reynaud, Y. A. Kang, D. Carlin, F. J. Calero-Nieto, A. D. Leavitt, J. A. Stuart, B. Göttgens, E. Passegué, Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35-46 (2015).
58. E. Montecino-Rodriguez, Y. Kong, D. Casero, A. Rouault, K. Dorshkind, P. D. Pioli, Lymphoid-Biased Hematopoietic Stem Cells Are Maintained with Age and Efficiently Generate Lymphoid Progeny. *Stem Cell Reports* **12**, 584-596 (2019).
59. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nature Protocols* **4**, 1184-1191 (2009).
60. F. Anjos-Afonso, F. Buettner, S. A. Mian, H. Rhys, J. Perez-Lloret, M. Garcia-Albornoz, N. Rastogi, L. Ariza-McNaughton, D. Bonnet, Single cell analyses identify a highly regenerative and homogenous human CD34+ hematopoietic stem cell population. *Nat Commun* **13**, 2048 (2022).
61. S. Z. Xie, L. Garcia-Prat, V. Voisin, R. Ferrari, O. I. Gan, E. Wagenblast, K. B. Kaufmann, A. G. X. Zeng, S. I. Takayanagi, I. Patel, E. K. Lee, J. Jargstorf, G. Holmes, G. Romm, K. Pan, M. Shoong, A. Vedi, C. Luberto, M. D. Minden, G. D. Bader, E. Laurenti, J. E. Dick, Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal. *Cell Stem Cell* **25**, 639-653 e637 (2019).

62. A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
63. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. Artyomov, A. Sergushichev, Fast gene set enrichment analysis. *bioRxiv*, (2021).
64. S. I. Shen Li. (Bioconductor version: Release (3.14), 2021).

ACKNOWLEDGEMENTS.
We thank the Oxford University Biomedical Services for help with animal studies, WIMM FACS facility for assistance with cell sorting, and the WIMM CBRG for computational support.

Funding:
This work was supported by a TUBITAK scholarship to M.A., MRC Unit Grants to C.N. (MC_UU_12009/7) and P.V. (MC_UU_12009/11), an MRC Discovery Award to P.V and C.N., the NIHR/Oxford Biomedical Research Centre Haematology Theme and Blood Cancer UK Specialist Program Grant 13001 to P.V. The WIMM FACS Core Facility is supported by the MRC HIU, MRC MHU (MC_UU_12009), NIHR Oxford BRC and the John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170) and by WIMM Strategic Alliance awards (G0902418 and MC_UU_12025).

Author contributions:
M.A., B.S., M.B., Y.M., M.M., Y.Z., S.T. and S.A.C. performed the experiments. M.A., G.G., S.T. and C.N. analyzed the data, R.B. generated and validated the barcoded lentiviral library and barcode detection, N.A.J. collected bone marrow from aged individuals, C.N. and P.V. conceived the experiments, and C.N., P.V., M.A. and G.G. wrote the manuscript.

Competing interests:
The authors do not have any competing interests.

Data and materials availability:
All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. Request for materials and code should be directed to C.N. (claus.nerlov@imm.ox.ac.uk).
Sequencing data generated has been deposited in GEO (accession number GSE188889). Underlying tabulated data for all figures can be found in Supplementary Data S7.

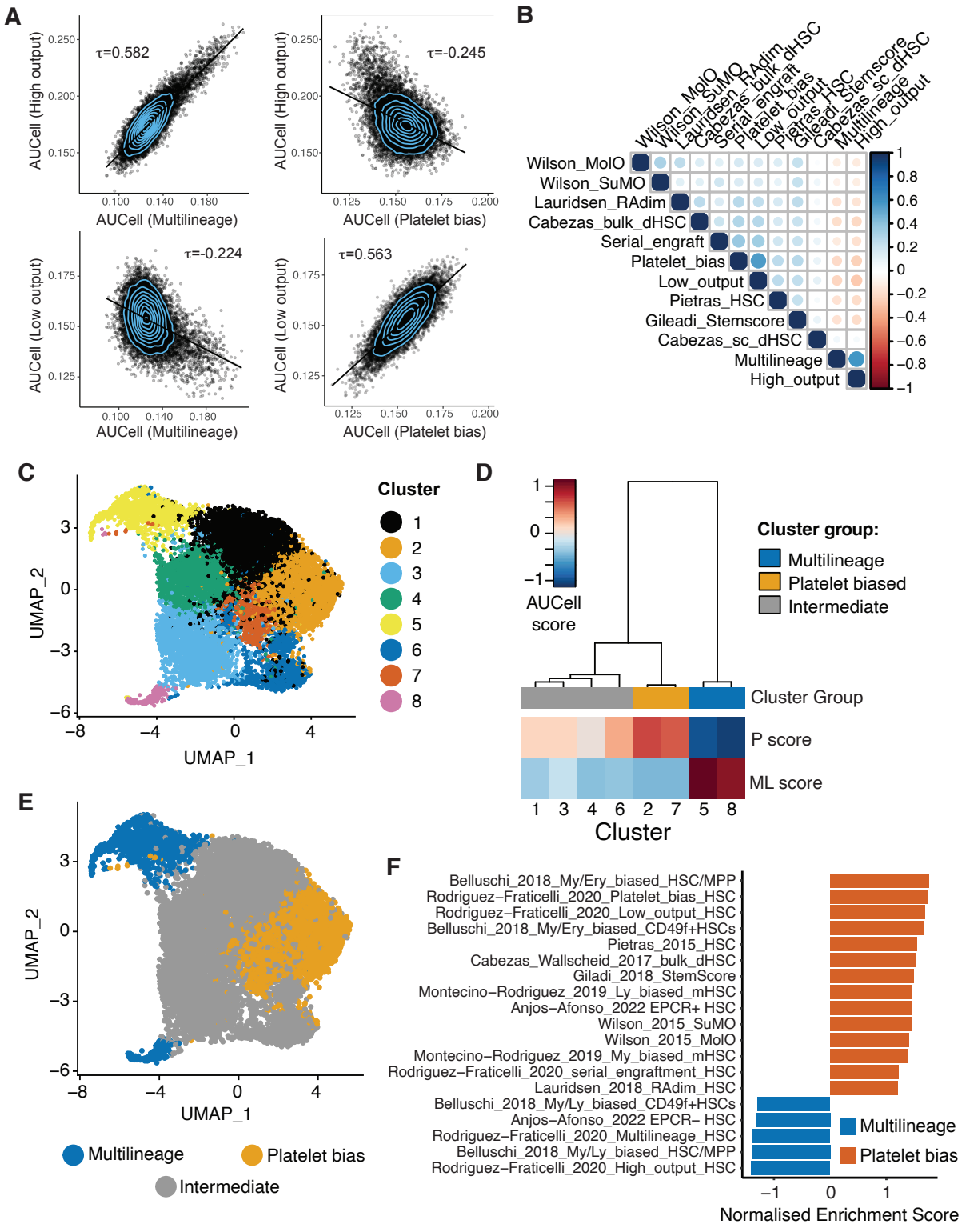
A
τ=0.582
τ=-0.245
τ=-0.224
τ=0.563
AUCell (High output)
AUCell (Low output)
AUCell (Multilineage)
AUCell (Platelet bias)
B
Wilson_MolO
Wilson_SuMO
Lauridsen_RAdim
Cabezas_bulk_dHSC
Serial_engraft
Platelet_bias
Low_output
Pietras_HSC
Gileadi_Stemscore
Cabezas_sc_dHSC
Multilineage
High_output
C
Cluster
UMAP_1
UMAP_2
D
AUCell score
Cluster group:
Multilineage
Platelet biased
Intermediate
Cluster Group
P score
ML score
Cluster
E
Multilineage
Platelet bias
Intermediate
F
Belluschi_2018_My/Ery_biased_HSC/MPP
Rodriguez−Fraticelli_2020_Platelet_bias_HSC
Rodriguez−Fraticelli_2020_Low_output_HSC
Belluschi_2018_My/Ery_biased_CD49f+HSCs
Pietras_2015_HSC
Cabezas_Wallscheid_2017_bulk_dHSC
Giladi_2018_StemScore
Montecino−Rodriguez_2019_Ly_biased_mHSC
Anjos−Afonso_2022 EPCR+ HSC
Wilson_2015_SuMO
Wilson_2015_MolO
Montecino−Rodriguez_2019_My_biased_mHSC
Rodriguez−Fraticelli_2020_serial_engraftment_HSC
Lauridsen_2018_RAdim_HSC
Belluschi_2018_My/Ly_biased_CD49f+HSCs
Anjos−Afonso_2022 EPCR− HSC
Rodriguez−Fraticelli_2020_Multilineage_HSC
Belluschi_2018_My/Ly_biased_HSC/MPP
Rodriguez−Fraticelli_2020_High_output_HSC
Multilineage
Platelet bias
Normalised Enrichment Score

Figure 1

Fig. 1. Computational identification of platelet biased HSCs in adult human bone marrow.
A) Correlation between AUCell scores of the indicated transcriptional signatures defining HSC lineage bias and cellular output in transcriptomes of single human HSCs from adult bone marrow (N=13633, 3 independent donors). For each comparison the Kendal τ-b correlation coefficient and the associated P-value are shown.
B) Kendal correlation coefficients of the indicated gene signatures from published datasets with in the single cell transcriptomes from (A). The scale shows the correlation coefficient.
C) UMAP representation and unsupervised clustering of the single cell transcriptomes from (A).
D) Clusters from (C) were hierarchically clustered of using median AUCell scores of platelet-biased and multi-lineage signatures (Supplementary Fig. S2B), generating 3 distinct cluster groups.
E) Localization of the lineage-bias associated cluster groups in the UMAP representation from (C).
F) GSEA normalized enrichment from the comparison of the platelet-bias and multilineage cluster groups from (D) using the gene signatures from (B) as well as other gene sets associated with HSC lineage bias and self-renewal (see Data S4). Only gene signatures achieving a normalized enrichment score with FDR < 0.25 are shown.
FDR: False discovery rate.

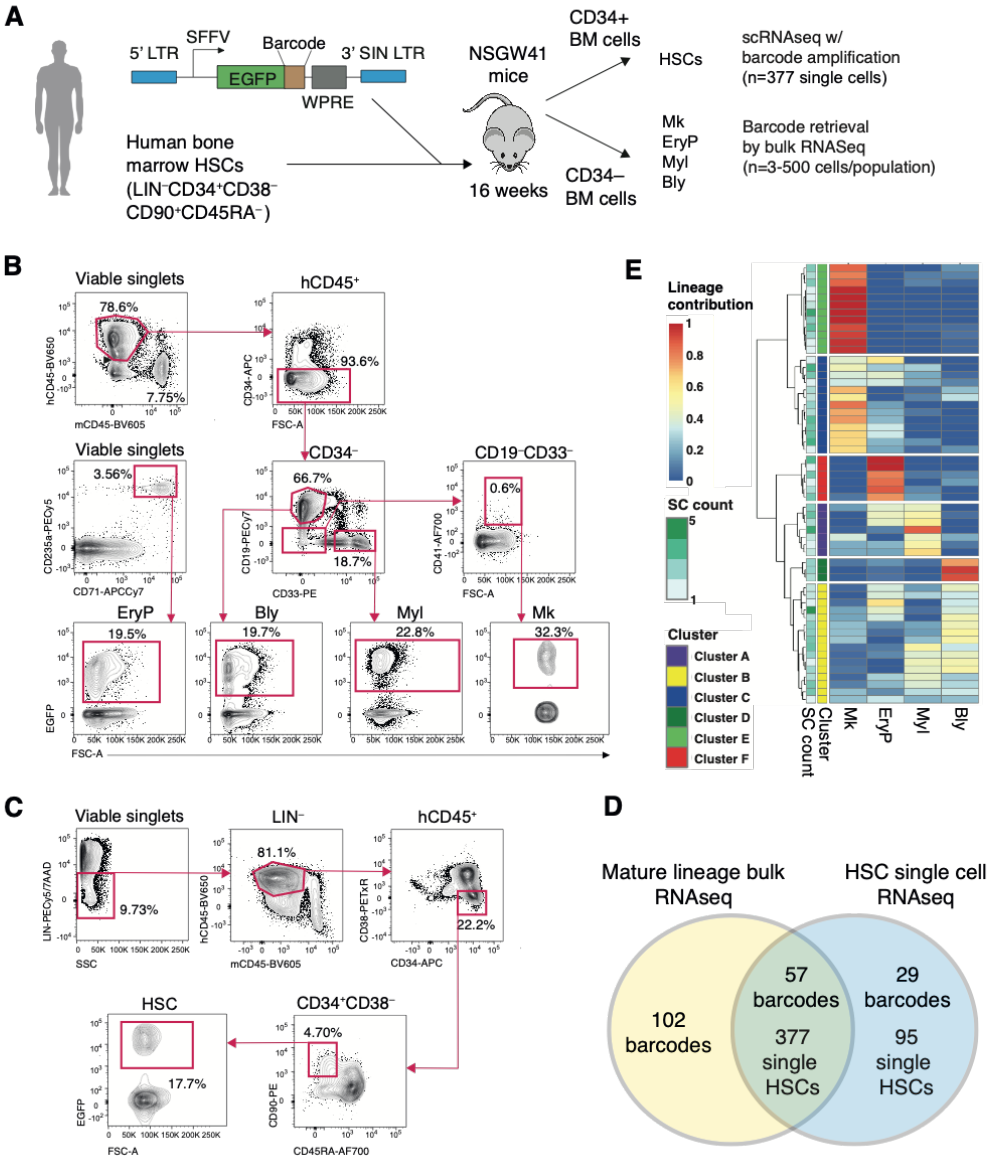
A
5' LTR
SFFV
Barcode
3' SIN LTR
EGFP
WPRE
Human bone marrow HSCs (LIN⁻CD34⁺CD38⁻ CD90⁺CD45RA⁻)
NSGW41 mice
16 weeks
CD34+ BM cells
CD34– BM cells
HSCs
Mk
EryP
Myl
Bly
scRNAseq w/ barcode amplification (n=377 single cells)
Barcode retrieval by bulk RNASeq (n=3-500 cells/population)
B
Viable singlets
78.6%
7.75%
hCD45⁺
93.6%
CD34⁻
66.7%
18.7%
CD19⁻CD33⁻
0.6%
3.56%
EryP
19.5%
Bly
19.7%
Myl
22.8%
Mk
32.3%
C
Viable singlets
9.73%
LIN⁻
81.1%
hCD45⁺
22.2%
HSC
17.7%
CD34⁺CD38⁻
4.70%
D
Mature lineage bulk RNAseq
HSC single cell RNAseq
102 barcodes
57 barcodes
377 single HSCs
29 barcodes
95 single HSCs
E
Lineage contribution
SC count
Cluster
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E
Cluster F
Mk
EryP
Myl
Bly

Figure 2

Fig. 2. Lentiviral barcoding of adult human bone marrow HSCs.

A) Experimental workflow and structure of the barcode lentiviral construct. 5' LTR: 5' long terminal repeat; SFFV: spleen focus forming virus; EGFP: enhanced green fluorescent protein; WPRE**,** Woodchuck hepatitis virus regulatory element; 3' SIN LTR: 3' self-inactivating LTR. The barcode is situated downstream of the EGFP and is expressed as proviral mRNA.

B) Representative flow cytometry plots showing the gating strategy used to purify the cellular output of barcoded human HSCs from murine bone marrow. EryP: erythroid progenitor; Bly, B-lymphocytes; Myl: myeloid cells; Mk: megakaryocytes. Percentages shown are the proportion of the gated population.

C) Representative flow cytometry plots showing the gating strategy used to purify barcoded human HSCs from murine bone marrow

D) The overlap between cellular barcodes retrieved from bulk RNA sequencing of the lineage output isolated as in (B) and from single cell RNAseq of human HSCs retrieves as in (C).

E) Heatmap showing hierarchical clustering of barcoded hHSC clones based on their lineage output. Each row represents an individual barcoded hHSC clone. SC count: shows the number of hHSCs within each barcoded clone Each column shows the contribution of the hHSC clone to different lineages (megakaryocyte, Mk; erythroid, EryP; myeloid, Myl; and B lymphocytes, Bly). The heat signature shows the normalized lineage contribution of each hHSC clone as fraction of total output. Based on lineage outputs, the hHSC clones fall into six clusters A-F, which are colour-coded as shown.

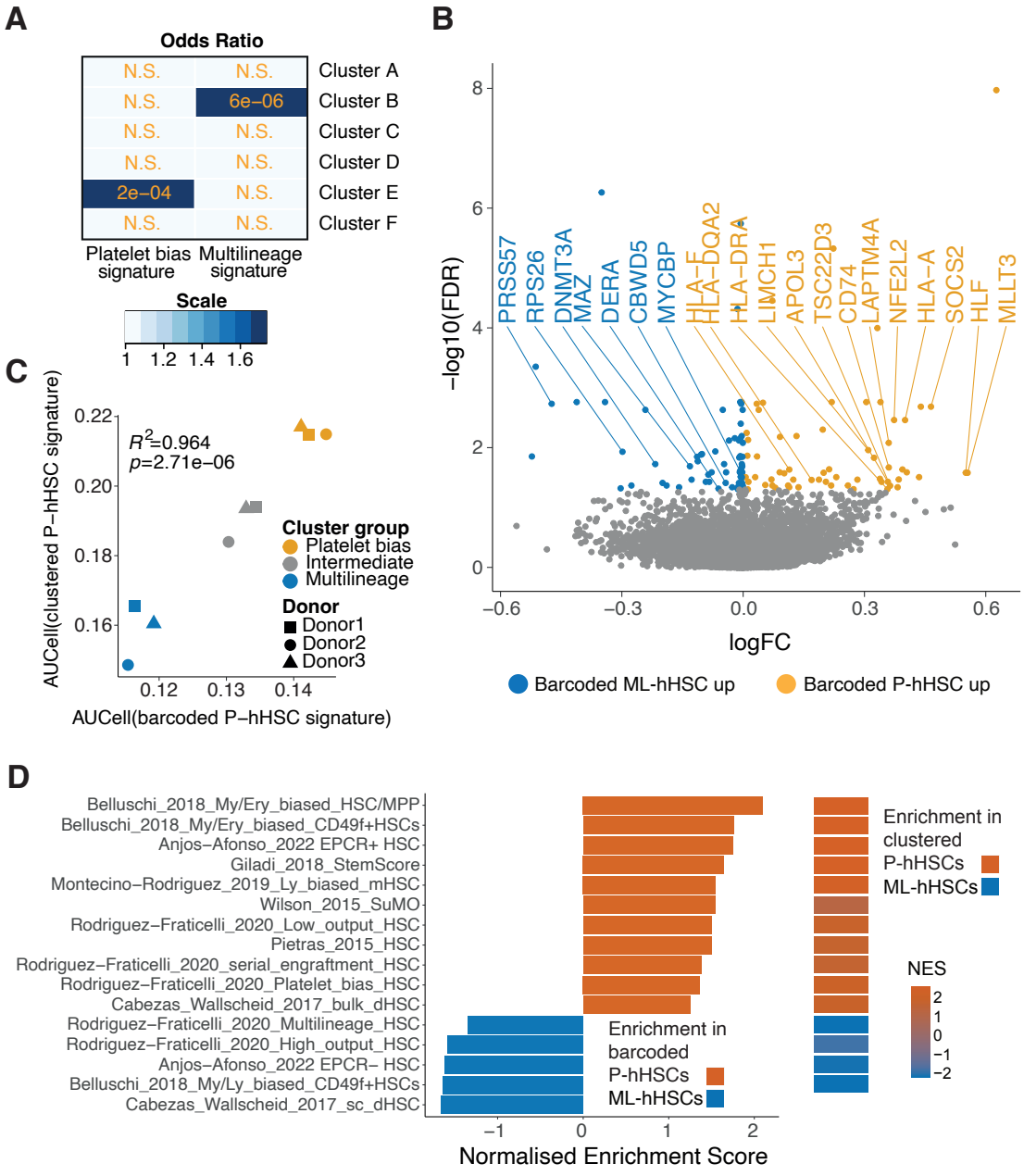
A
Odds Ratio
N.S.
6e−06
2e−04
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E
Cluster F
Platelet bias signature
Multilineage signature
Scale
1 1.2 1.4 1.6
B
−log10(FDR)
logFC
PRSS57
RPS26
DNMT3A
MAZ
DERA
CBWD5
MYCBP
HLA−E
HLA−DQA2
HLA−DRA
LIMCH1
APOL3
TSC22D3
CD74
LAPTM4A
NFE2L2
HLA−A
SOCS2
HLF
MLLT3
Barcoded ML-hHSC up
Barcoded P-hHSC up
C
AUCell(clustered P−hHSC signature)
AUCell(barcoded P−hHSC signature)
R2=0.964
p=2.71e−06
Cluster group
Platelet bias
Intermediate
Multilineage
Donor
Donor1
Donor2
Donor3
D
Belluschi_2018_My/Ery_biased_HSC/MPP
Belluschi_2018_My/Ery_biased_CD49f+HSCs
Anjos−Afonso_2022 EPCR+ HSC
Giladi_2018_StemScore
Montecino−Rodriguez_2019_Ly_biased_mHSC
Wilson_2015_SuMO
Rodriguez−Fraticelli_2020_Low_output_HSC
Pietras_2015_HSC
Rodriguez−Fraticelli_2020_serial_engraftment_HSC
Rodriguez−Fraticelli_2020_Platelet_bias_HSC
Cabezas_Wallscheid_2017_bulk_dHSC
Rodriguez−Fraticelli_2020_Multilineage_HSC
Rodriguez−Fraticelli_2020_High_output_HSC
Anjos−Afonso_2022 EPCR− HSC
Belluschi_2018_My/Ly_biased_CD49f+HSCs
Cabezas_Wallscheid_2017_sc_dHSC
Enrichment in barcoded P-hHSCs
ML-hHSCs
Normalised Enrichment Score
Enrichment in clustered P-hHSCs
ML-hHSCs
NES

Figure 3

Fig. 3. Comparison of platelet-biased HSCs identified by barcoding and computational clustering.

A) Heatmap of the odds ratio quantifying the magnitude of the overlap between upregulated cluster marker genes detected in the barcoded hHSC dataset and the P-hHSC and ML-hHSCS signatures defined clustering in the young donor hHSC dataset. Odds ratio scale and p-value of the Fisher's exact test used to quantify the magnitude of the overlap are shown if the p-value is < 0.05.

B) Volcano plot showing genes differentially expressed between platelet biased and multilineage hHSC clones identified by barcoding (FDR<0.05). Gene IDs for genes also present in the P-hHSC and ML-hHSC signatures identified by clustering in Fig. 1 are shown.

C) Correlation between the AUCell scores of the P-hHSC signatures defined by barcoding (x-axis) and clustering in the young donor HSC dataset (y-axis). AUCell scores have been aggregated as medians at the level of both cluster group and donor. The Pearson's R correlation coefficient and associated P-value are shown.

D) Comparison of P-hHSCs and ML-hHSCs defined by barcoding using GSEA of stemness and proliferation signatures from Fig. 1F. NESs are shown on the x-axis. Only gene signatures achieving an FDR<0.25 are included. The NESs for the same analysis using the P-hHSC and ML-hHSC populations defined by clustering in Fig. 1F is shown on the right for comparison, with NES values given by the scale bar.

NES: Normalised enrichment score. FDR: False discovery rate.

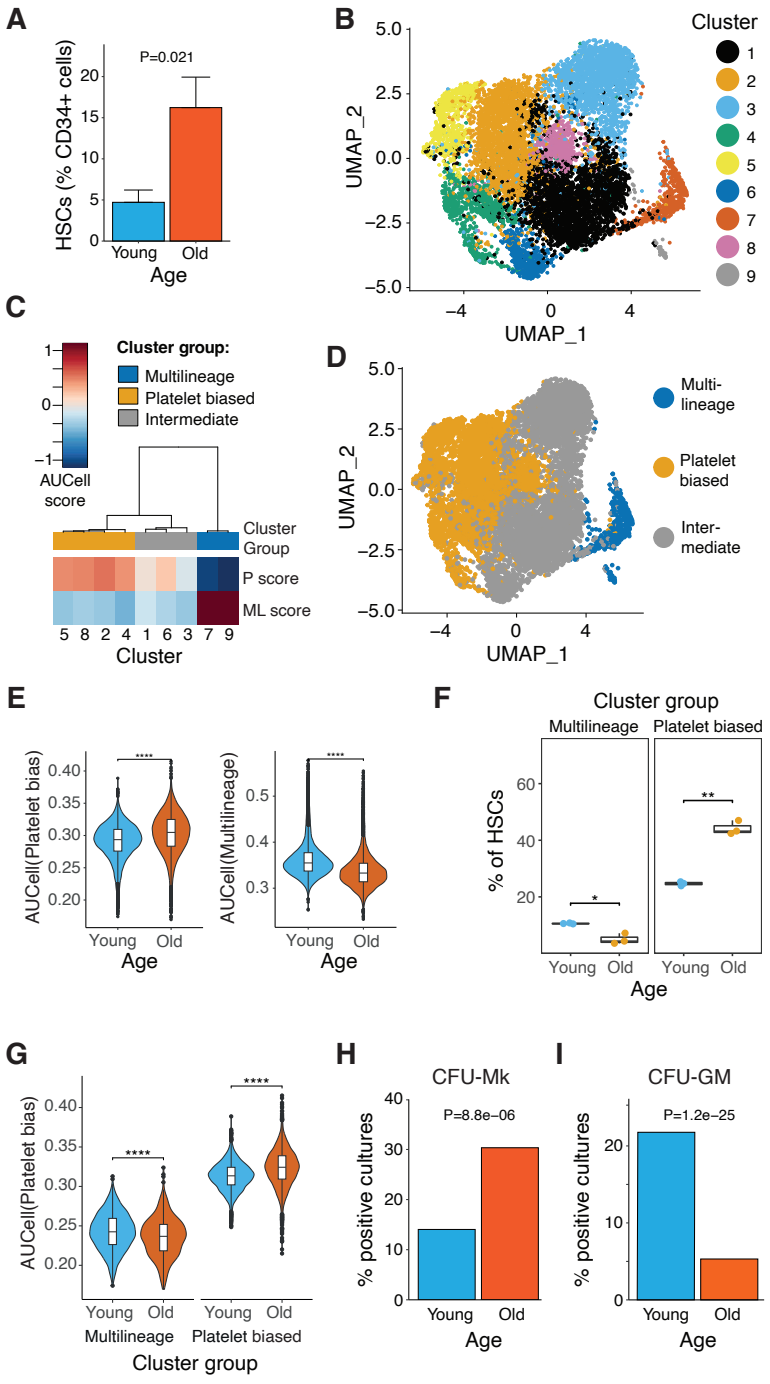
A
HSCs (% CD34+ cells)
P=0.021
Young
Old
Age
B
UMAP_2
UMAP_1
Cluster
1
2
3
4
5
6
7
8
9
C
Cluster group:
Multilineage
Platelet biased
Intermediate
AUCell score
Cluster Group
P score
ML score
5 8 2 4 1 6 3 7 9
Cluster
D
Multi-lineage
Platelet biased
Inter-mediate
UMAP_2
UMAP_1
E
AUCell(Platelet bias)
AUCell(Multilineage)
Young
Old
Age
F
Cluster group
Multilineage
Platelet biased
% of HSCs
Young
Old
Age
G
AUCell(Platelet bias)
Young
Old
Multilineage
Platelet biased
Cluster group
H
CFU-Mk
P=8.8e−06
% positive cultures
Young
Old
Age
I
CFU-GM
P=1.2e−25
% positive cultures
Young
Old
Age

Figure 4

Fig. 4. In silico identified platelet biased HSCs in human adult bone marrow exhibit an enrichment in old healthy donors.

A) Frequency of immunophenotypic hHSCs (% of CD34+ bone marrow mononuclear cells) in the young (N=3) and old (N=3) donors used for generating the single cell transcriptomic datasets. P-value is shown (Student's t-test).

B) UMAP representation and unsupervised clustering of the single cell transcriptomes from old donor hHSCs (N = 9981, 3 independent donors)

C) Grouping of clusters from (B) by hierarchical clustering of using median AUCell scores following unbiased clustering in cluster groups according to patterns of gene expression summarised by AUCell scores. Refer to Supplementary Fig. S6C for the per cluster distribution of the AUCell scores computed from the platelet bias and multilineage HSC gene signatures.

D) Localization of the platelet biased and multilineage cluster groups in the UMAP representation from (B).

E) The distribution of the AUCell scores generated using the age-independent P- and ML- hHSC signatures at the level of the young and old donor hHSC datasets. ****: P≤0.0001 (Wilcoxon rank-sum test).

F) The frequency of the P- and ML- hHSC reported as percentage of the HSC compartment of each donor at the level of the young and old donor hHSC datasets. *: P≤0.05, **:P≤0.01 (Student t test)

G) The distribution of the AUCell scores generated using the age-independent P- and ML- hHSC signatures at the level of the P- and ML- hHSC identified in the young and old donor hHSC datasets. ****: P≤0.0001 (Wilcoxon rank-sum test).

H) The frequency of cultures positive for megakaryocytes using young and old CD34+ BM cells (10 cells/culture). P-value is shown (Chi-square test)

I) The frequency of cultures positive for myeloid cells using young and old CD34+ BM cells (10 cells/culture). P-value is shown (Chi-square test)

Supplementary Materials.

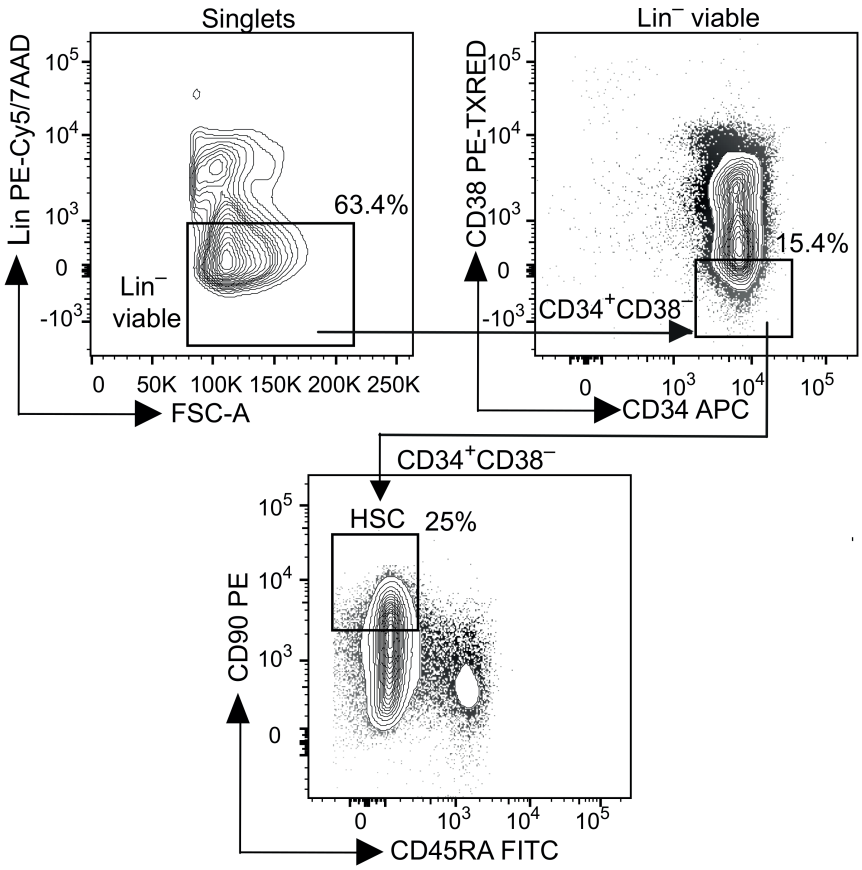
Singlets
Lin PE-Cy5/7AAD
63.4%
Lin⁻ viable
FSC-A
Lin⁻ viable
CD38 PE-TXRED
15.4%
CD34⁺CD38⁻
CD34 APC
CD34⁺CD38⁻
HSC 25%
CD90 PE
CD45RA FITC

Figure S1

Fig. S1. Gating strategy for bone marrow hHSC purification.
Human bone marrow hHSCs were FACS purified using the sorting strategy outlined. The gated populations as percentage of the parental gate are shown.

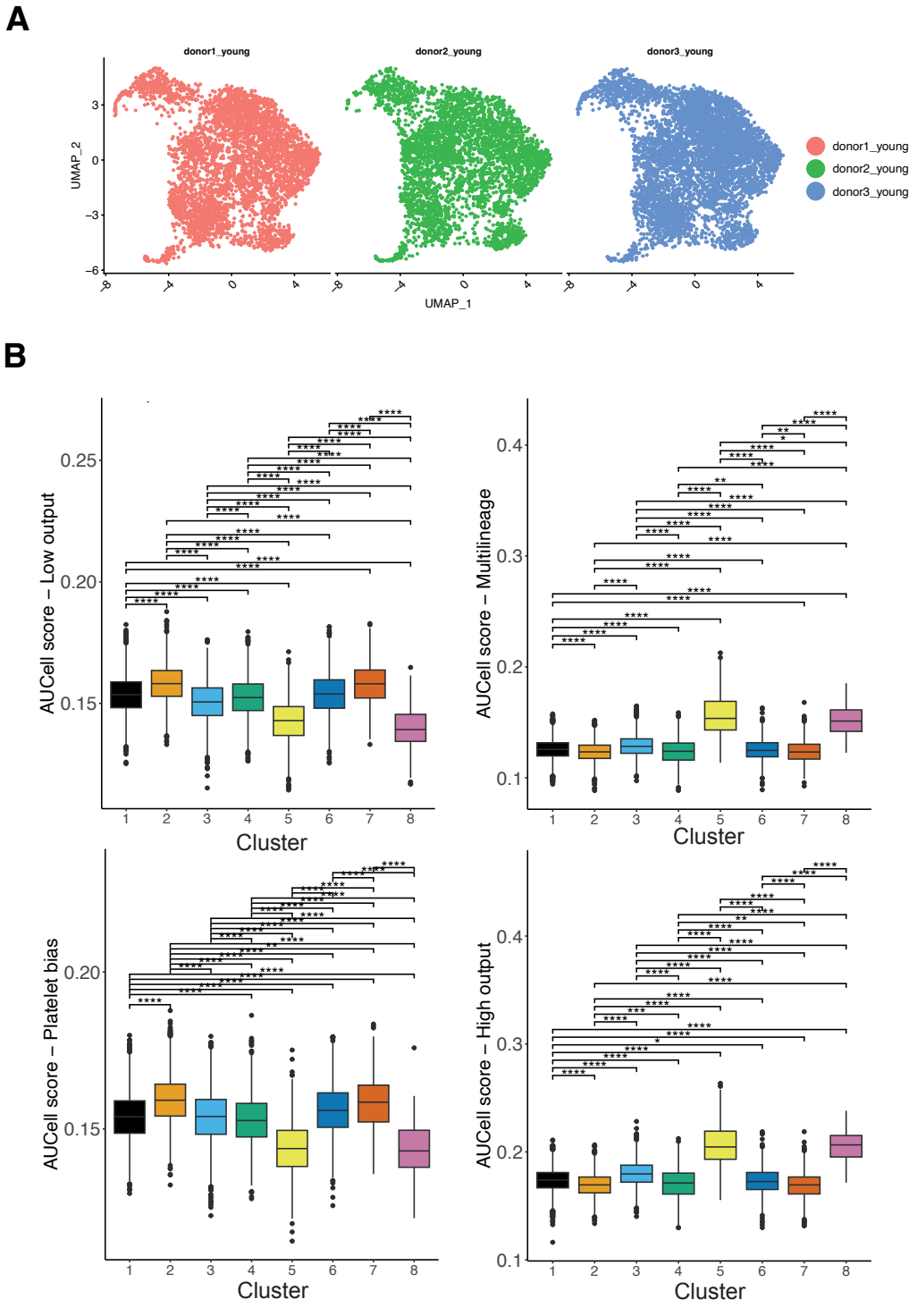
A
donor1_young
donor2_young
donor3_young
UMAP_2
UMAP_1
donor1_young
donor2_young
donor3_young
B
AUCell score – Low output
AUCell score – Multilineage
AUCell score – Platelet bias
AUCell score – High output
Cluster

Figure S2

Fig. S2. Clustering of human HSCs.
A) UMAP projection of the data from each individual young hHSC donor
B) AUCell score distribution for the platelet bias, multilineage output as well as high and low output HSC signatures postulated in (*8*), after murine to human gene symbol conversion, in the hHSC clusters defined in Fig. 1C. Pairwise comparison using the Wilcoxon rank-sum test – multiple testing adjusted p-values represented: **** adjusted P <= 0.0001, *** adjusted P <= 0.001, ** adjusted P < 0.01, * adjusted P < 0.05.

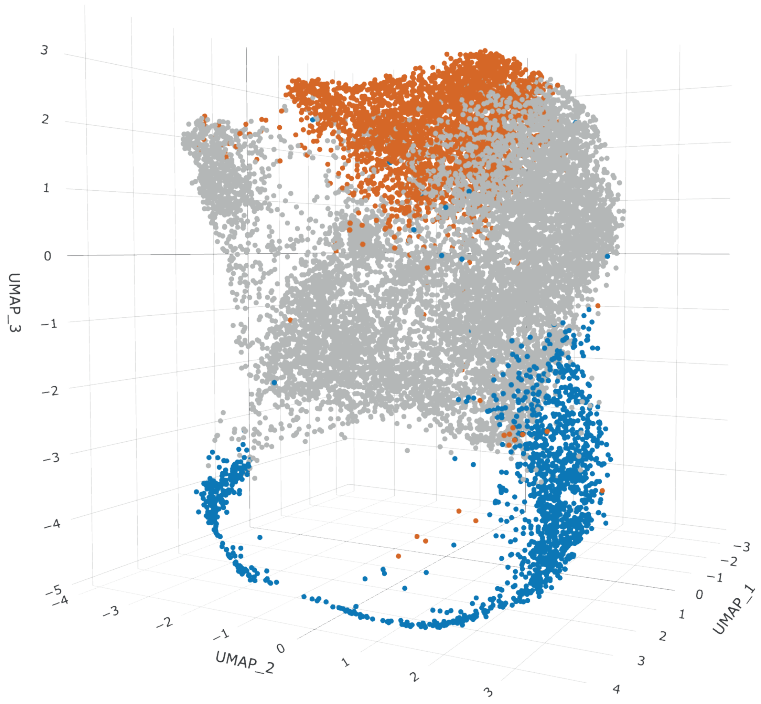

Figure S3

Fig. S3. 3D UMAP projection of hHSC cluster groups. The 3D UMAP projection shows P-hHSCs (orange), ML-hHSC (blue) and intermediate hHSCs (grey). Note the continuity of the ML-hHSC population in the 3D projection.

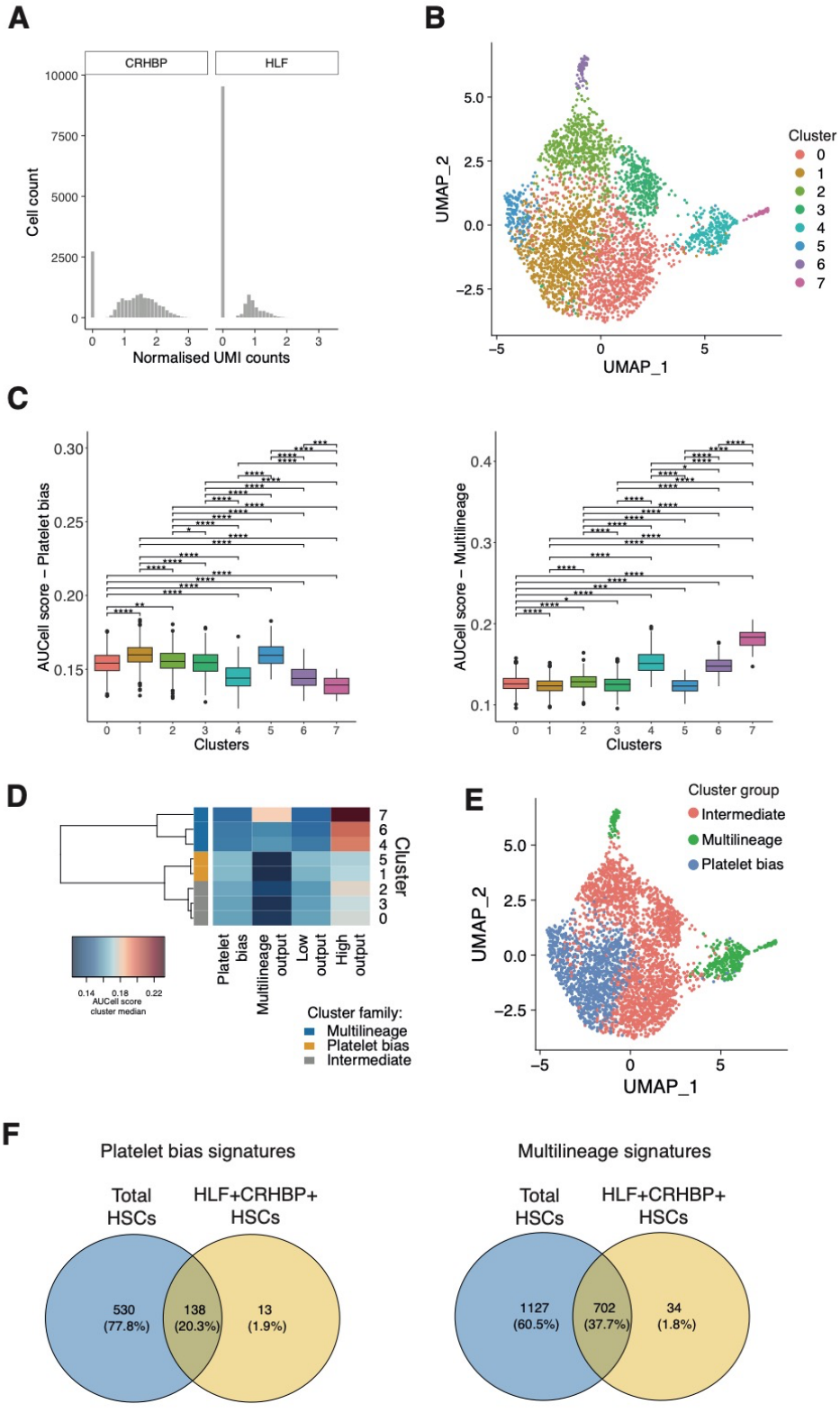
A
CRHBP
HLF
Cell count
Normalised UMI counts
B
UMAP_2
UMAP_1
Cluster
C
AUCell score – Platelet bias
AUCell score – Multilineage
Clusters
D
Cluster
Platelet bias
Multilineage output
Low output
High output
AUCell score cluster median
Cluster family:
Multilineage
Platelet bias
Intermediate
E
Cluster group
Intermediate
Multilineage
Platelet bias
F
Platelet bias signatures
Total HSCs
HLF+CRHBP+ HSCs
530 (77.8%)
138 (20.3%)
13 (1.9%)
Multilineage signatures
1127 (60.5%)
702 (37.7%)
34 (1.8%)

Figure S4

Fig. S4. Analysis of young CRHBP+HLF+ hHSCs.

A) UMI counts of CHRBP (left) and HLF (right) in 10X Chromium single cell RNAseq of young hHSCs. 3500 of 13633 cells were found to express both genes.

B) UMAP representation and unsupervised clustering of the CRHBP+HLF+ single cell transcriptomes from (A).

C) AUCell score distribution for the platelet bias and multilineage output HSC signatures postulated in (*8*) , after murine to human gene symbol conversion, in the hHSC clusters defined in (B). Pairwise comparison using the Wilcoxon rank-sum test – multiple testing adjusted p-values represented: **** adjusted $P <= 0.0001$, *** adjusted $P <= 0.001$, ** adjusted $P < 0.01$, * adjusted $P < 0.05$.

D) Clusters from (B) were hierarchically clustered of using median AUCell scores of platelet-biased and multi-lineage signatures, as well as high and low output HSC signatures, generating 3 distinct cluster groups.

E) Localization of the lineage-bias associated cluster groups in the UMAP representation from (B).

F) Venn diagrams showing the overlap between the human platelet (left) and multi-lineage gene expression signatures (right) generated by comparison of P- and ML-HSC cluster groups generated using the total young HSC population, or the young CHRBP+HLF+ HSC subset.

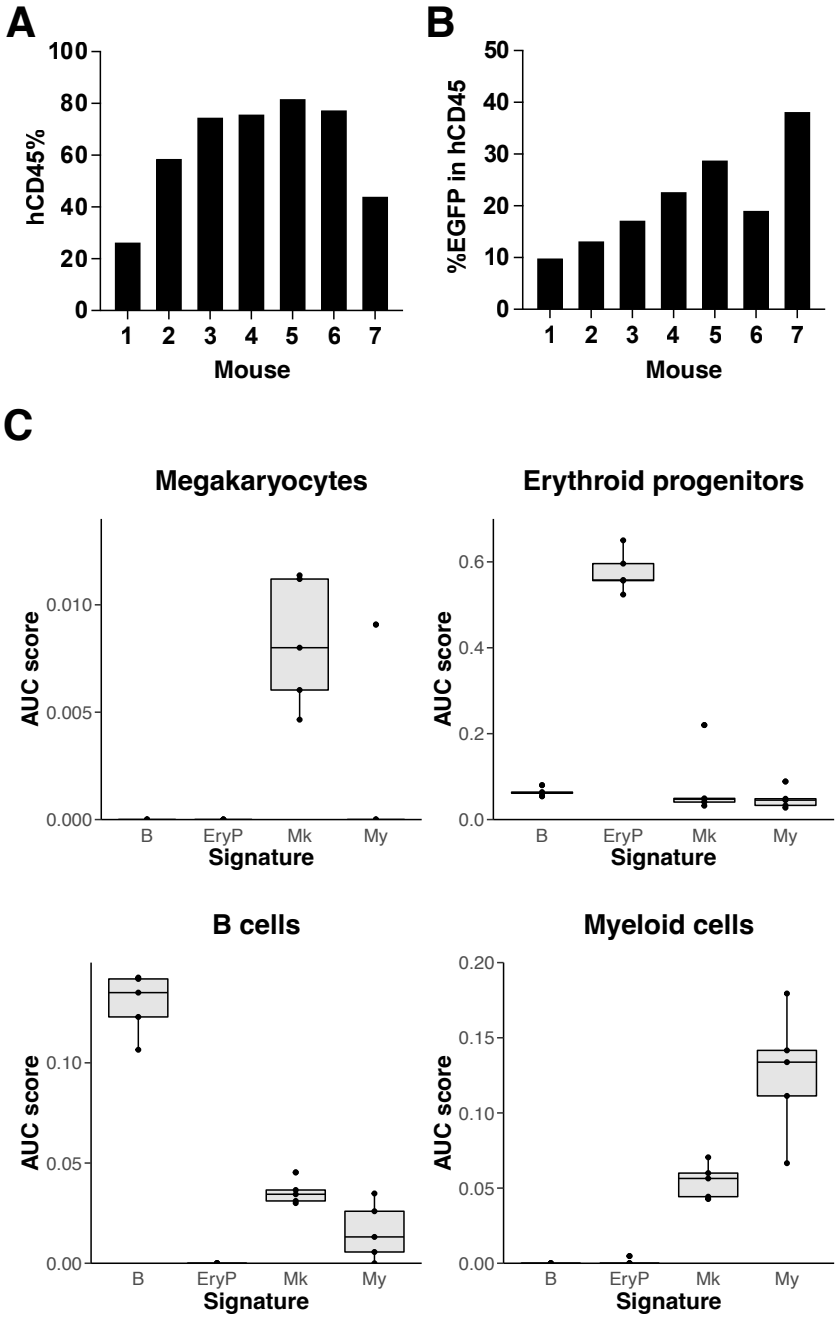
A
hCD45%
Mouse
B
%EGFP in hCD45
Mouse
C
Megakaryocytes
Erythroid progenitors
B cells
Myeloid cells
AUC score
Signature
B
EryP
Mk
My

Figure S5

Fig. S5. Human HSC engraftment and mature cell type identity.
A) Human chimerism within the murine bone marrow leukocyte population from mice used for barcode analysis. The bars show the percentage of hCD45+ cells within the viable singlet gate as defined in Fig. 2B.
B) Bar graph showing the percentage of EGFP+ cells within the hCD45+ population, as defined in (A).
C) Box and whisker plot showing AUCell scores for lineage-specific signatures in bulk RNAseq of the indicated mature cell types. N=5 per cell type.

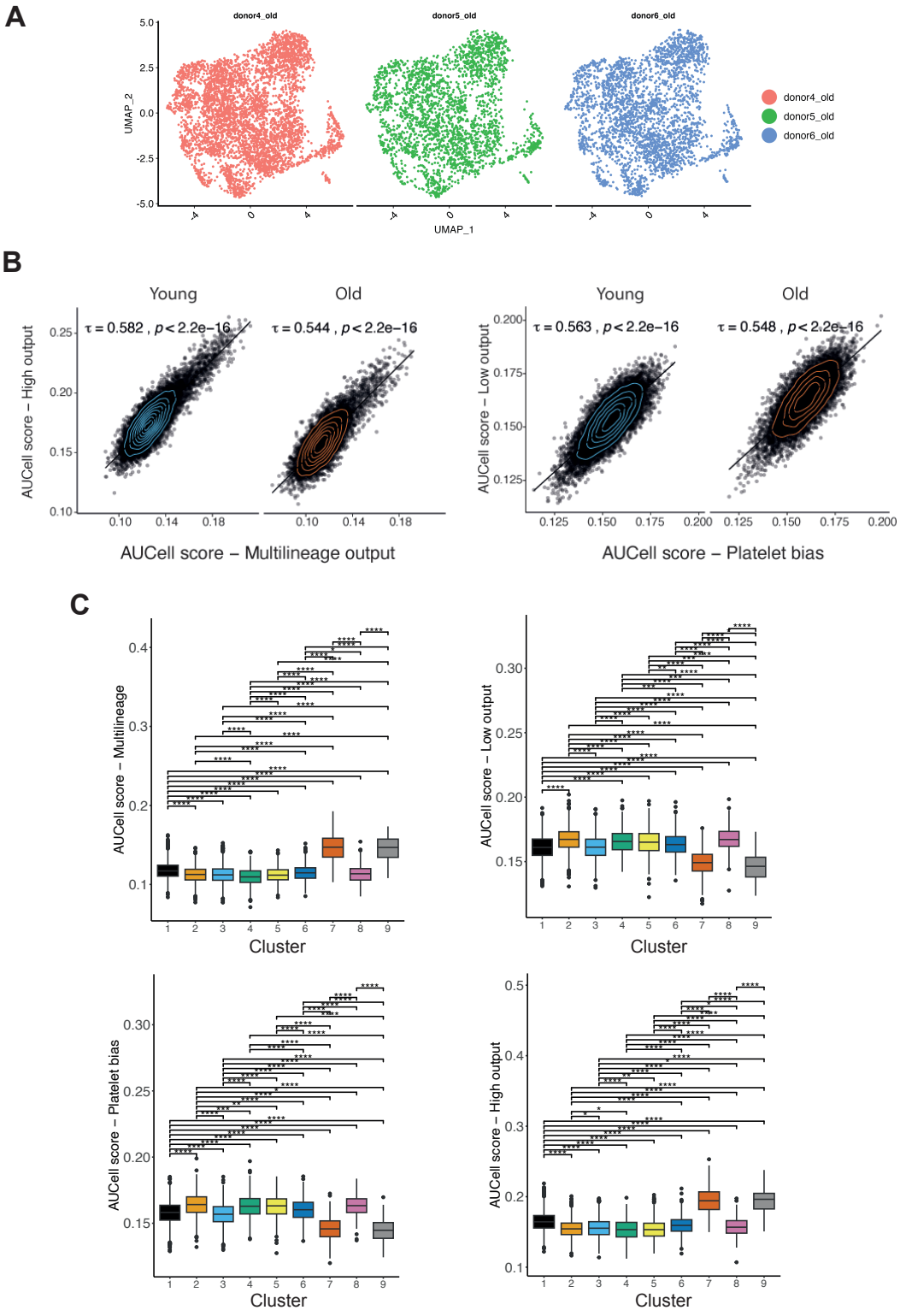
A
donor4_old
donor5_old
donor6_old
UMAP_2
UMAP_1
B
Young
Old
τ = 0.582 , p < 2.2e−16
τ = 0.544 , p < 2.2e−16
AUCell score – High output
AUCell score – Multilineage output
τ = 0.563 , p < 2.2e−16
τ = 0.548 , p < 2.2e−16
AUCell score – Low output
AUCell score – Platelet bias
C
AUCell score – Multilineage
AUCell score – Low output
AUCell score – Platelet bias
AUCell score – High output
Cluster

Figure S6

Fig. S6. Analysis of old hHSCs.

A) UMAP projection of old hHSCS from the 3 individual donors.

B) Correlation between AUCell scores of the indicated transcriptional signatures defining HSC lineage bias and cellular output in transcriptomes of single human HSCs from young (left panel) and old (right panel) adult bone marrow (N=13884, N=10301, respectively, 3 independent donors/condition). For each comparison the Kendal τ-b correlation coefficient and the associated P-value is shown.

C) AUCell score distribution for the indicated gene signatures in the hHSC clusters performed as in Fig. 4B. Pairwise comparison Wilcoxon rank-sum test – multiple testing adjusted p-values represented: **** adjusted P≤0.0001, *** adjusted P≤0.001, ** adjusted P≤0.01, * adjusted P≤0.05


A
Young P-HSC signature
Old P-HSC signature
299
369
512
Young ML-HSC signature
Old ML-HSC signature
716
1113
416
B
Belluschi_2018_My/Ery_biased_HSC/MPP
Rodriguez−Fraticelli_2020_Platelet_bias_HSC
Belluschi_2018_My/Ery_biased_CD49f+HSCs
Lauridsen_2018_RAdim_HSC
Montecino−Rodriguez_2019_Ly_biased_mHSC
Montecino−Rodriguez_2019_My_biased_mHSC
Anjos−Afonso_2022 EPCR+ HSC
Rodriguez−Fraticelli_2020_Low_output_HSC
Giladi_2018_StemScore
Wilson_2015_MolO
Cabezas_Wallscheid_2017_bulk_dHSC
Cabezas_Wallscheid_2017_sc_dHSC
Belluschi_2018_My/Ly_biased_CD49f+HSCs
Rodriguez−Fraticelli_2020_Multilineage_HSC
Rodriguez−Fraticelli_2020_High_output_HSC
Belluschi_2018_My/Ly_biased_HSC/MPP
Enriched in:
Old ML−hHSC
Old P−hHSC
−1
0
1
2
Normalised Enrichment Score
C
TNFA_SIGNALING_VIA_NFKB
ALLOGRAFT_REJECTION
KRAS_SIGNALING_UP
IL6_JAK_STAT3_SIGNALING
ESTROGEN_RESPONSE_EARLY
COAGULATION
Tirosh_2016_G2M_phase
Tirosh_2016_S_phase
MTORC1_SIGNALING
DNA_REPAIR
OXIDATIVE_PHOSPHORYLATION
MYC_TARGETS_V1
GLYCOLYSIS
G2M_CHECKPOINT
E2F_TARGETS
Enriched in:
Old ML−hHSC
Old P−hHSC
−1
0
1
2
Normalised Enrichment Score
TNFA_SIGNALING_VIA_NFKB
INTERFERON_GAMMA_RESPONSE
INTERFERON_ALPHA_RESPONSE
IL6_JAK_STAT3_SIGNALING
ESTROGEN_RESPONSE_EARLY
COAGULATION
ALLOGRAFT_REJECTION
INFLAMMATORY_RESPONSE
KRAS_SIGNALING_DN
KRAS_SIGNALING_UP
NOTCH_SIGNALING
HEDGEHOG_SIGNALING
MYC_TARGETS_V2
GLYCOLYSIS
OXIDATIVE_PHOSPHORYLATION
PI3K_AKT_MTOR_SIGNALING
DNA_REPAIR
MTORC1_SIGNALING
Tirosh_2016_S_phase
MYC_TARGETS_V1
G2M_CHECKPOINT
E2F_TARGETS
Enriched in:
Young ML−hHSC
Young P−hHSC
−1
0
1
Normalised Enrichment Score

Figure S7

Fig. S7. Comparison of young and old lineage-biased hHSCs by GSEA.
A) Venn diagrams illustrating the size of the overlap between the P-hHSC gene signatures determined in the young and the old donor hHSC datasets as well as the overlaps between the ML-hHSC signatures determined in the young and the old donor hHSC datasets.
B) GSEA of the differential gene expression analysis results between P-hHSCs versus ML-HSCs from the old donor hHSC dataset focusing on the expression patterns of the gene sets explored in the analysis illustrated in Fig 1F. Gene signatures achieving a normalized enrichment score with an FDR < 0.25 are shown.
C) Analysis as in (B) with the addition of exploring the same hHSC type comparison in the young hHSC dataset and using the MSigDB Hallmark pathway gene set.

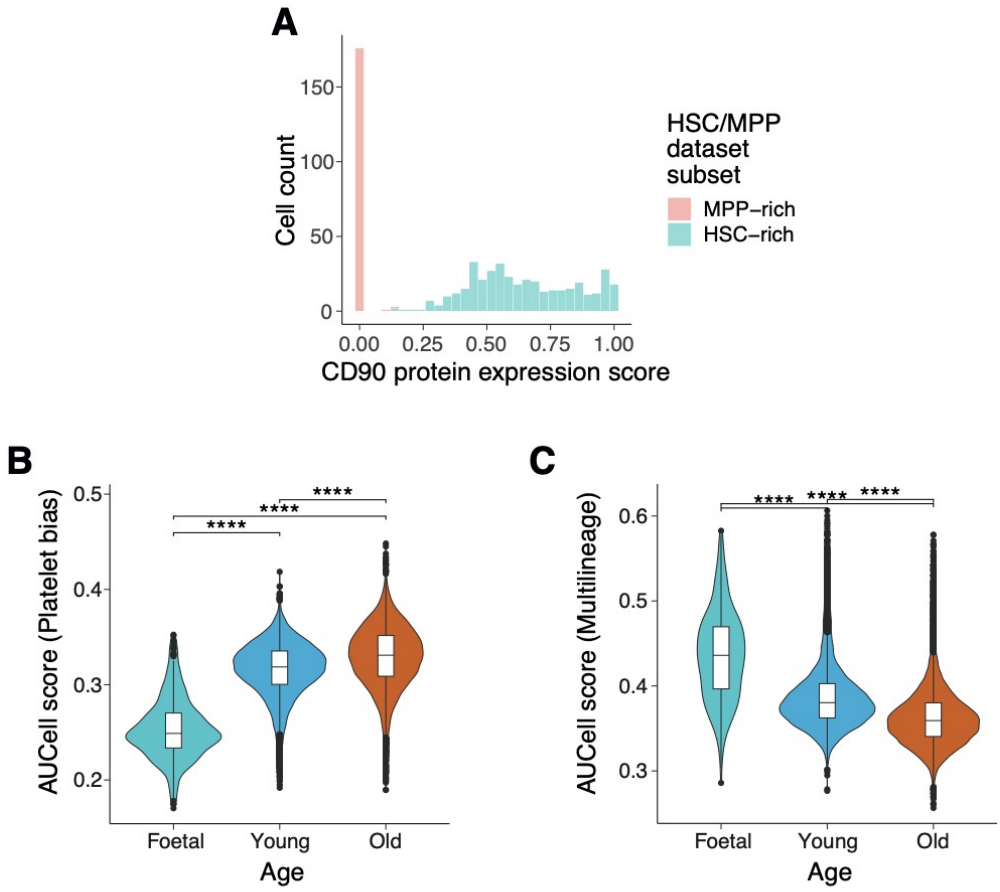
A
Cell count
150
100
50
0
0.00
0.25
0.50
0.75
1.00
CD90 protein expression score
HSC/MPP dataset subset
MPP−rich
HSC−rich
B
AUCell score (Platelet bias)
0.5
0.4
0.3
0.2
****
****
****
Foetal
Young
Old
Age
C
AUCell score (Multilineage)
0.6
0.5
0.4
0.3
****
****
****
Foetal
Young
Old
Age

Figure S8

Fig. S8. Analysis of fetal hHSCs.

A) Histogram of the distribution of the CD90 protein expression score in the HSC/MPP human fetal bone marrow CD34+ enriched MNC CITE-seq dataset. HSCs were defined as CD90+ for analysis as described below.

B,C) AUCell score distribution for the human age independent platelet bias hHSC signature (B) and human age independent multilineage HSC signature (C) signatures at the level of each the whole human HSC compartment, as recorded in the young, old and fetal donor datasets . Pairwise comparison using the Wilcoxon rank-sum test – multiple testing adjusted p-values represented: ****: adjusted $P \leq 0.0001$.

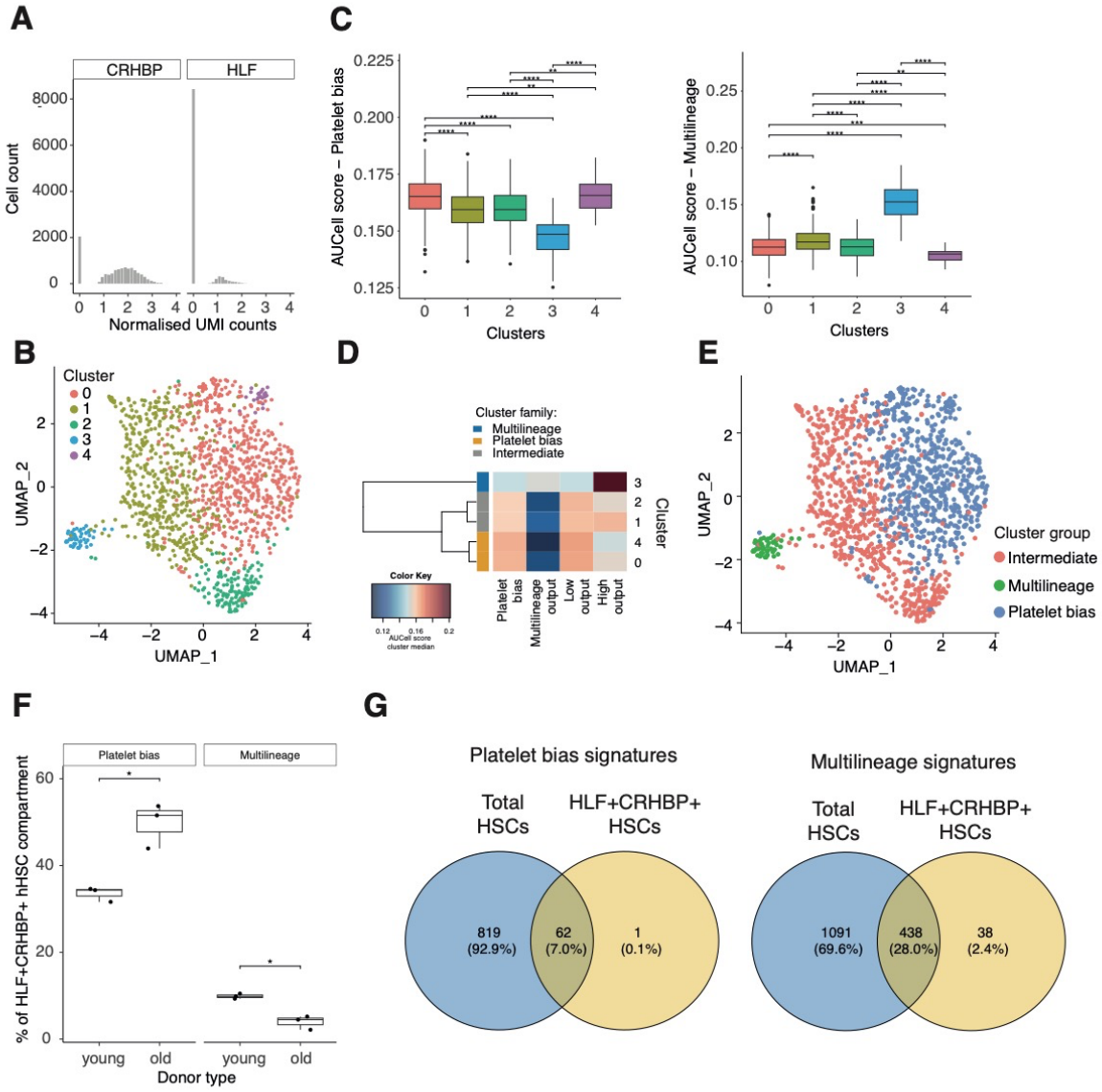
A
CRHBP
HLF
Cell count
Normalised UMI counts
C
AUCell score – Platelet bias
Clusters
AUCell score – Multilineage
Clusters
B
Cluster
UMAP_2
UMAP_1
D
Cluster family:
Multilineage
Platelet bias
Intermediate
Color Key
Platelet bias
Multilineage output
Low output
High output
Cluster
E
UMAP_2
UMAP_1
Cluster group
Intermediate
Multilineage
Platelet bias
F
Platelet bias
Multilineage
% of HLF+CRHBP+ hHSC compartment
young
old
young
old
Donor type
G
Platelet bias signatures
Total HSCs
HLF+CRHBP+ HSCs
819 (92.9%)
62 (7.0%)
1 (0.1%)
Multilineage signatures
Total HSCs
HLF+CRHBP+ HSCs
1091 (69.6%)
438 (28.0%)
38 (2.4%)

Figure S9

Fig. S9. Analysis of old CRHBP+HLF+ hHSCs.

A) UMI counts of CHRBP (left) and HLF (right) in 10X Chromium single cell RNAseq of old hHSCs. 1309 of 9981 cells were found to express both genes.

B) UMAP representation and unsupervised clustering of the CRHBP+HLF+ single cell transcriptomes from (A).

C) AUCell score distribution for the platelet bias and multilineage output HSC signatures postulated in (*8*) , after murine to human gene symbol conversion, in the hHSC clusters defined in (B). Pairwise comparison using the Wilcoxon rank-sum test – multiple testing adjusted p-values represented: **** adjusted P <= 0.0001, *** adjusted P <= 0.001, ** adjusted P < 0.01, * adjusted P < 0.05.

D) Clusters from (B) were hierarchically clustered of using median AUCell scores of platelet-biased and multi-lineage signatures, as well as high and low output HSC signatures, generating 3 distinct cluster groups.

E) Localization of the lineage-bias associated cluster groups in the UMAP representation from (B).

F) Quantification of the proportion of young and old CRHBP+HLF+ hHSCs within the platelet biased (left) and multi-lineage (right) cluster groups as a percentage of the total CRHBP+HLF+ hHSC population. P-values for the comparison of young and old hHSCs are shown: * P<0.05 (Student's t-test).

G) Venn diagrams showing the overlap between the human platelet (left) and multi-lineage signatures (right) generated by comparison of P- and ML-HSC cluster groups generated using the total old HSC population, or the old CHRBP+HLF+ HSC subset.

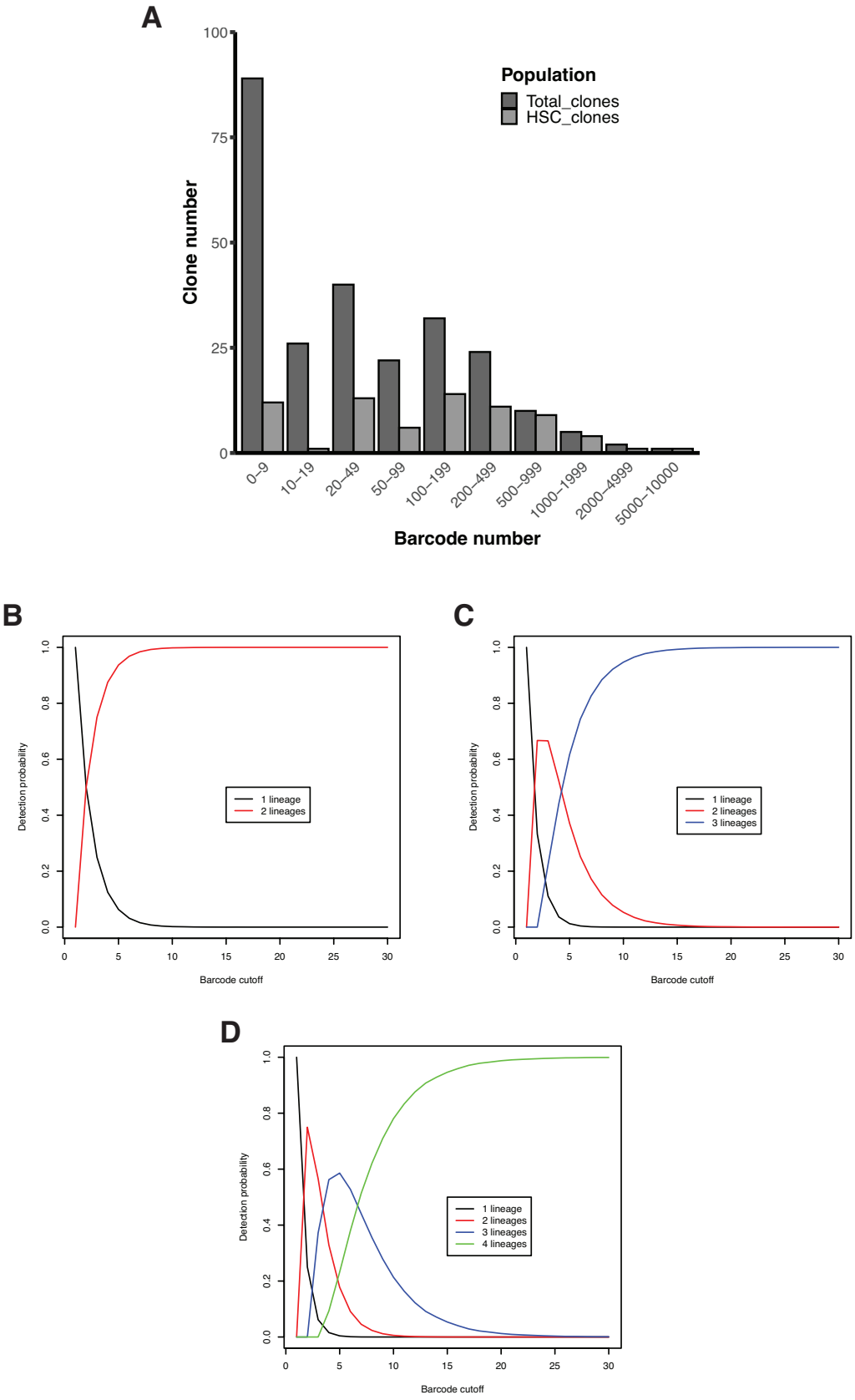
A
Population
Total_clones
HSC_clones
Clone number
100
75
50
25
0
0–9
10–19
20–49
50–99
100–199
200–499
500–999
1000–1999
2000–4999
5000–10000
Barcode number
B
Detection probability
1 lineage
2 lineages
Barcode cutoff
C
Detection probability
1 lineage
2 lineages
3 lineages
Barcode cutoff
D
Detection probability
1 lineage
2 lineages
3 lineages
4 lineages
Barcode cutoff

Figure S10

Fig. S10. Distribution of barcode reads in HSC clones.

A) Bar graph showing the number of barcodes detected with a number of reads within the indicated intervals by the bulk RNAseq (total clones), as well as the number of these barcodes that were detected in the single cell RNAseq of purified HSCs.

B) Modelling of the probability of detecting one or two lineages from an HSC clone with two lineage outputs as a function of the number of detected barcode reads, using random sampling and assuming equal probability of barcodes being found in each lineage.

C) Modelling of the probability of detecting one, two or three lineages from an HSC clone with three lineage outputs as a function of the number of detected barcode reads, calculated as in (B).

D) Modelling of the probability of detecting one, two, three or four lineages from an HSC clone with four lineage outputs as a function of the number of detected barcode reads, calculated as in (B).

Table S1. Donor used for 10X Chromium sequencing

Donors	Age	Gender	Number of cells purified	Total recovered cells	Mean reads per cell	Cells post QC
1	21	Male	12,000	4,327	172,155	4,279
2	24	Male	12,000	3,909	136,597	3,845
3	22	Male	12,000	5,648	94,207	5,509
4	68	Male	13,130	4,684	107,314	4,514
5	71	Male	6,658	2,638	97,334	2,560
6	75	Male	7,078	2,989	87,326	2,907

Table S2. List of antibodies used for FACS staining

Antigen	Fluorochrome	Clone	Dilution	Company	Catalogue no.	RRID
CD10	BV421	HI10a	1/20	BioLegend	312218	AB_2561833
CD10	PE-Cy5	HI10a	1/80	Biolegend	312206	AB_314917
CD117	BV605	104D2	1/100	BioLegend	313218	AB_2562025
CD11b	PE-Cy5	ICRF44	1/80	BioLegend	301308	AB_314160
CD123	PE-Cy7	6H6	1/20	BioLegend	306010	AB_493576
CD127	BV605	A019D5	1/20	BioLegend	351334	AB_2562022
CD14	PE-Cy5	61D3	1/80	Thermo Fisher Scientific	15-0149-42	AB_2573058
CD14	PE	61D3	1/100	Thermo Fisher Scientific	12-0149-42	AB_10598367
CD15	FITC	W6D3	1/500	BioLegend	323004	AB_756010
CD19	PE-Cy5	HIB19	1/160 (for sorting)	BioLegend	302210	AB_314240
CD19	PE-Cy7	HIB19	1/500 (for B cell readout)	BioLegend	302216	AB_314246
CD19	PE-Cy5	HIB19	1/80	Biolegend	302210	AB_314240
CD19	PE-Cy7	HIB19	1/100	Biolegend	302216	AB_314246
CD2	PE-Cy5	RPA-2.10	1/80	Biolegend	300210	AB_314034
CD20	PE-Cy5	2H7	1/160	Biolegend	302308	AB_314256
CD235ab	PE-Cy5	HIR2	1/2500	BioLegend	306606	AB_314624
CD235ab	PE-Cy5	HIR2	1/500 (for EryP cell readout)	Biolegend	306606	AB_314624
CD3	PE-Cy5	HIT3a	1/80	BioLegend	300310	AB_314046
CD33	PE	WM53	1/100	Biolegend	303404	AB_314348
CD34	APC	581	1/160	Biolegend	343510	AB_1877153
CD38	PE-Texas Red	HIT2	1/40	Thermo Fisher Scientific	MHCD3817	AB_10392545
CD4	PE-Cy5	RPA-T4	1/80	BioLegend	300510	AB_314078
CD41	AF700	HIP8	1/40 (for MkP analysis)	BioLegend	303728	AB_2566539
CD41	AF700	HIP8	1/100 (for Mk readout)	Biolegend	303728	AB_2566539

Table S3. List of primers used in single-cell and bulk RNA sequencing

Primer ID	**Primer sequence (5'- 3')**
Reverse custom primer	GAAAGCCATACGGGAAGCAATAGC
Forward custom primer	CGGCATGGACGAGCTGTACCAG
Reverse custom primer with ISPCR adapters	AAGCAGTGGTATCAACGCAGAGTGAAAGCCATACGGGAAGCAATAGC
Forward custom primer with ISPCR adapters	AAGCAGTGGTATCAACGCAGAGTCGGCATGGACGAGCTGTACCAG
ISPCR Primers	AAGCAGTGGTATCAACGCAGAGT
TSO	AAGCAGTGGTATCAACGCAGAGTACATrGrG+G

Table S4. HSC transplantation and mature lineage cell retrieval.

Experiment	Number of HSCs transplanted	Transduction efficiency (%)	Number of cells bulk sequenced per population				Unique barcodes retrieved from bulk RNAseq
			Bly	Myl	Mk	EryP	
1	4,000	5	300	300	300	300	3
2	14,000	16	500	500	500	500	4
3	13,000	18	500	500	500	500	23
4	13,000	18	500	500	500	500	25
5	13,000	18	500	500	500	500	39
6	13,000	18	500	500	500	500	59
7	13,000	18	500	500	500	500	9

Table S5. Single HSC transcriptomes sequenced and barcodes retrieved from xenografted mice.

Experi-ment	Total number of single HSCs sequenced	Total barcodes retrieved from HSC scRNAseq	HSC barcodes shared with bulk RNAseq	Number of HSCs transcriptomes with shared barcodes
1	39	3	2	37
2	77	6	2	55
3	65	11	7	58
4	53	12	10	45
5	130	30	18	105
6	44	14	11	40
7	49	10	7	37

Data S1: Genes differentially expressed between platelet-biased and multilineage hHSCs defined by clustering.
Genes differentially expressed between platelet-biased and multilineage hHSCs defined by clustering in young donors and old donors. Positive log fold change (FC) values highlight genes upregulated in platelet-biased hHSCs compared to multilineage hHSCs and column varLogFC contains an approximation of the variance of logFC. Column Pr..Chisq. contains the likelihood ratio test p-values and column fdr contains the Benjamini & Hochberg method adjusted p-values. Column coef contains the point estimate of the mixed effects hurdle model used for MAST DGEA for each gene, columns ci.hi and ci.lo contain the upper and lower bound of the confidence interval of the mixed effects model used for MAST DGEA for each gene and column z, the z score of the values contained in the coef column. Column Age highlights in which donor type the expression of the genes mentioned in the table was tested.

Data S2: Genes differentially expressed between platelet-biased and multilineage hHSCs defined by barcoding.
Genes differentially expressed between platelet-biased and multilineage hHSCs defined by barcoding. Positive average per cell population log2 FC values highlight genes upregulated in platelet-biased hHSCs defined by barcoding compared to multilineage hHSCs defined by barcoding. Columns pct.1 and pct.2 contain the percentage of cells expressing the tested gene in the platelet-biased and multilineage hHSC respectively. Column p_val contains the likelihood ratio test p-values and column BH_padj contains the Benjamini & Hochberg method adjusted p-values.

Data S3: Gene signatures used for computing AUCell scores.
Gene signatures used for computing AUCell scores. For the gene sets extracted from publications, the surname of the first author and the year of the publication as well as the cell populations in which the genes are upregulated is stated. Otherwise only the name of the cell populations in which the genes are upregulated is stated.

Data S4: Gene sets used for gene set enrichment analysis.
Gene sets used for gene set enrichment analysis. The column nomenclature is done in the same style as Data S3 with the addition that the Hallmark gene signatures (MSigDB) are in the columns stat contain the word Hallmark (capitalised).

Data S5: Normalized barcode frequencies.
Normalised barcode frequencies observed in differentiated lineages for barcodes also retrieved in HSCs. The barcodes names with suffix indicating recipient mouse number, as well as the normalised frequencies for B-cell (B), erythoid (EryP), megakaryocytic (Mk) and myeloid (My) cells is shown.

Data S6: Cluster-specific gene expression signatures.
Genes significantly up-regulated in each hHSC cluster from Fig. 2E compared to all other hHSCs in the experiment. Table shows cluster identity, gene ID, the log2 fold change, frequency of expression in cluster of interest (pct.1) and in the remining hHSCs (pct.2), as well

as the Cs defined by barcoding. Positive average per cell population log2 FC values highlight genes upregulated in platelet-biased hHSCs defined by barcoding compared to multilineage hHSCs defined by barcoding. Columns pct.1 and pct.2 contain the percentage of cells, as well as unadjusted (p_val) and adjusted (p_val_adj, Benjamini & Hochberg method) P-values.

Data S7: Source data. Tabs show the source data for the indicated figure panels. Where relevant the statistical test and analysis is shown.