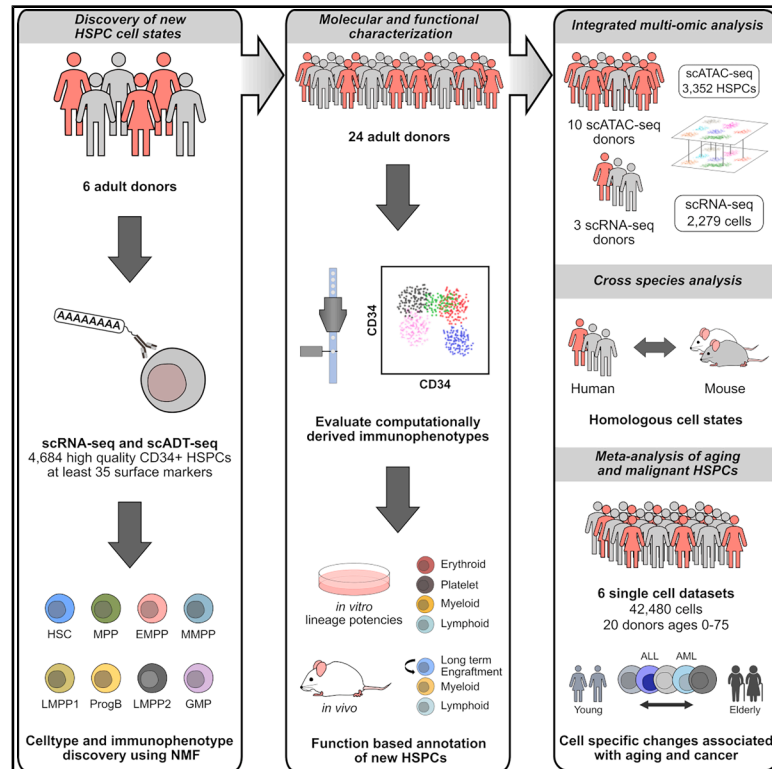


# A single-cell framework identifies functionally and molecularly distinct multipotent progenitors in adult human hematopoiesis

## Graphical abstract



## Authors

Asiri Ediriwickrema, Yusuke Nakauchi, Amy C. Fan, ..., Aaron M. Newman, Andrew J. Gentles, Ravindra Majeti

## Correspondence

rmajeti@stanford.edu

## In brief

Ediriwickrema et al. conducted a systematic evaluation of human bone marrow cells and provide an updated model of adult hematopoiesis, which includes distinct, immunophenotypically defined, multipotent and oligopotent progenitors. They also provide an extended biomolecular characterization of these cells and a framework for utilizing this reference in future research.

## Highlights

- Traditional human HSPCs contain functionally distinct sub-populations
- Adult human HSPCs express unique cell-type-specific gene expression changes with aging
- Human and mouse HSPCs exist in similar cellular states and differentiation trajectories



## Article

# A single-cell framework identifies functionally and molecularly distinct multipotent progenitors in adult human hematopoiesis

Asiri Ediriwickrema,<sup>1,2,3,4</sup> Yusuke Nakauchi,<sup>1,2,4</sup> Amy C. Fan,<sup>1,2,4,5</sup> Thomas Köhnke,<sup>1,2,4</sup> Xiaoyi Hu,<sup>1,2</sup> Bogdan A. Luca,<sup>6,7</sup> YeEun Kim,<sup>5,6</sup> Sreejith Ramakrishnan,<sup>8</sup> Margaret Nakamoto,<sup>8</sup> Daiki Karigane,<sup>1,2,4,9</sup> Miles H. Linde,<sup>1,2,4,5</sup> Armon Azizi,<sup>1,2,4</sup> Aaron M. Newman,<sup>7</sup> Andrew J. Gentles,<sup>4,6,7</sup> and Ravindra Majeti<sup>1,2,4,10,\*</sup>

<sup>1</sup>Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

<sup>2</sup>Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA, USA

<sup>3</sup>Cancer Biology Graduate Program, Stanford University, Stanford, CA, USA

<sup>4</sup>Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

<sup>5</sup>Immunology Graduate Program, Stanford University School of Medicine, Stanford, CA, USA

<sup>6</sup>Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

<sup>7</sup>Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA

<sup>8</sup>BD Biosciences, San Jose, CA 95131, USA

<sup>9</sup>Japan Society for the Promotion of Science, Tokyo, Japan

<sup>10</sup>Lead contact

\*Correspondence: [rmajeti@stanford.edu](mailto:rmajeti@stanford.edu)

<https://doi.org/10.1016/j.celrep.2025.116236>

## SUMMARY

Hematopoietic multipotent progenitors (MPPs) regulate blood cell production to meet the evolving demands of an organism. Adult human MPPs remain ill defined, whereas mouse MPPs are well characterized, with distinct immunophenotypes and lineage potencies. Using multi-omic single-cell analyses and functional assays, we identified distinct human MPPs within Lin<sup>−</sup>CD34<sup>+</sup>CD38<sup>dim</sup>/lo adult bone marrow with unique bio-molecular and functional properties. These populations were prospectively isolated based on expression of CD69, CLL1, and CD2 in addition to classical markers like CD90 and CD45RA. We identified a CD69<sup>+</sup> MPP with long-term engraftment and multilineage differentiation potential, a CLL1<sup>+</sup> myeloid-biased MPP, and a CLL1<sup>−</sup>CD69<sup>−</sup> erythroid-biased MPP. We used this updated hematopoietic stem and progenitor cell (HSPC) profile to study human and mouse bone marrow cells and observe unique cell-type-specific homology between species and cell-type-specific changes associated with human aging. By identifying and functionally characterizing adult MPP sub-populations, we provide a framework for future studies in hematopoiesis.

## INTRODUCTION

Hematopoiesis is a complex process that generates differentiated blood cells from a multipotent hematopoietic stem cell (HSC).<sup>1</sup> Maintaining hematopoiesis is critical for the needs of an organism through times of stress, infection, and illness. In humans, multipotent progenitors (MPPs) are important regulators of blood cell production and are key to maintaining appropriate cell output throughout life. The definition of human MPPs lacks clarity, which limits our understanding of human hematopoiesis. Historically, blood cells have been defined through prospective isolation using fluorescence-activated cell sorting (FACS) and functional evaluation using *in vitro* and *in vivo* assays.<sup>2,3</sup> These studies have led to the hierarchical description of hematopoiesis containing HSCs at the apex capable of long-term self-renewal and multilineage differentiation potential.<sup>2,4–7</sup> Historically, HSCs have been described as long-term HSCs (LT-HSCs) and

short-term HSCs (ST-HSCs) based on differential ability for multipotent engraftment upon serial transplantations. The HSC can generate increasingly abundant hematopoietic stem and progenitor cells (HSPCs) and, ultimately, terminally differentiated blood cells. This hierarchical model has been revised frequently<sup>8</sup>; however, it is believed that HSCs give rise to MPPs and lymphoid-primed MPPs (LMPPs), which have decreasing self-renewal capacity but persistent lymphoid and myeloid potential.<sup>5,6,9</sup> MPPs and LMPPs differentiate into increasingly lineage-restricted oligopotent and unipotent progenitors. Several oligopotent progenitor (OPP) sub-populations have been described, including common myeloid progenitors (CMPs), megakaryocyte-erythroid progenitors (MEPs), granulocyte-monocyte progenitors (GMPs), and common lymphoid progenitors (CLPs, or B cell natural killer [NK] cell progenitors).<sup>5,6,8,10–12</sup> The immunophenotypic and functional heterogeneities within OPPs have been re-evaluated, leading



to a better understanding of the underlying cellular content of these sub-populations.<sup>6,13</sup>

Importantly, adult human MPPs are functionally and molecularly heterogeneous within purified sub-populations using current FACS methods, and they have not been significantly refined based on their immunophenotypic profiles.<sup>5,13–16</sup> This reflects how little is known about human MPPs, especially when compared to the mouse. Murine hematopoiesis contains distinct MPP subsets,<sup>17,18</sup> which undergo functional changes with stress and aging. Specifically, four murine MPPs, MPP1–4, have been identified: MPP1 is considered a metabolically active HSC subset,<sup>19</sup> MPP2 and MPP3 are considered myeloid-biased MPPs,<sup>19</sup> and MPP4 is considered a lymphoid-biased MPP.<sup>20</sup> Similar immunophenotyping studies in humans are limited and often involve cord blood (CB),<sup>5,13,21</sup> which can be functionally different than adult bone marrow (BM).<sup>13,22–25</sup> Further, studies in both humans and mice suggest that early progenitors, defined by traditional immunophenotypes, are lineage primed and, therefore, question the existence of immunophenotypically defined OPPs like the CMP.<sup>13,15</sup>

Although it has been important to use immunophenotypes to purify and study HSPCs, this approach has led to the idea that hematopoiesis progresses through a stepwise process.<sup>2</sup> Recent studies re-evaluating hematopoiesis using single-cell methods challenge this view, as they present primitive hematopoiesis as a more gradual and continuous process.<sup>7,8,14,26</sup> These observations are important but are not radical, as surface marker expression is continuous, and discrete bins of expression are not always present. Regardless, the hierarchical resolution of mouse hematopoiesis using surface markers, particularly the MPP compartment, is finer than in humans and has led to significant advances in our understanding of hematopoiesis. The need remains to resolve primitive human MPPs into distinct, immunophenotypically defined, functionally distinct sub-populations to prospectively isolate and study these cells.

Analysis of single-cell RNA sequencing (scRNA-seq) data from FACS-isolated murine HSCs and MPPs suggests that these cells also exist in a continuum, despite being immunophenotypically distinct.<sup>15,27</sup> It is therefore important to not only adapt prior methods from the study of mouse hematopoiesis to human HSPCs but also to consider a multi-modal framework to improve our ability to both compartmentalize and study hematopoiesis and associated blood disorders.<sup>28,29</sup> Although multiple scRNA-seq studies on healthy human hematopoiesis have been published over the past several years,<sup>6,14,30–37</sup> only a few have incorporated multiple analytes like concurrent scRNA-seq and surface marker expression analysis using antibody-derived tags (ADTs; scADT-seq).<sup>31,35,37,38</sup> Importantly, the functional evaluation of adult MPP sub-populations observed in these studies has been limited.

Here, we present a systematic approach for analyzing multi-omic single-cell data from adult human BM and provide a framework for rigorously evaluating the HSPC compartment of adult hematopoiesis using computational methods and functional assays (Figure S1). As a result, we observe greater phenotypic heterogeneity within CD34+ cells in adult BM compared to CB and define 4 functionally distinct progenitors based on the expression of CD90, CD69, and CLL1 within the lineage-negative

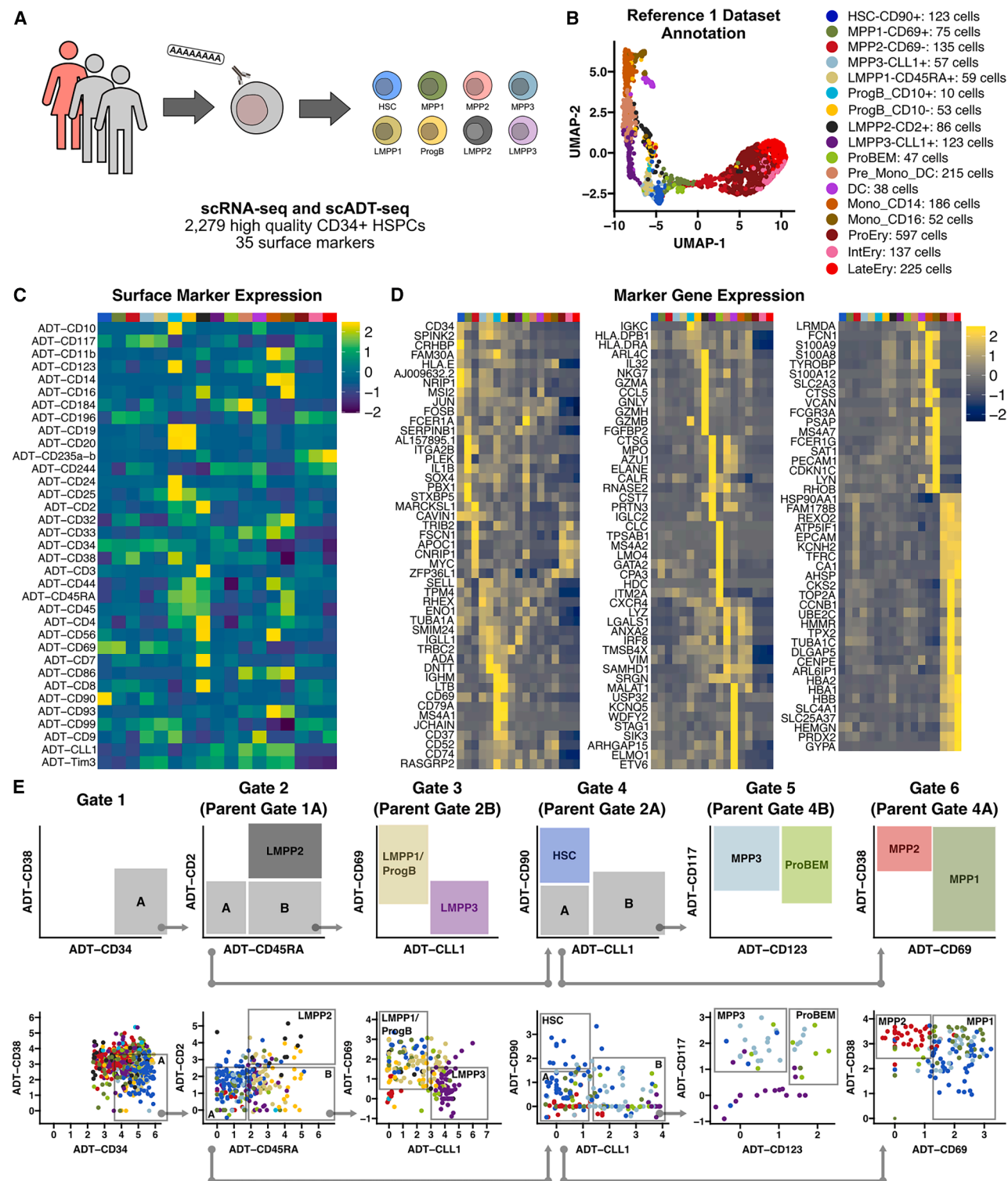
(Lin<sup>−</sup>)CD34+CD38dim/loCD45RA−CD2− compartment. Using complementary *in vitro* and *in vivo* assays, we observe that (1) the Lin−CD34+CD38dim/loCD90−CD45RA−CLL1−CD69+ adult BM population contains a human MPP with long-term engraftment capacity and 7 lineage differentiation potentials, (2) the Lin−CD34+CD38dimCD90−CD45RA−CLL1−CD69− population contains an erythroid-biased human MPP, and (3) the Lin−CD34+CD38dimCD90−CD45RA−CLL1+ population contains a myeloid-biased MPP. Both the myeloid- and erythroid-biased MPPs have significantly decreased engraftment potential *in vivo*. We also observe 3 distinct OPPs within the Lin−CD34+CD38dim/loCD45RA+ adult BM population with differential CD2, CD69, and CLL1 surface marker expression: CD2+ LMPPs with both myeloid and lymphoid potencies, CD2−CLL1− LMPPs with high lymphoid potency, and CD2−CLL1+ GMPs with minimal lymphoid potency. Integrated single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) analysis revealed distinct patterns of chromatin accessibility among the HSPCs. Using the updated HSPC profile as a reference, we also performed a cross-species analysis, which allowed us to identify similar cell states between mouse and human HSPCs. Finally, we performed a large-scale ( $n = 42,480$  cells) computational analysis across 20 human donors (ages 0–75) and observed cell-type-specific transcriptional changes associated with aging, and we evaluated those changes in common blood cancers. Our work provides additional insights into adult human HSPC biology with implications for future multi-omic studies.

## RESULTS

### Clustering and computational purification of human adult HSPC populations

To investigate human adult HSPC populations, we conducted concurrent scRNA-seq and scADT-seq on 10,297 cells isolated from 3 normal adult human BM donors (Figure S1). Prior scRNA-seq studies suggest that early HSPC sub-populations in adult BM are difficult to sub-cluster.<sup>14</sup> However, the inclusion of single-cell ADT information allowed us to not only remove poor-quality cells that persisted despite standard quality control (QC) methods (Figures S2A and S2B) but also to improve subsequent clustering and integration. As an exploratory analysis of the dataset, we first estimated the number of clusters within canonical HSPC sub-populations<sup>5</sup> using both gene and surface marker expression. Single cells from 3 donors (Table S1) were computationally isolated using ADT counts and conventional markers used in FACS-based gating (Table S2).<sup>30</sup> The number of sub-clusters for each computationally purified sub-population was estimated using the Gap statistic<sup>39</sup> on integrated features from concurrent gene expression and ADT data (Figures S2E and S2F). We observed stable sub-clustering within the CD34+CD38dim/lo sub-population,<sup>5</sup> which motivated the subsequent unsupervised sub-clustering analysis.

We evaluated two clustering methods, the Louvain algorithm<sup>40</sup> and non-negative matrix factorization (NMF),<sup>41</sup> and observed that NMF-based clustering generated more sub-clusters within the HSPC compartment. We therefore performed NMF, which revealed several candidate sub-clustering levels based on the



**Figure 1. Candidate HSPC sub-populations were identified using NMF and express unique immunophenotypes**

(A) Schematic of single-cell analysis workflow.

(B) Uniform manifold approximation and projection (UMAP) projection of the reference single-cell dataset (reference 1) and final cell counts for each cluster.

(legend continued on next page)



cophenetic index (Figure S2G).<sup>41,42</sup> To identify the optimal level of sub-clusters, we first split the data into a training and a test dataset (STAR Methods) and generated a signature matrix using the training dataset for deconvolution with CIBERSORTx.<sup>43</sup> The test dataset was then over-clustered ( $k = 18$ ; Figure S2H), and transcriptionally similar clusters were collapsed to optimize the deconvolution performance with CIBERSORTx in pseudo-bulk samples. This semi-supervised approach resulted in 15 high-quality clusters. We subsequently evaluated the strength of our sub-clustering approach by performing intra-dataset cell-annotation transfer (Figure S2I). This reference profile (semi-supervised NMF 15) performed similarly to, if not better than, alternative clustering approaches and published annotations<sup>31,32,35</sup> based on bulk and cluster-specific F1 scores while maintaining stable clustering parameters (Figures S2J and S2K). Our clustering approach allowed HSPC sub-population discrimination, whereas published annotations and the unsupervised Louvain approach generally combined these cells into a single cluster (Figures S2L–S2P).

The preliminary cluster annotations were derived based on the differential expression of marker genes and proteins (Figures 1B–1D; Table S3). We identified 4 clusters within the canonical MPP gate—MPP1, MPP2, MPP3, and a basophil (Ba), eosinophil (Eo), and mast cell (Mc) progenitor (ProBEM)—and 4 clusters within the canonical LMPP gate: LMPP1, LMPP2, ProB, and LMPP3. To evaluate cluster validity, we quantified donor integration using Seurat's canonical correlation analysis (CCA), scGen, and Scanorama (Figures S3A–S3D). The CCA method was superior to other methods as measured by the local inverse Simpson's index (Figures S3E and S3F).<sup>44–47</sup>

After creating this HSPC reference, we derived a sorting strategy for these HSPCs by iteratively calculating differentially expressed ADTs for each cluster (Figure 1E; Table S2). We then evaluated if immunophenotypically defined HSPCs exist in early transcriptional cell states. To address this question, we performed an iterative trajectory analysis using dynverse<sup>48</sup> and partition-based graph abstraction<sup>49</sup> (PAGA; Figure S3G). Cells were first computationally purified using the updated sorting strategy and labeled based on their immunophenotype. When evaluating HSPCs only, we see an enrichment of HSCs, MPP1s, and LMPP1s in milestone 1 (Figure S3H; Table S6). To further evaluate MPP development, we performed a subsequent analysis on HSCs and MPPs (MPP1-2, myeloid-primed MPP [MMPP], and LMPP1). We see that HSCs and MPPs reside in earlier milestones, whereas erythroid-primed MPPs (EMPPs), MMPPs, and LMPP1s track concordantly down developmental milestones (Figures S3I and S3J). When looking only at immunophenotypically defined cells based on ADT expression, we see expected patterns of enrichment within sequential milestones based on the known developmental hierarchy of hematopoiesis. Further, the differential ADTs that were informative for isolating HSPCs were also identified as important milestone-defining ADTs (Table S6). However, immunophenotypically defined

HSCs and MPPs exist across multiple milestones, which suggests that they occupy a transcriptional continuum, and the updated sorting strategy can enrich for cells that reside in early transcriptional states (Figures S3K and S3L). Further, computational purification using the updated sorting strategy produced greater target population enrichment and lower population diversity compared to the traditional sorting strategy (Figures S4A and S4B). Improvements in purity were higher within the traditional MPP sub-population (Figure S4C). From the analysis, we identified CD69, CLL1, and CD2 as important antigens for purifying HSPCs, in addition to conventional markers CD34, CD38, CD45RA, and CD90.

We also observed a candidate B cell progenitor, ProB, in the canonical LMPP population, based on both gene and surface markers (Figures 1E, S7A, and S7B; Table S3). Prior work suggests that CD117, CD10, and CD19 are important markers for discriminating B cell progenitors,<sup>50</sup> so we evaluated ProB cells for CD117, CD10, and CD19 expression (Figure S4D) and sub-annotated them based on CD10 expression (Figure 1B). The majority of ProB cells express low levels of CD117 and CD19 and do not express CD10 (ProB\_CD10<sup>−</sup>), and only a rare population of ProB cells express CD10 (ProB\_CD10<sup>+</sup>). Both CD19 and CD10 expressions were lower in ProB cells than in lymphocytes. We also observed that CD2 expression marked a candidate LMPP, LMPP2, which expresses cytotoxic genes at higher levels compared to other CD34<sup>+</sup> HSPCs (Figure S4E). The complete immunophenotypes are reported in Table S2. Lineage-defining marker expression compared to all cell types, including lymphocytes, and cell cycle scores is presented in Figures S4D–S4F. As expected, HSCs exhibit low S and G2M scores, whereas the remaining HSPCs upregulate both S and G2M signatures, suggesting cell cycle entry at variable levels. Overall, our observations support a semi-supervised approach, i.e., unsupervised NMF clustering followed by manual curation with CIBERSORTx, for identifying cell types in our data.

### An updated flow cytometry-based sorting strategy for purifying and studying primitive HSPCs

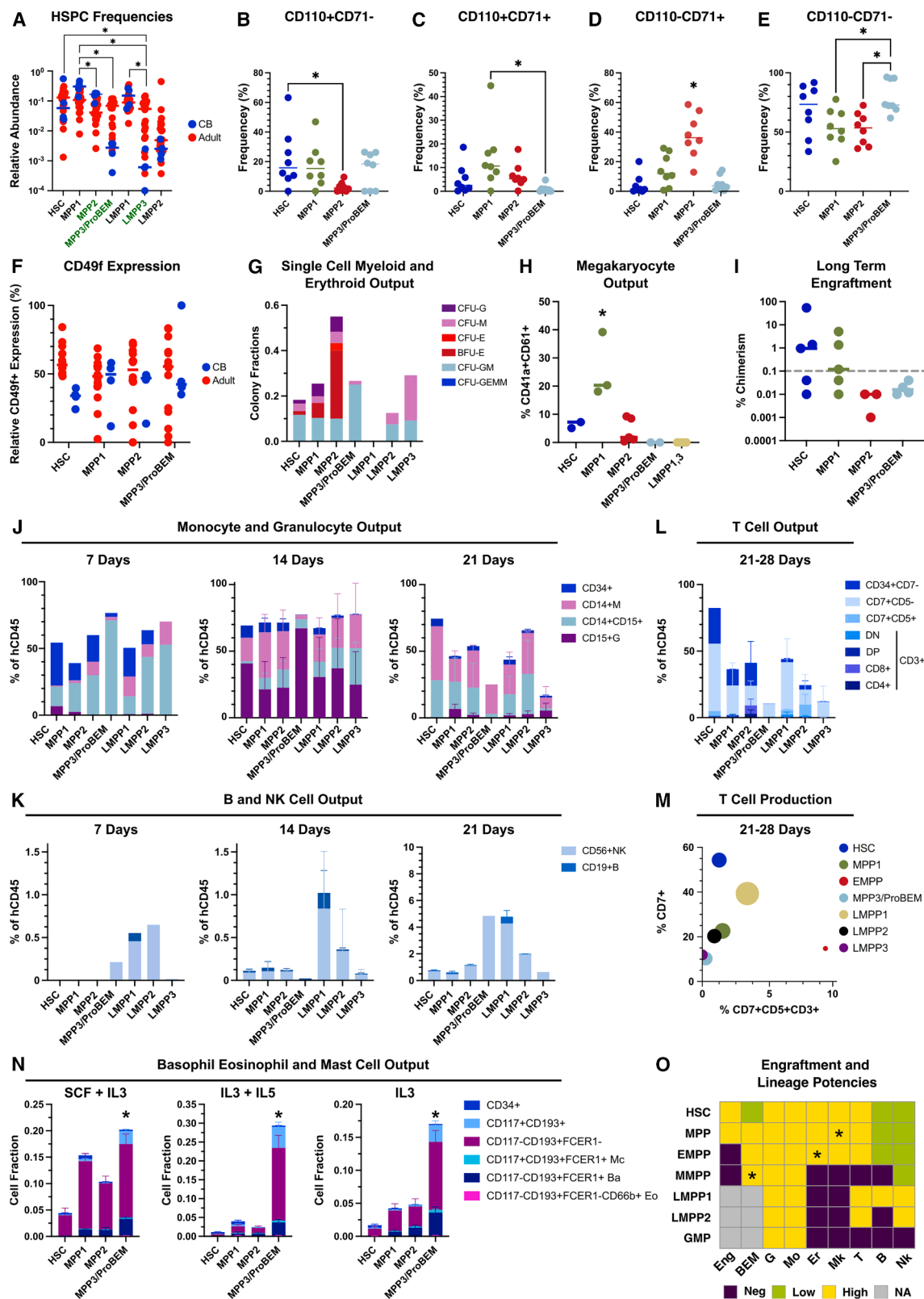
As a primary goal of this study is to understand the functional characteristics of the MPP and LMPP sub-populations, we designed a FACS strategy (Figure S5) using the computationally derived immunophenotypes (Table S2; STAR Methods). We used this sorting strategy to evaluate the proportion of each cell type within the Lin<sup>−</sup>CD34<sup>+</sup>CD38dim/lo sub-population across 24 adult BM and 4 CB donors (Figure 2A) to compare the cellular composition between different donor sources. We observed an increase in MPP3s/ProBEMs and LMPP3s and a decrease in MPP2s in adult BM compared to CB. The greater degree of heterogeneity in adult BM is partly marked by the expression of CLL1 in CD34<sup>+</sup>CD38dim/lo cells compared to CB (Figure S2D). Prior work has demonstrated the utility of CD110, CD71, and CD49f in identifying HSC and MPP subsets in CB.<sup>13,51</sup> We therefore evaluated the expression of these markers

(C) Heatmap showing average ADT expression for each cluster.

(D) Heatmap showing average gene expression for top features in each cluster.

(E) A schematic illustrating an updated sorting strategy for purifying HSPC sub-populations.

Clusters are coded by color on the top of the heatmap (C and D) and in the ADT plots (E) using the legend in (B).



(legend on next page)

among the HSPCs and observed greater CD71-expressing cells and fewer CD110-expressing cells in MPP2s compared to HSCs, MPP1s, and MPP3s (Figures 2B–2E). Most HSCs and MPP1s expressed higher CD110 compared to MPP2 and MPP3s. There was no significant difference between the frequency of CD49f-expressing cells within HSCs and MPP cell types in either adult BM or CB (Figure 2F).

Although candidate HSPCs were rare in adult BM, we were able to purify these sub-populations using the updated sorting strategy for the subsequent assays (Table S7). In single-cell methylcellulose-based colony-forming unit (CFU) assays, HSCs, MPP1s, and MPP2s had both myeloid and erythroid clonogenic potential, whereas MPP3s, LMPP2s, and LMPP3s had only granulocyte and monocyte potential (Figure 2G). MPP2s demonstrated the highest clonogenic output, whereas LMPP1s did not expand in this assay. To evaluate megakaryocyte production, we assayed purified HSPCs *in vitro* in a serum-free liquid culture system. After 7 days of culture, MPP1s produced the greatest relative percentage of CD41a+CD61+ megakaryocyte progenitors (Figure 2H). HSCs and MPP2s produced minimal CD41a+CD61+ megakaryocyte progenitors, whereas MPP3s, LMPP1s, and LMPP3s did not produce megakaryocyte progenitors.

HSPC populations were subsequently evaluated for long-term engraftment in *NOD/SCID/IL2Rγ<sup>-/-</sup>* (NSG) mice (Table S8). Purified HSCs and candidate MPPs were injected intrafemorally, and human chimerism was measured between 17 and 18 weeks.<sup>52</sup> HSCs and MPP1s were able to produce long-term engraftment compared to MPP2s and MPP3s/ProBEMs (Figure 2I). In all engrafted mice, we observed the presence of immunophenotypically defined HSCs, MPP1s, MPP2s, and MPP3s/ProBEMs and both CD33+ and CD19+ cells within the human CD45+HLA+ sub-population.

To characterize lineage potencies across all HSPCs, we performed a series of *in vitro* assays. Combined lymphoid and myeloid differentiation potential was assessed in liquid culture on OP9 cells with SCF, G-CSF, FLT3L, interleukin (IL)-2, DuP-697, and IL-15.<sup>6</sup> Purified HSPCs were cultured for 3 weeks, and aggregate cellular output was quantified in Figures 2J and 2K. All cell types produced hCD45+CD14+ monocyte-macrophages (Ms) and CD15+ neutrophil-granulocytes (Gs), with a noticeable burst of Gs at 14 days. LMPP1-2 cells produced the

greatest amount of hCD45+CD14–CD15–SSA-lowCD56+ NK cells, whereas HSCs, MPP1s, MPP2s, and MPP3s/ProBEMs produced moderate to minimal NK cells. LMPP1 cells also produced the greatest amount of hCD45+CD14–CD15–SSA-lowCD19+ B cells, whereas HSCs, MPP1s, and MPP2s demonstrated moderate to minimal B cell production.

T cell potency was assessed using an *in vitro* liquid culture system with OP9 stromal cells expressing human notch receptor ligand DL4 (OP9-hDL4).<sup>53</sup> Purified HSPCs were cultured with cytokines SCF, FLT3L, and IL-7 for 21–28 days.<sup>6</sup> HSCs, MPP1s, MPP2s, LMPP1s, and LMPP2s produced moderate hCD45+SSA-lowCD7+CD5+CD3+ cells, with a few samples producing minimal CD4+, CD8+, or CD4+CD8+ double-positive (DP) cells (Figure 2L). MPP3s produced very few immunophenotypic T cells, and LMPP3s produced no immunophenotypic T cells. LMPP1 cells generated the greatest total cellular production after 21–28 days in culture (Figure 2M). BEM production was determined using an *in vitro* serum-free culture system. HSPCs were expanded and evaluated for BEM production after 7 days. MPP3s produced the greatest relative BEM production, followed by MPP1s and MPP2s (Figure 2N).

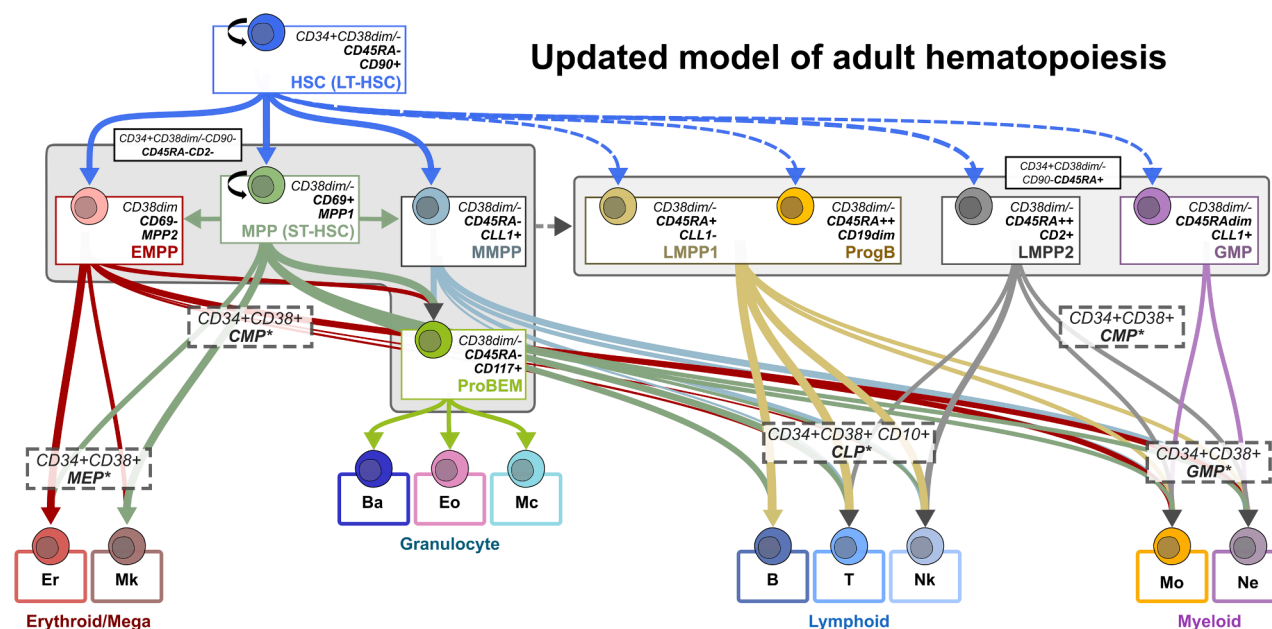
Both HSCs and MPP1s demonstrated long-term engraftment and 7 lineage differentiation potentials. MPP1s demonstrated greater megakaryocyte differentiation potential, MPP2s demonstrated greater erythroid potential, and MPP3s/ProBEMs demonstrated greater BEM potential compared to other HSPCs *in vitro*. Although LMPP1s had the greatest lymphoid potential *in vitro*, all HSPCs evaluated in our assays, except for LMPP3s, contain a degree of lymphoid potency. Based on these results, MPP1 was annotated as an MPP, MPP2 was annotated as an EMPP, MPP3 was annotated as a MMPP, and LMPP3 was annotated as a GMP (Figure 2O).

### An updated model of adult human hematopoiesis

An updated model of adult hematopoiesis based on our computational and functional analyses is presented in Figure 3. Each HSPC's differentiation potential is designated by arrows, with the thickness corresponding to the strength of the lineage potency. Annotations and arrows that were not directly studied in our functional assays are marked with dashed lines and were incorporated into the model based on our computational analysis and prior observations. The canonical CD90+ HSC can be

**Figure 2. Computationally identified HSPCs can be purified using flow cytometry and exhibit distinct functional properties**

- (A) HSPC frequencies across 24 adult bone marrow and 4 cord blood donors.  
(B–E) Cells within each respective gate were evaluated for CD71 and CD110 expression, and relative frequencies are displayed.  
(F) CD49f expression of cells within each respective gate.  
(G) Single-cell methylcellulose colony-forming unit (CFU) assay of clonogenic output and lineage differentiation potencies of adult HSPCs.  
(H) Proportions of CD41a+CD61+ megakaryocyte progenitors produced after 7-day *in vitro* expansion of each adult HSPC.  
(I) Human to mouse chimerism 17–18 weeks post-primary transplantation of purified HSPCs. Chimerism >0.1% was considered positive for long-term engraftment.  
(J and K) Monocyte and granulocyte (J) and B and NK cell (K) output after 7–21 days of expansion in liquid culture with OP9 cells and SGF15 media.  
(L) T cell output after 21–28 days of expansion in liquid culture with OP9-hDL4 cells and SF7 media. CD7+CD5+CD3+ cells were split based on CD4 and CD8 expression, and the remaining cells did not express CD3.  
(M) T cell production after 21–28 days of expansion. The size of each circle correlates with hCD45+/SSA-low cell production normalized to 100 cells seeded.  
(N) Ba, Eo, and Mc output after 7 days of expansion with SCF, IL-3, or IL-5.  
(O) Summary of engraftment and lineage differentiation potencies for all HSPCs.  
DN, double negative; DP, double positive. Pooled data are represented as mean ± SEM. Statistical significance was tested using a t test or ANOVA with multiple comparison adjustments. \*Adjusted *p* < 0.05.



**Figure 3. Updated model of adult hematopoiesis**

A flow chart illustrating an updated framework for organizing the adult hematopoietic hierarchy. Arrows depict differentiation potencies with relative thickness highlighting the strength of that potency. Dashed lines designate cell populations and differentiation trajectories that have been previously described or were evaluated computationally in the current study. The updated model highlights the heterogeneity within the canonical MPP and LMPP populations in adult bone marrow. We propose that within the canonical Lin- $CD34+CD38dim/loCD90-CD45RA-$  MPP sub-population, the MPP resides in  $CLL1-CD69+$  cells, and  $CLL1+$  cells are myeloid biased, whereas  $CLL1-CD69-$  cells are erythroid biased. Within the canonical Lin- $CD34+CD38dim/loCD90-CD45RA+$  LMPP population, we propose that  $CD2-CLL1-LMPP1$ s have the strongest lymphoid potency,  $CD2+LMPP2$ s have both lymphoid and myeloid potency with high expression of cytotoxic genes, and  $CD2-CLL1+$  cells are primarily GMPs.

further purified by removing  $CD2+$  and  $CLL1+$  cells. The 4 clusters within the canonical  $CD90-CD45RA-$  MPP gate can be purified using  $CD69$  and  $CLL1$  and show distinct functional properties. The canonical  $CD90-CD45RA+$  LMPPs contain 4 additional clusters that can be purified using  $CD2$ ,  $CLL1$ , and  $CD69$ .

### A mixed-model classifier improves cell annotation transfer across scRNA-seq datasets

After identifying functionally and molecularly distinct HSPCs, we tested if we could identify these populations in external scRNA-seq datasets. An important requirement for this analysis is accurate cell annotation transfer. To identify the optimal cell classifier, we adapted a prior method<sup>54</sup> to incorporate both model- and cell-type-specific metrics (Figure S6). The optimal cell-type-specific method was incorporated into the final classifier, i.e., a mixed-model classifier, and was evaluated further using an inter-dataset analysis (Figure S6E). We evaluated the performance of the mixed-model classifier on 14 external scRNA-seq datasets.<sup>31,32,35,36,55-64</sup> The model performed well on both adult and pediatric BM cells (Figure S7A); however, the performance was significantly reduced when annotating mobilized peripheral blood (mPB) and CB (Figure S7B). This would suggest that mPB and CB HSPCs have more distinct cell states compared to adult and pediatric HSPCs. We also evaluated model performance by comparing our reference annotation assignments to the original cell annotations in 6 datasets containing this information and

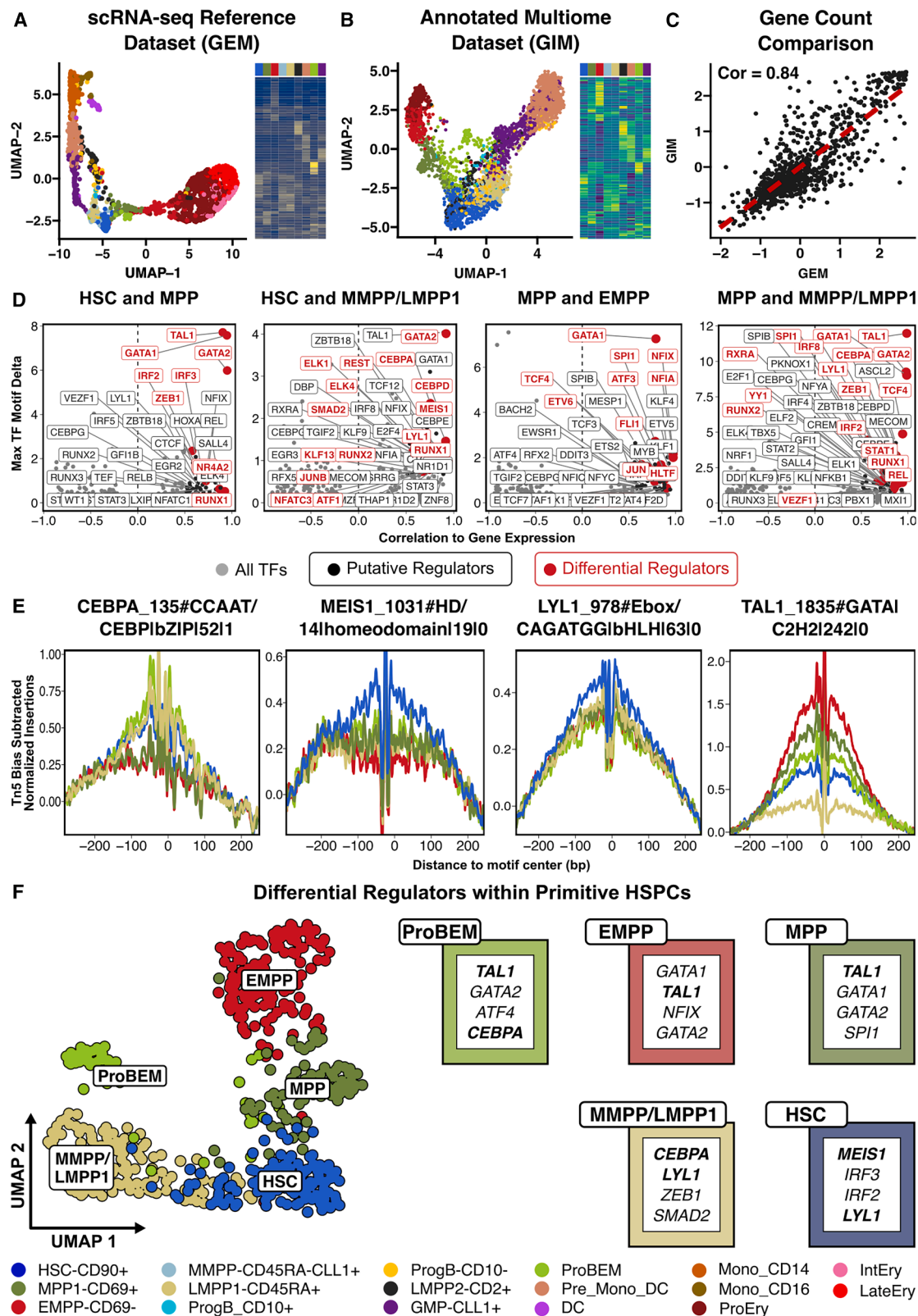
found good concordance (Figure S7C). Our annotation assignments allowed for further deconvolution of progenitor cell states in the external data. Heatmaps illustrating marker gene expression for our reference dataset and the integrated external datasets are presented in Figure S7D. The optimized mixed-model classifier was used for the subsequent analysis of external human and mouse scRNA-seq data.

### HSPC states can be isolated in a second cohort of adult donors

To evaluate the validity of the updated HSPC reference, we profiled a second cohort of adult donors ( $n = 3$ , reference 2) using the same workflow (Figure S1D). After implementing our updated QC methods, we recovered 2,405 high-quality HSPCs. Using the mixed-model classifier, we were able to identify similar cell states (Figure S8A) with high confidence as measured by the F1 score using the inter-dataset approach (Figure S8B). Using these labels, we observed similar clustering metrics compared to reference 1 (Figures S8C and S8D). We then computationally purified and annotated cells in references 1 and 2 using ADT information and the updated sorting strategy (Figure 1E) and compared average gene expression across datasets. Analogous cell populations in each reference showed highly correlated average gene expression (Figures S8E-S8G), which supports the reproducibility of our observations.

Cells from reference 2 were profiled using an expanded ADT panel (Table S5), which included previously validated HSC and





(legend on next page)



MPP markers like CD49f and CD52.<sup>13,16,51</sup> We therefore evaluated CD52 and CD49f expression in the MPP sub-gate and observed poor discrimination of cells using these markers (Figures S9A–S9C). Separating cells based on CD69 expression produced a significantly more homogenous population compared to CD49f and CD52 (Figure S9D). The MPPs were also gated based on CD52 or CD49f expression with respect to CD69 (Figures S9E and S9F) to determine if those markers could produce transcriptionally distinct sub-populations. Rather, we observed that CD69+ MPPs separated based on CD52 and CD49f expression remained transcriptionally similar, suggesting the limited utility of these markers to identify additional MPP sub-populations (Figures S9G and S9H).

### MPP populations are associated with unique changes in TF networks

Transcription factors (TFs) modulate specific gene regulatory networks (GRNs), which are critical for determining cell identity, function, and lineage commitment.<sup>15,30,65,66</sup> Studies evaluating TF GRNs in hematopoiesis suggest that specific cell fate decisions are regulated by circuits involving several TFs rather than the sequential expression of individual factors.<sup>67,68</sup> These observations may be particularly relevant for HSPCs, which can be regulated by small variations in several unique TFs.<sup>66,68</sup> We therefore evaluated if HSPC states could be characterized by individual TFs or a combination of factors. Motif analyses have been useful for evaluating differential GRNs and cell states<sup>30</sup>; however, TF families often share similar DNA-binding motifs, which confounds the analysis. To address these questions and the associated limitations, we used our updated HSPC reference to analyze published combined scATAC-seq and single-nucleus RNA-seq (snRNA-seq) data (ATAC-RNA) from healthy adult BMMCs and a curated motif reference.<sup>69–71</sup> ATAC-RNA cells were integrated with scRNA-seq cells using snRNA-seq data with cytosolic gene expression (scRNA-seq) (Figures 4A and 4B), which enabled cross-platform cell annotation transfer and gene expression integration of ATAC-RNA cells. The expression of key genes from the gene integration matrix (GIM) for the annotated ATAC-RNA cells are compared to their expected expression from the reference gene expression matrix (GEM) (Figures 4A and 4B). Averaged normalized counts per cluster for these differentially expressed genes were highly correlated between the two gene expression matrices (Figure 4C).

The integrated dataset was used to evaluate correlated motif accessibility with TF gene expression (putative regulators). Given the underlying noise associated with cross-platform cell annotation transfer and gene expression inference, we identified puta-

tive regulators that also varied significantly across differentiation trajectories based on our functional studies (differential regulators) to identify TFs that were most likely to influence HSPC GRNs (Figure 4D). As the cell types observed from these studies reside in canonical HSCs and MPPs (HSCs, MPPs, EMPPs, MMPPs/LMPP1s, and ProBEMs), we focused the TF GRN analysis on these HSPCs. Representative footprint plots for important TFs identified using the differential regulator approach are provided in Figure 4E, and a summary of the results is illustrated in Figure 4F. We observe that *MEIS1* and *LYL1* motifs are more accessible in HSCs compared to MPPs and MMPPs/LMPP1s (Figure 4E). We also see that TF accessibility for primitive HSCs/MPPs is lower than that of more differentiated progenitors, i.e., *TAL1* in EMPPs (Figure 4E). We also evaluated whether the functional capacity of the HSPCs associates with known lineage-defining TF GRNs. Motif accessibility for key TFs, including *GATA1*, *SPI1*, and *IRF8*, shows expected variations among HSPCs (Figure S10A) based on the lineage potencies observed from our functional studies (Figure 2), which is further supported by the observed GRNs for key marker genes like *HBB*, *CCR3*, *TBX21*, *ITGAM*, *ITGA2B*, *FCER1A*, *IL7R*, and *CCR2* (Figure S10B). Therefore, the integrated analysis supports the observed functional potencies and suggests that a network of lineage-defining TFs gains activity in primitive progenitors based on chromatin accessibility associated with specific cell states.

### An iterative integration of single-cell gene expression can identify homologous cell states between mouse and human HSPCs

Research in murine hematopoiesis has identified immunophenotypically distinct HSPCs with unique functional characteristics.<sup>17,18</sup> These include sub-populations of murine HSCs, i.e., CD45+EPCR+CD48–CD150+ (ESLAM) HSCs, LT-HSCs, ST-HSCs,<sup>72,73</sup> and MPPs with varying lineage potencies, i.e., MPP2-4<sup>17</sup> and MPP Ly I/II.<sup>29</sup> Because our analysis provided evidence that lineage-biased MPPs exist in humans as well, we next attempted to identify their homologous cell states in murine BM to support future studies in hematopoiesis. To perform this analysis, we evaluated two single-cell datasets with concurrent single-cell mRNA and surface marker quantification in young adult (2–3 months)<sup>73</sup> and adult (8–10 months) mice BM.<sup>29</sup> These data contained HSC and MPP annotations, which allowed for a cell-type-specific cross-species analysis with our reference human HSPC profiles (Figure S11). We first performed a gene ortholog-to-ortholog conversion of the mouse data and quantified murine expression patterns of gene signatures associated with human cells (Figure S12A). Gene signatures were calculated

#### Figure 4. Adult HSPCs express distinct changes in chromatin accessibility

(A) UMAP of scRNA-seq reference and heatmap of top differentially expressed genes.

(B) UMAP of scATAC-seq cells labeled based on integration with scRNA-seq data and heatmap of the differentially expressed genes within the gene integration matrix.

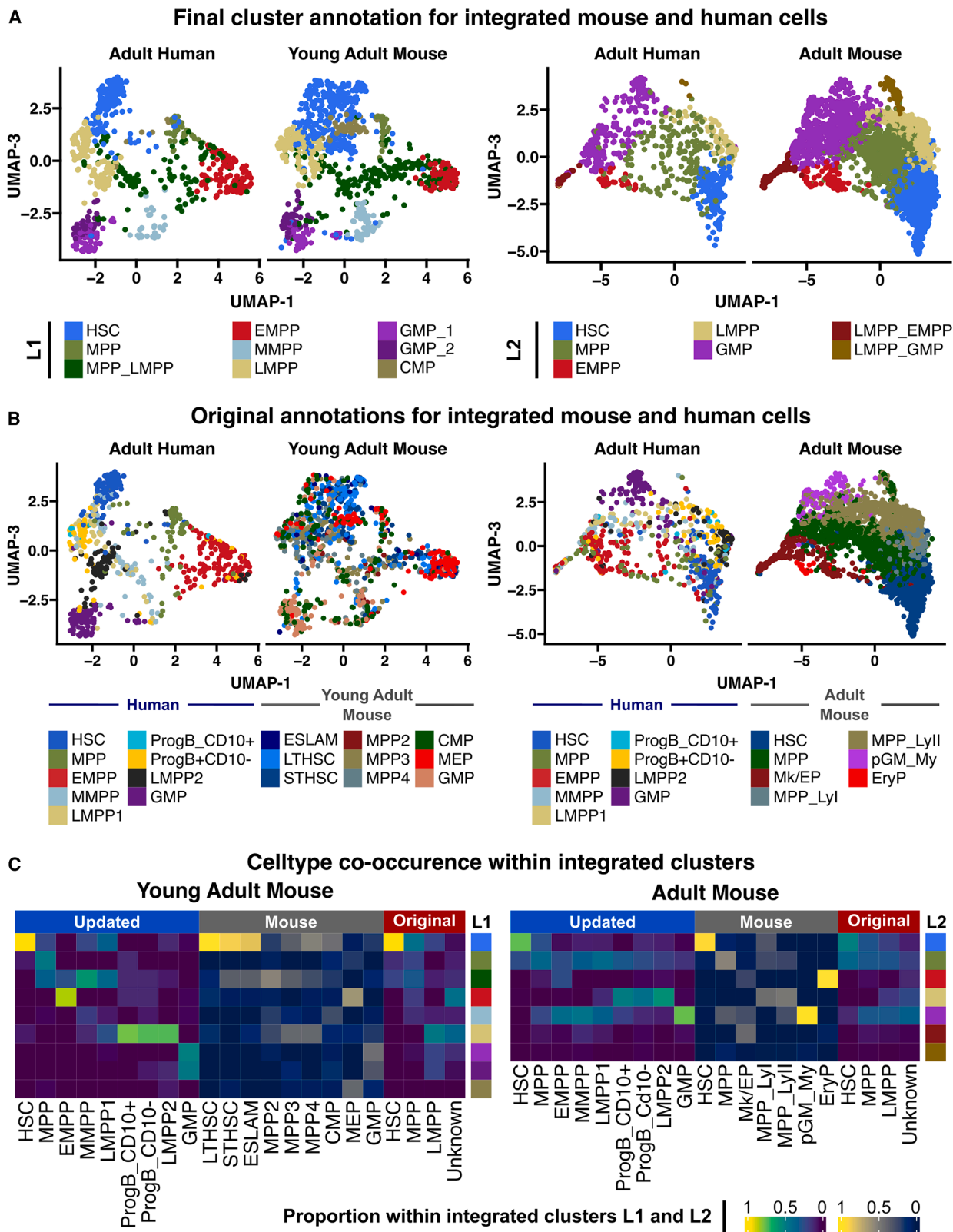
(C) Correlation between averaged normalized counts per cluster for top differentially expressed features in the GIM (scATAC-seq dataset) and GEM (scRNA-seq reference dataset).

(D) Putative and differential regulators between designated HSPCs. Positive values are up in the first cell type indicated in the title.

(E) Footprints for representative differential regulators.

(F) Summary of key differential regulators that define each specific HSPC. Each box contains significant regulators compared to the labeled HSPC. Entries in bold are present in multiple comparisons.

GIM, gene integration matrix; GEM, gene expression matrix.



(legend on next page)

per murine cell based on upregulated gene expression programs associated with differentiated human blood cells.<sup>74–76</sup> Although we observed the expected patterns of gene expression changes across murine cells, the differences in the primitive HSPC populations were gradual. We then annotated the murine cells using our mixed-model classifier (Figure S12B). In young adult mice, LT- and ST-HSCs were annotated as human HSCs; however, the ESLAM HSC and MPP2–3 annotations were more heterogeneous. Murine MPP4s appear to resemble human LMPP1s, murine MPP2s were mostly labeled as human MPPs and LMPP1s, and murine MPP3s were mostly labeled as LMPP1s. In adult mice, there was also a strong one-to-one association between HSCs. Murine granulocyte and monocyte progenitors (pGM\_Mys) and CLPs were predominantly labeled as GMPs and LMPPs, respectively; however, the assignments for the remaining cell types were more inconsistent (Figure S12B).

To improve our ability to identify homologous cell states, we subsequently performed an iterative integration of the murine and human data (Figure S11). We integrated each murine dataset with the human reference using CCA and identified conserved features between mouse and human HSPCs (Figure S13A; Table S11). The data were then re-integrated using only conserved features and sub-clustered to achieve a similar number of clusters to cell types present in the original murine dataset (Figure S13B). The final clusters, L1 and L2, were annotated based on marker genes and their resemblance to the human HSPC states (Figure 5A). For comparison, a separate projection displays the integrated cells based on the original species-specific annotation (Figure 5B).

We then performed a co-occurrence analysis within each cluster (L1 and L2) to determine homologous cell states. This analysis is illustrated in Figure 5C, which displays the proportion (column comparison) of human HSPCs that co-occur in the same integrated clusters (L1 and L2) as murine HPSCs (row comparison). We evaluated human HSPCs based on our updated sorting strategy (“updated”) and the classical sorting strategy (“original”).<sup>5,30</sup> In young adult mice, we again saw homology between HSCs and moderate homology between EMPP and MEP, ProB and MPP4, and GMP states. In adult mice, we also saw homology between HSC states and between human GMPs and mouse pGM\_My cells (Figure 5C). There was moderate homology between ProBs, LMPP2s, and MPP Ly/I cells. To determine conserved gene expression programs in hematopoiesis, we performed another milestone-based trajectory analysis. This analysis identified similar differentiation trajectories between species (Figures 6A–6H). For each cross-species analysis between young adult and adult mice, we identified key differentiation nodes comprising a heterogeneous population of MPPs and OPPs. The most dynamic features within these nodes that were also conserved across species and age were determined (Table S11) and are illustrated in Figure 6I. From this analysis, we observe that TFs like *GATA1*, *GATA2*, and *PBX1* and surface

markers like *CD34*, *FLT3*, and *CD74* are conserved dynamic features between mouse and human hematopoiesis. A summary of homologous cell states based on L1 and L2 clustering is illustrated in Figure 6J, and their differentially expressed genes are provided in Table S11. These observations indicate that while there are differences between HSPC populations in mouse and human hematopoiesis, L1 and L2 clustering can assist in identifying homologous states between species.

### HSPCs undergo cell-type-specific changes in gene expression during aging

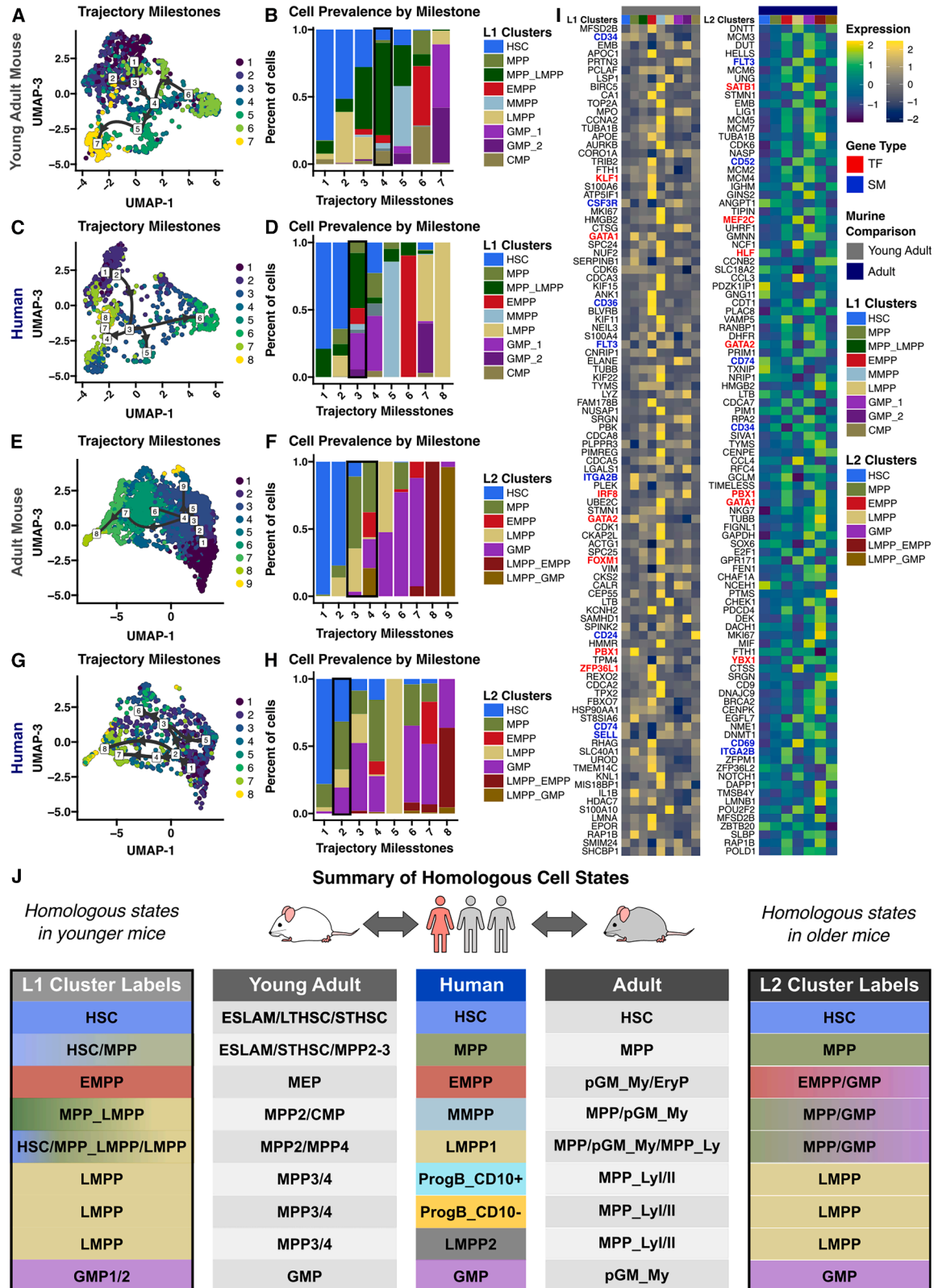
The long-term goal for updating the HSPC classification framework is to advance the study of human health and disease. Aging is a risk factor for numerous diseases and is characterized by multiple biological hallmarks.<sup>77</sup> In hematopoiesis, the relative contribution of these individual hallmarks to aging is unclear, as is the interconnection between these processes and the associated cell type. We hypothesized that by performing a cell-type-specific analysis of human aging, we could overlay a cell-specific context to derangements observed in aging. To address this hypothesis, we evaluated whether the updated HSPC reference can be used to identify gene signatures associated with human aging (Figure 7A). Unlike in mice, we cannot prospectively study aging human HSPCs. However, with the emergence of numerous scRNA-seq datasets, we can indirectly evaluate aging with greater resolution using primary human data. To perform this analysis, we first identified six single-cell datasets from human BM with age and cell-type annotations.<sup>32,35,36,55,56,59</sup> Each dataset was labeled using the mixed-model classifier, and after integration, we obtained 42,480 cells with high-quality annotations across 20 donors aged 0–75 (Figure 7B).

Cells were sub-sampled from the integrated dataset based on their annotations to perform a cell-type-specific analysis. We also sub-sampled a similar number of cells across the entire dataset to resemble a comparable bulk sample (“bulk”). The final cell-type-specific metrics are provided in Table S12. We first performed a weighted Spearman correlation comparing gene expression within each cell type to both increasing age (aging) and decreasing age (youth) to identify important genes associated with aging and youth. These genes were then evaluated for important biological processes using an enrichment analysis (Figure 7C). A cell-type-specific approach identified more significant terms compared to a similar analysis across all cell types (bulk; Figures 7C and 7D). In aging HSCs, we observe enrichment in terms associated with antigen processing and presentation, DNA damage and repair response, and RNA processing and splicing (Figure S14). In aging MPPs, we see enrichment in terms involved in TP53 signal transduction, RNA processing, and RNA splicing. In differentiated progenitors, including LMPP1s and GMPs, we observe enrichment in terms associated with inflammation, RNA processing, RNA splicing, DNA damage response and repair, and cell cycle. We were unable to identify

**Figure 5. Human MPPs and OPPs exist in homologous states with corresponding murine HSPCs**

(A and B) UMAPs of integrated human and mouse cells colored based on (A) updated cluster labels L1 and L2 of integrated cells and (B) original cell annotations for the mouse and human cells.

(C) Human and mouse HSPC co-occurrence plots. Each row is colored based on integrated cluster assignments L1 and L2 (see A), and each entry is colored based on the proportion (column comparison) of the original cluster within the updated cluster assignments.



(legend on next page)



terms associated with young age when analyzing all cell types together (bulk; Figure 7C); however, by using a cell-type-specific approach, we observe several cell-type-specific enrichments in terms associated with cell cycle, RNA biology, metabolism, and proteostasis (Figure S14). To further investigate the specific mechanisms within these broader categories, we focused on two pathways, RNA processing and splicing and DNA damage repair and processing, as they were enriched in multiple cell types. We curated gene signatures associated with each process (Table S13)<sup>78–86</sup> and calculated cell-type-specific modules based on these signatures. From this analysis, we observe that elderly HSCs and MPPs show a significant decrease in critical processes involved in DNA and mRNA biology, which may contribute to aging-related dysfunction (Figures 7E–7G).

Identifying important features that can predict individual cell age can be a useful tool for studying aging cells in single-cell datasets. To identify simpler cell-type-specific gene signatures compared to the correlation analysis, we performed recursive feature elimination (RFE) using random forest (RF) regression on a training dataset comprising 50% of the cells. The cell-type-specific approach identified simpler signatures that improved age predictions in the test dataset based on root-mean-square error (RMSE) and coefficient of determination ( $R^2$ ) (Figures 7D and 7H).

We then evaluated whether HSPCs in blood cancers, including acute myeloid leukemia (AML) and pediatric acute lymphoblastic leukemia (ALL), express an aging phenotype. Published single-cell datasets from patients with AML<sup>32</sup> and pediatric ALL<sup>87</sup> were labeled using our mixed-model classifier. Cell types based on our annotations were sub-sampled and integrated with their healthy counterparts (Figure 7B). We calculated cell-type-specific aging modules using the optimal features derived from our RFE analysis and compared them to healthy cells (Figure 7I). In both AML and pediatric ALL, which are aggressive myeloid and lymphoid blood cancers, LMPP1s expressed a higher LMPP1 aging score. There were too few ALL HSCs and GMPs for this analysis; however, HSC and GMP aging scores were not increased in AML. Prior work has shown that leukemia stem cell populations in AML reside in LMPP populations, as determined by their immunophenotype, and may express similar gene expression patterns.<sup>9</sup> Our results suggest that these cells may also express a more aged phenotype, which is unrelated to their chronological age (Table S12). Overall, these results support a cell-type-specific analysis for identifying gene expression programs that are relevant in human aging and cancer.

## DISCUSSION

In this study, we built a framework for characterizing adult human MPPs using multi-omic single-cell data from normal human BM and computationally identifying HSPC cell states with distinct

immunophenotypes (Figure 1), gene expression (Table S6), and chromatin accessibility (Figure 4). These cells were prospectively isolated using flow cytometry and demonstrated unique functional properties (Figure 2O). We show that within the classical MPP pool, only CLL1–CD69+ cells are capable of long-term engraftment and 7 lineage differentiation potentials, whereas CLL1–CD69– and CLL1+ cells are erythroid-primed and myeloid-primed MPPs, respectively, with highly diminished engraftment capacity. Within the classical LMPP sub-population, we show that CD2–CLL1– cells demonstrate high lymphoid potency, CD2+ cells produce both lymphoid and myeloid cells, and CLL1+ cells contain primarily GMPs. Megakaryocyte production is restricted primarily to MPPs, erythrocyte production is enriched in EMPPs, lymphoid potency is greatest in LMPPs, and all evaluated HSPCs can produce granulocytes and monocytes (Figure 2O). It is worth noting that while the sub-populations purified by the updated sorting strategy (Figures 1B and S5) did enrich for the target cluster of interest, the final gates were not homogenous. This suggests that HSPCs exist in a gradient of cell states, and our sorting strategy primarily enriches for cells with certain characteristics rather than purifying for a homogenous population (Figure S3). This is an inherent limitation, as the continuous nature of hematopoiesis is often not captured in discrete populations purified using FACS methods, which is common in both mouse and human studies involving HSPCs.<sup>13,17,51,73</sup>

Despite the limitations of our study, the updated annotation and functional characterization of human HSPCs allowed us to build a valuable reference for evaluating external human and mouse data, which led to several key observations. Importantly, cell-type-specific approaches improved cell annotation transfer (Figures S6 and S7), which facilitated the subsequent analysis of external single-cell datasets. We see that the cell states derived from our reference dataset are present in several external datasets and are analogous to available external dataset annotations (Figure S7). Of note, our ability to label cells in CB and mPB was highly diminished, which again highlights the underlying molecular differences between different hematopoietic tissues. Ultimately, our approach provides increased resolution for identifying HSPC states compared to published annotations.

Historically, human MPP populations have been less characterized compared to murine MPPs. This has led to difficulties in appropriately inferring the human relevance of observations made in mice. Such difficulties are in part due to mouse studies being performed predominantly in a single genetic strain, i.e., C57BL/6, whereas human studies include genetically distinct individuals. Regardless, our work provides insights into the heterogeneity within the human MPP sub-populations, and our cross-species analysis with murine HSPCs identifies homologous cell states between mice and humans. We see that murine

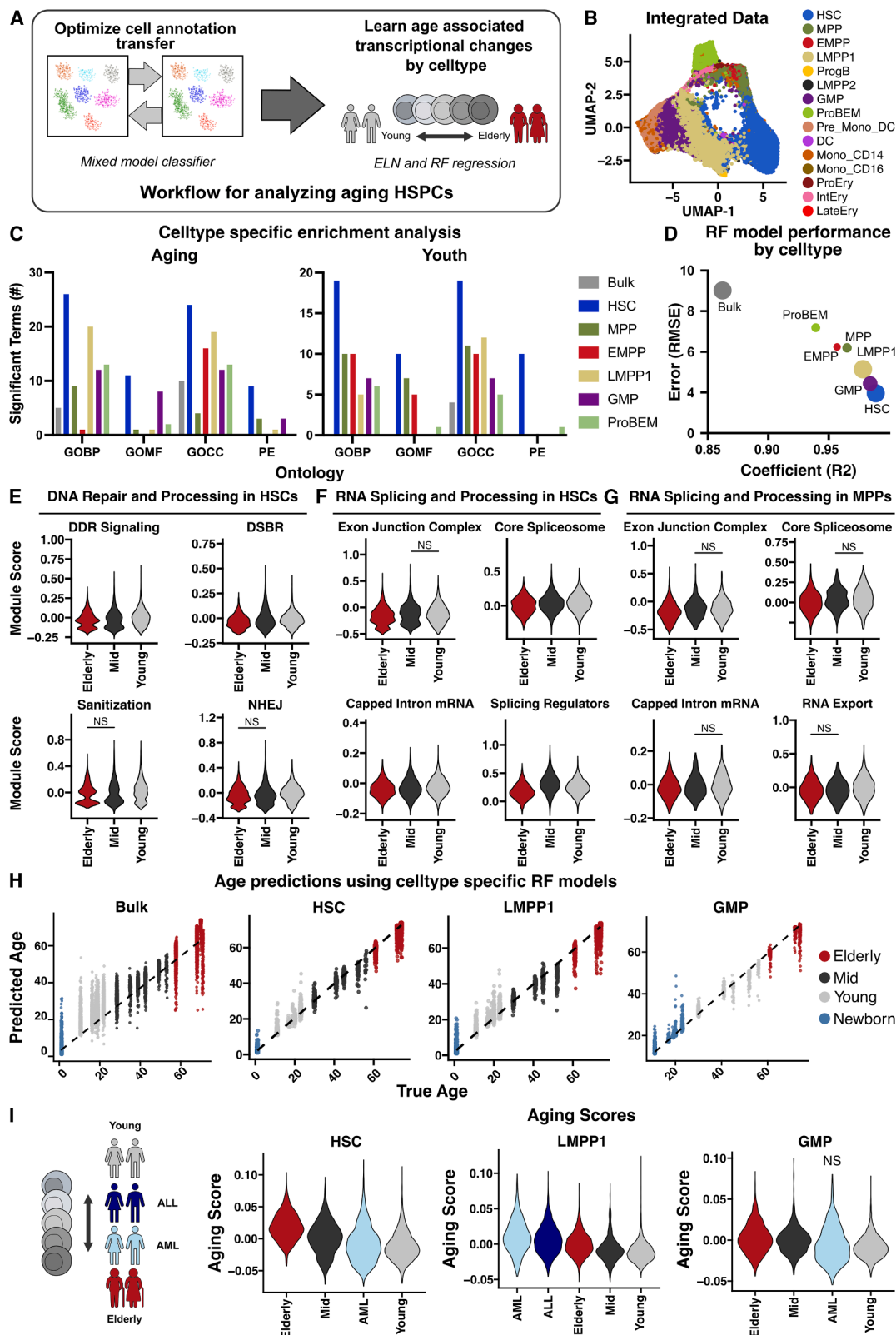
**Figure 6. Milestone analysis of integrated human and mouse hematopoiesis shows conserved gene expression programs across species** (A–H) UMAPs and cell-type prevalence of young adult mouse (A and B), adult mouse (E and F), and human (C, D, G, and H) cells categorized based on inferred developmental milestones.

(I) The expression of important features associated key differentiation nodes (outlined in black in B, D, F, and H) are illustrated in heatmaps for both cross-species comparisons.

(J) Two cell-type cluster references were derived based on homologous states between humans and young adult (L1) and older adult mice (L2).

SM, surface marker; TF, transcription factor.





(legend on next page)

HSCs and more differentiated progenitors show a stronger one-to-one concordance with their analogous human states compared to the MPPs and OPPs (Figures 5 and 6). Although these observations highlight the underlying complexity and heterogeneity within both mouse and human MPPs, they also provide a framework for connecting murine and human cell states, which is important since murine immunophenotypes may change significantly based on mouse age and genetic strain.<sup>18</sup>

Finally, we demonstrate the utility of our reference profile and mixed-model classifier for studying human disease by performing a meta-analysis across 42,480 human HSPCs aged 0–75. Using a cell-type-specific approach, we were able to nominate important biological processes in human aging and identified cell-type-specific associations for these derangements (Figure 7). Additionally, our approach allowed us to identify cell-type-specific aging scores that revealed interesting observations between the aging phenotype and blood cancers (Figure 7I), as AML and ALL LMPP1s, rather than HSCs, express a significantly higher aging score compared to healthy donors, even in pediatric cases where the average age of the donor was 5.2 years. We summarized the observations from this study into an updated model of adult hematopoiesis (Figure 3), which highlights the heterogeneity within the human MPP and OPP sub-populations and their associated functional capacities. Overall, our findings and methodologies provide important insights into adult MPP heterogeneity and a valuable resource for future research in hematopoiesis.

### Limitations of the study

This study provides significant advancements in our understanding of adult human hematopoiesis, as we observe that cell states determined primarily by gene expression can be enriched within immunophenotypically defined sub-populations. It was therefore possible to use these surface markers to purify cells, although heterogeneous, to study their functional characteristics. However, our analysis was limited by the scarcity of primary adult human tissue and the diminished engraftment potential of adult BM compared to CB.<sup>22–25,57,88</sup> In fact, adult BM cells have not been reported to reliably produce multilineage engraftment over serial transplantations in immunodeficient mice. Moreover, we had to evaluate populations of cells, rather than single cells, for our functional assays. This, of course, impairs our ability to definitively determine if our purified cells exist as a homogenous population or as a collection of cells with various functional capabilities, even though this approach is consistent

with historical advancements and refinements in HSPC immunophenotyping. Further, we were unable to show human cell engraftment in secondary transplants for either HSCs or CLL1–CD69+ MPPs. Therefore, we did not relabel the MPPs as HSCs since positive multilineage engraftment over serial transplantations is required to define HSCs. Regardless of these limitations, we were able to observe significant differences among the purified cells, which will be informative when designing and implementing translational research in human hematopoiesis. Future work will need to address the shortcomings inherent in our functional assays and continue to use complementary cell sources, i.e., CB and mPB, and murine models. Although the reference profile provides an important framework for studying hematopoiesis, our computational analyses only nominate potential relevant biological processes. Future work will need to expand on these observations with the necessary mechanistic and functional correlates.

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Ravindra Majeti ([rmajeti@stanford.edu](mailto:rmajeti@stanford.edu)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- Single-cell sequencing data have been deposited at GEO and are publicly available as of the date of publication. The accession numbers are listed in the [key resources table](#).
- All original code used in this work is publicly available as of the date of the publication. The link is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

### ACKNOWLEDGMENTS

We would like to thank Feifei Zhao for lab management and all members of the Majeti and Gentles labs for supporting this study. This work was supported by the National Institutes of Health (NIH) under awards 1R01HL142637 and 1R01CA251331 and the Stanford Ludwig Center for Cancer Stem Cell Research and Medicine (R.M.). A.J.G. was supported by the National Cancer Institute (NCI) of the NIH under awards R21CA238971, U01CA264611, and R01CA276828. A.E. was supported by the NCI under award F32CA250304, the Advanced Residency Training Program at Stanford, and the American Society of Hematology (ASH) Scholar Award. A.C.F. was supported by a Stanford Graduate Fellowship, NSF Graduate Research Fellowship Program, and Stanford Lieberman Fellowship. T.K. was supported by the Leukemia and

### Figure 7. HSPCs demonstrate unique cell-type-specific transcriptional changes during aging

(A) Workflow for analysis of aging HSPCs.  
(B) UMAP of integrated HSPCs with confident cell annotations.  
(C) Enrichment analysis was performed using informative genes identified from cell-type-specific Spearman correlation analysis for aging and youth. Bar plots show the number of significant terms for each cell-type-specific analysis.  
(D) RF model performance metrics after RFE for predicting cell-type-specific age on the test data.  
(E–G) Expression of gene modules associated with DNA repair and processing in HSCs (E) and RNA splicing and processing in HSCs (F) and MPPs (G).  
(H) Correlation between the predicted age from the RF model after RFE and true age on the test data.  
(I) Cell-type-specific aging scores for healthy and cancer cells annotated using the mixed-model classifier.  
Statistical significance was tested using a t test with multiple comparison adjustments. All scores are significantly different unless marked not significant (NS). DDR, DNA damage response; DSB, double-stranded break repair; ELN, elastic net; GO, Gene Ontology; GOBP, GO biological processes; GOCC, GO cellular component; GOMF, GO molecular functions; NHEJ, non-homologous end joining; PE, pathway enrichment (Reactome); RF, random forest; RFE, recursive feature elimination.

Lymphoma Society Special Fellow Award. X.H. was supported by the ASH Medical Student Physician Scientist Award. B.A.L. was supported by the NIH under award K99CA276901. Y.K. was supported by Stanford Immunology Baker Fellowship and by a KFAS Overseas PhD Scholarship from the Korea Foundation for Advanced Studies. M.H.L. was supported by the Blavatnik Family Fellowship. Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## AUTHOR CONTRIBUTIONS

Conceptualization, A.E. and R.M.; methodology, A.E., A.J.G., A.M.N., and R.M.; investigation, A.E., Y.N., A.C.F., X.H., S.R., M.N., and A.A.; writing – original draft, A.E. and R.M.; writing – review & editing, A.E., A.J.G., A.M.N., and R.M.; funding acquisition, R.M.; supervision, A.E., A.C.F., T.K., B.A.L., Y.K., D.K., M.H.L., A.J.G., A.M.N., and R.M.

## DECLARATION OF INTERESTS

R.M. is on the advisory boards of Kodikaz Therapeutic Solutions, Orbital Therapeutics, Pheast Therapeutics, 858 Therapeutics, Prelude Therapeutics, Mubadala Capital, and Aculeus Therapeutics. R.M. is a co-founder and equity holder of Pheast Therapeutics, MyeloGene, and Orbital Therapeutics. S.R. and M.N. previously worked at BD Biosciences.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Description of healthy adult donors
  - Animal care
- **METHOD DETAILS**
  - A multi-modal framework for studying adult human HSPCs
  - Single cell captures for whole transcriptome and surface marker quantification
  - Sample processing, integration, and quality control
  - Clustering and diversity analysis
  - NMF and Louvain clustering
  - Trajectory analysis
  - Surface marker identification
  - Fluorescence-activated cell sorting
  - Single cell colony formation assay
  - Megakaryocyte expansion assay
  - Xenotransplantation assays
  - OP9 and OP9-hDL4 maintenance
  - OP9 co-culture lymphoid-myeloid cell differentiation assay
  - OP9-hDL4 co-culture T cell differentiation assay
  - Basophil, eosinophil and mast cell expansion assay
  - Integrated single cell ATAC sequencing analysis
  - Deriving a mixed model classifier for cell label transfer
  - Cross-species analysis
  - Aging analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2025.116236>.

Received: March 22, 2024  
Revised: February 19, 2025

Accepted: August 12, 2025  
Published: September 3, 2025

## REFERENCES

1. Eaves, C.J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. *Blood* 125, 2605–2613.
2. Weissman, I.L., and Shizuru, J.A. (2008). The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases. *Blood* 112, 3543–3553.
3. Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. *Cell Stem Cell* 10, 120–136.
4. Baum, C.M., Weissman, I.L., Tsukamoto, A.S., Buckle, A.-M., and Peault, B. (1992). Isolation of a candidate human hematopoietic stem-cell population. *Proc. Natl. Acad. Sci. USA* 89, 2804–2808.
5. Majeti, R., Park, C.Y., and Weissman, I.L. (2007). Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. *Cell Stem Cell* 1, 635–645.
6. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., et al. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat. Immunol.* 19, 85–97.
7. Watcham, S., Kucinski, I., and Gottgens, B. (2019). New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing. *Blood* 133, 1415–1426.
8. Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. *Nature* 553, 418–426.
9. Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R., et al. (2011). Coexistence of LMPP-like and GMP-like Leukemia Stem Cells in Acute Myeloid Leukemia. *Cancer Cell* 19, 138–152.
10. Manz, M.G., Miyamoto, T., Akashi, K., and Weissman, I.L. (2002). Prospective isolation of human clonogenic common myeloid progenitors. *Proc. Natl. Acad. Sci. USA* 99, 11872–11877.
11. Galy, A., Travis, M., Cen, D., Chen, B., and Human, T. (1995). B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. *Immunity* 3, 459–473.
12. Belluschi, S., Calderbank, E.F., Ciauro, V., Pijuan-Sala, B., Santoro, A., Mende, N., Diamanti, E., Sham, K.Y.C., Wang, X., Lau, W.W.Y., et al. (2018). Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat. Commun.* 9, 4100.
13. Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., et al. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* 351, aab2116.
14. Velten, L., Haas, S.F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B.P., Hirche, C., Lutz, C., Buss, E.C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* 19, 271–281.
15. Paul, F., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., et al. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* 163, 1663–1677.
16. Zhang, Y., Xie, X., Huang, Y., Liu, M., Li, Q., Luo, J., He, Y., Yin, X., Ma, S., Cao, W., et al. (2022). Temporal molecular program of human hematopoietic stem and progenitor cells after birth. *Dev. Cell* 57, 2745–2760.e6.
17. Pietras, E.M., Reynaud, D., Kang, Y.A., Carlin, D., Calero-Nieto, F.J., Leavitt, A.D., Stuart, J.M., Göttgens, B., and Passegué, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* 17, 35–46.

18. Challen, G.A., Pietras, E.M., Wallscheid, N.C., and Signer, R.A.J. (2021). Simplified murine multipotent progenitor isolation scheme: Establishing a consensus approach for multipotent progenitor identification. *Exp. Hematol.* **104**, 55–63.
19. Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D.B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., et al. (2014). Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522.
20. Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A.M., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306.
21. Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593.
22. da Silva, C.L., Gonçalves, R., Porada, C.D., Ascensão, J.L., Zanjani, E.D., Cabral, J.M.S., and Almeida-Porada, G. (2009). Differences amid bone marrow and cord blood hematopoietic stem/progenitor cell division kinetics. *J. Cell. Physiol.* **220**, 102–111.
23. Kim, D.K., Fujiki, Y., Fukushima, T., Ema, H., Shibuya, A., and Nakauchi, H. (1999). Comparison of Hematopoietic Activities of Human Bone Marrow and Umbilical Cord Blood CD34 Positive and Negative Cells. *Stem Cells* **17**, 286–294.
24. Theunissen, K., and Verfaillie, C.M. (2005). A multifactorial analysis of umbilical cord blood, adult bone marrow and mobilized peripheral blood progenitors using the improved ML-IC assay. *Exp. Hematol.* **33**, 165–172.
25. Wang, J.C.Y., Doedens, M., and Dick, J.E. (1997). Primitive Human Hematopoietic Cells Are Enriched in Cord Blood Compared With Adult Bone Marrow or Mobilized Peripheral Blood as Measured by the Quantitative In Vivo SCID-Repopulating Cell Assay. *Blood* **89**, 3919–3924.
26. Haas, S. (2020). Hematopoietic Stem Cells in Health and Disease—Insights from Single-Cell Multi-omic Approaches. *Curr. Stem Cell Rep.* **6**, 67–76.
27. Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216.
28. Edir Wickrema, A., Gentles, A.J., and Majeti, R. (2023). Single-cell genomics in AML: extending the frontiers of AML research. *Blood* **141**, 345–355.
29. Konturek-Ciesla, A., Dhapola, P., Zhang, Q., Säwén, P., Wan, H., Karlsson, G., and Bryder, D. (2023). Temporal multimodal single-cell profiling of native hematopoiesis illuminates altered differentiation trajectories with age. *Cell Rep.* **42**, 112304.
30. Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203.
31. Granja, J.M., Klemm, S., McGinnis, L.M., Kathiria, A.S., Mezger, A., Corces, M.R., Parks, B., Gars, E., Liedtke, M., Zheng, G.X.Y., et al. (2019). Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465.
32. van Galen, P., Hovestadt, V., Wadsworth II, M.H., Hughes, T.K., Griffin, G.K., Battaglia, S., Verga, J.A., Stephansky, J., Pastika, T.J., Lombardi Story, J., et al. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265–1281.e24.
33. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. *eLife* **6**, e27041.
34. Setty, M., Kisieliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460.
35. Triana, S., Vonficht, D., Jopp-Saile, L., Raffel, S., Lutz, R., Leonce, D., Antes, M., Hernández-Malmierca, P., Ordoñez-Rueda, D., Ramasz, B., et al. (2021). Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589.
36. Ainciburu, M., Ezponda, T., Berastegui, N., Alfonso-Pierola, A., Vilas-Zornoza, A., San Martín-Uriz, P., Alignani, D., Lamo-Espinosa, J., San-Julian, M., Jiménez-Solas, T., et al. (2023). Uncovering perturbations in human hematopoiesis associated with healthy aging and myeloid malignancies at single cell resolution. *eLife* **12**, e79363.
37. Zhang, X., Song, B., Carlino, M.J., Li, G., Ferchen, K., Chen, M., Thompson, E.N., Kain, B.N., Schnell, D., Thakkar, K., et al. (2024). An immunophenotype-coupled transcriptomic atlas of human hematopoietic progenitors. *Nat. Immunol.* **25**, 703–715.
38. Lasry, A., Nadorp, B., Fornerod, M., Nicolet, D., Wu, H., Walker, C.J., Sun, Z., Witkowski, M.T., Tikhonova, A.N., Guillaumot-Ruano, M., et al. (2023). An inflammatory state remodels the immune microenvironment and improves risk stratification in acute myeloid leukemia. *Nat. Cancer* **4**, 27–42.
39. Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. Roy. Stat. Soc. B* **63**, 411–423.
40. Waltman, L., and van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471.
41. Luca, B.A., Steen, C.B., Matusiak, M., Azizi, A., Varma, S., Zhu, C., Przybyl, J., Espín-Pérez, A., Diehn, M., Alizadeh, A.A., et al. (2021). Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* **184**, 5482–5496.e28.
42. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinf.* **11**, 367.
43. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782.
44. Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. *Nat. Methods* **16**, 715–721.
45. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., and Theis, F.J. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50.
46. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296.
47. Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691.
48. Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554.
49. Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59.
50. Bendall, S.C., Davis, K.L., Amir, E.A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014).



Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157, 714–725.

51. Notta, F., Doulatov, S., Laurenti, E., Poepl, A., Jurisica, I., and Dick, J.E. (2011). Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. *Science* 333, 218–221.
52. Nakauchi, Y., Ediriwickrema, A., Martinez-Krams, D., Zhao, F., Rangavajhula, A., Karigane, D., and Majeti, R. (2023). Simplified Intrafemoral Injections Using Live Mice Allow for Continuous Bone Marrow Analysis. *JoVE J* 201, e65874. <https://doi.org/10.3791/65874>.
53. Mohtashami, M., Shah, D.K., Nakase, H., Kianizad, K., Petrie, H.T., and Zúñiga-Pflücker, J.C. (2010). Direct comparison of Dll1- and Dll4-mediated Notch activation levels shows differential lymphomyeloid lineage commitment outcomes. *J. Immunol.* 185, 867–876.
54. Abdelaal, T., Michielsen, L., Cats, D., Hoogduin, D., Mei, H., Reinders, M. J.T., and Mahfouz, A. (2019). A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol.* 20, 194.
55. Bunis, D.G., Bronevetsky, Y., Krow-Lucal, E., Bhakta, N.R., Kim, C.C., Nerella, S., Jones, N., Mendoza, V.F., Bryson, Y.J., Gern, J.E., et al. (2021). Single-Cell Mapping of Progressive Fetal-to-Adult Transition in Human Naive T Cells. *Cell Rep.* 34, 108573.
56. Griffin, G.K., Booth, C.A.G., Togami, K., Chung, S.S., Ssozi, D., Verga, J. A., Bouyssou, J.M., Lee, Y.S., Shanmugam, V., Hornick, J.L., et al. (2023). Ultraviolet radiation shapes dendritic cell leukaemia transformation in the skin. *Nature* 618, 834–841.
57. Mende, N., Bastos, H.P., Santoro, A., Mahbubani, K.T., Ciaurro, V., Calderbank, E.F., Quiroga Londoño, M., Sham, K., Mantica, G., Morishima, T., et al. (2022). Unique molecular and functional features of extramedullary hematopoietic stem and progenitor cell reservoirs in humans. *Blood* 139, 3387–3401.
58. Oetjen, K.A., Lindblad, K.E., Goswami, M., Gui, G., Dagur, P.K., Lai, C., Dillon, L.W., McCoy, J.P., and Hourigan, C.S. (2018). Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight* 3, e124928.
59. Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., Meira, A.R., et al. (2021). Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Rep.* 36, 109698.
60. Thongon, N., Ma, F., Santoni, A., Marchesini, M., Fiorini, E., Rose, A., Adema, V., Ganan-Gomez, I., Groarke, E.M., Gutierrez-Rodriguez, F., et al. (2021). Hematopoiesis under telomere attrition at the single-cell resolution. *Nat. Commun.* 12, 6850.
61. Vanuytsel, K., Villacorta-Martin, C., Lindstrom-Vautrin, J., Wang, Z., Garcia-Beltran, W.F., Vrbanc, V., Parsons, D., Lam, E.C., Matte, T.M., Dowrey, T.W., et al. (2022). Multi-modal profiling of human fetal liver hematopoietic stem cells reveals the molecular signature of engraftment. *Nat. Commun.* 13, 1103.
62. Zheng, S., Papalexi, E., Butler, A., Stephenson, W., and Satija, R. (2018). Molecular transitions in early progenitors during human cord blood hematopoiesis. *Mol. Syst. Biol.* 14, e8041.
63. Dussiau, C., Boussaroque, A., Gaillard, M., Bravetti, C., Zaroili, L., Knosp, C., Friedrich, C., Asquier, P., Willems, L., Quint, L., et al. (2022). Hematopoietic differentiation is characterized by a transient peak of entropy at a single-cell level. *BMC Biol.* 20, 60.
64. Kaufmann, K.B., Zeng, A.G.X., Coyaud, E., Garcia-Prat, L., Papalexi, E., Murison, A., Laurent, E.M.N., Chan-Seng-Yue, M., Gan, O.I., Pan, K., et al. (2021). A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness. *Nat. Immunol.* 22, 723–734.
65. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132, 631–644.
66. Schütte, J., Wang, H., Antoniou, S., Jarratt, A., Wilson, N.K., Riepsaame, J., Calero-Nieto, F.J., Moignard, V., Basilico, S., Kinston, S.J., et al. (2016). An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *eLife* 5, e11469.
67. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W. N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. *Cell* 144, 296–309.
68. Beck, D., Thoms, J.A.I., Perera, D., Schütte, J., Unnikrishnan, A., Knezevic, K., Kinston, S.J., Wilson, N.K., O'Brien, T.A., Göttgens, B., et al. (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood* 122, e12–e22.
69. Luecken, M., Burkhardt, D., Cannoodt, R., and et al. (2021). In A Sandbox for Prediction and Integration of DNA, RNA, and Proteins in Single Cells, 1, J. Vanschoren and S. Yeung, eds..
70. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* 53, 403–411.
71. Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. *Nature* 583, 729–736.
72. Kent, D.G., Copley, M.R., Benz, C., Wöhler, S., Dykstra, B.J., Ma, E., Cheyne, J., Zhao, Y., Bowie, M.B., Zhao, Y., et al. (2009). Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood* 113, 6342–6350.
73. Nestorowa, S., Hamey, F.K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N.K., Kent, D.G., and Göttgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 128, e20–e31.
74. Nakauchi, Y., Azizi, A., Thomas, D., Corces, M.R., Reinisch, A., Sharma, R., Cruz Hernandez, D., Köhnke, T., Karigane, D., Fan, A., et al. (2022). The cell type specific 5hmC landscape and dynamics of healthy human hematopoiesis and TET2-mutant pre-leukemia. *Blood Cancer Discov.* 3, 346–367.
75. Psaila, B., Wang, G., Rodriguez-Meira, A., Li, R., Heuston, E.F., Murphy, L., Yee, D., Hitchcock, I.S., Sousos, N., O'Sullivan, J., et al. (2020). Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets. *Mol. Cell* 78, 477–492.e8.
76. Aschenbrenner, A.C., Mouktaroudi, M., Krämer, B., Oestreich, M., Antonakos, N., Nuesch-Germano, M., Gkizeli, K., Bonaguro, L., Reusch, N., Baßler, K., et al. (2021). Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med.* 13, 7.
77. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The Hallmarks of Aging. *Cell* 153, 1194–1217.
78. Will, C.L., and Lührmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harbor Perspect. Biol.* 3, a003707.
79. Le Hir, H., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* 17, 41–54.
80. Köhler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nat. Rev. Mol. Cell Biol.* 8, 761–773.
81. Busch, A., and Hertel, K.J. (2012). Evolution of SR protein and hnRNP splicing regulatory factors. *WIREs RNA* 3, 1–12.
82. Björk, P., and Wieslander, L. (2017). Integration of mRNP formation and export. *Cell. Mol. Life Sci.* 74, 2875–2897.
83. Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D. (2016). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* 26, 52–64.
84. Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506.



85. Ciccio, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* 40, 179–204.
86. Gad, H., Koolmeister, T., Jemth, A.-S., Eshtad, S., Jacques, S.A., Ström, C.E., Svensson, L.M., Schultz, N., Lundbäck, T., Einarsdottir, B.O., et al. (2014). MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* 508, 215–221.
87. Caron, M., St-Onge, P., Sontag, T., Wang, Y.C., Richer, C., Ragoussis, I., Sinnett, D., and Bourque, G. (2020). Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. *Sci. Rep.* 10, 8079.
88. Holyoake, T.L., Nicolini, F.E., and Eaves, C.J. (1999). Functional differences between transplantable human hematopoietic stem cells from fetal liver, cord blood, and adult marrow. *Exp. Hematol.* 27, 1418–1427.
89. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
90. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902.e21.
91. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
92. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* 28, 1–26.
93. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 8, 329–337.e4.
94. Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* 9, giaa151.
95. Benaglia, T., Chauveau, D., Hunter, D.R., and Young, D.S. (2009). mixtools: An R Package for Analyzing Mixture Models. *J. Stat. Software* 32, 1–29.
96. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141.
97. Lun, A. (2024). bluster: Clustering Algorithms for Bioconductor, R package version 1.18.0. Bioconductor, <https://bioconductor.org/packages/bluster>. <https://doi.org/10.18129/B9.bioc.bluster>.
98. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.
99. Erickson, J.R., Mair, F., Bugos, G., Martin, J., Tyznik, A.J., Nakamoto, M., Mortimer, S., and Prlic, M. (2020). AbSeq protocol using the nano-well cartridge-based rhapsody platform to generate protein and transcript expression data on the single-cell level. *STAR Protoc.* 1, 100092.
100. Sun, B., Bugarin-Estrada, E., Overend, L.E., Walker, C.E., Tucci, F.A., and Bashford-Rogers, R.J.M. (2021). Double-jeopardy: scRNA-seq doublet/multiplet detection using multi-omic profiling. *Cell Rep. Methods* 1, 100008.
101. Burel, J.G., Pomaznoy, M., Lindestam Arlehamn, C.S., Weiskopf, D., da Silva Antunes, R., Jung, Y., Babor, M., Schulten, V., Seumois, G., Greenbaum, J.A., et al. (2019). Circulating T cell-monocyte complexes are markers of immune perturbations. *eLife* 8, e46045.
102. Joly, E., and Hudrisier, D. (2003). What is trogocytosis and what is its purpose? *Nat. Immunol.* 4, 815.
103. Mulè, M.P., Martins, A.J., and Tsang, J.S. (2022). Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nat. Commun.* 13, 2099.
104. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* 367, 405–411.
105. Dixon, P. (2003). VEGAN, A Package of R Functions for Community Ecology. *J. Veg. Sci.* 14, 927–930.
106. de Kanter, J.K., Lijnzaad, P., Candelli, T., Margaritis, T., and Holstege, F. C.P. (2019). CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. *Nucleic Acids Res.* 47, e95.
107. Boufe, K., Seth, S., and Batada, N.N. (2020). scID Uses Discriminant Analysis to Identify Transcriptionally Equivalent Cell Types across Single-Cell RNA-Seq Data with Batch Effect. *iScience* 23, 100914.
108. Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.
109. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 20, 163–172.
110. Tan, Y., and Cahan, P. (2019). SingleCellNet: A Computational Tool to Classify Single Cell RNA-Seq Data Across Platforms and Across Species. *Cell Syst.* 9, 207–213.e2.
111. Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. *Nat. Methods* 15, 359–362.
112. Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297.
113. Alquicira-Hernandez, J., Sathe, A., Ji, H.P., Nguyen, Q., and Powell, J.E. (2019). scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* 20, 264.
114. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
CD3	BD Biosciences	Cat#555341; RRID: AB_395747
CD4	BD Biosciences	Cat#555348; RRID: AB_395753
CD7	BD Biosciences	Cat#555362; RRID: AB_395765
CD8	BD Biosciences	Cat#555368; RRID: AB_395771
CD11b	BD Biosciences	Cat#555389; RRID: AB_395790
CD14	BD Biosciences	Cat#340585; RRID: AB_400065
CD16	BD Biosciences	Cat#555408; RRID: AB_395808
CD19	BD Biosciences	Cat#555414; RRID: AB_395814
CD20	BD Biosciences	Cat#555624; RRID: AB_395990
CD56	BD Biosciences	Cat#555517; RRID: AB_395907
GPA (CD235a)	BD Biosciences	Cat#559944; RRID: AB_397387
CD38	BD Biosciences	Cat#355790; RRID: AB_399969
CD90	Biolegend	Cat#328108; RRID: AB_893429
CD123	Biolegend	Cat#306006; RRID: AB_314580
CD34	BD Biosciences	Cat#340667; RRID: AB_400531)
CD10	Biolegend	Cat#312212; RRID: AB_2146550
CD45RA	Biolegend	Cat#304134; RRID: AB_2563814
CD45	BD Biosciences	Cat#560367; RRID: AB_1645573
Anti-mouse CD45.1	eBioscience	Cat#11-0453-85; RRID: AB_465059
Anti-mouse TER-119	eBioscience	Cat#47-5921-82; RRID: AB_1548786
CD33	BD Biosciences	Cat#555450; RRID: AB_395843
CD19	BD Biosciences	Cat#340667; RRID: AB_400531
HLA-ABC	BD Biosciences	Cat#568004; RRID: AB_2916810
CD49f	BD Biosciences	Cat#563271; RRID: AB_2738110
CD110	Biolegend	Cat#393807; RRID: AB_2750424
CD71	Biolegend	Cat#334109; RRID: AB_2563116
CD41a	BD Biosciences	Cat#559777; RRID: AB_398671
CD61	BD Biosciences	Cat#561912; RRID: AB_10892635
CD56	Biolegend	Cat#362533; RRID: AB_2565632
CD45	BD Biosciences	Cat#560367; RRID: AB_1645573
CD3	BD Biosciences	Cat#566518; RRID: AB_2744378
CD4	BD Biosciences	Cat#347327; RRID: AB_400283
CD5	eBioscience	Cat#25-0059-42; RRID: AB_1582282
CD7	BD Biosciences	Cat#555362; RRID: AB_395765
CD8	Biolegend	Cat#344713; RRID: AB_2044005
CD34	Biolegend	Cat#343510; RRID: AB_1877153
CD56	Biolegend	Cat#362533; RRID: AB_2565632
CD45	BD Biosciences	Cat#560367; RRID: AB_1645573
CD15	BD Biosciences	Cat#551376; RRID: AB_398501
CD14	BD Biosciences	Cat#564054; RRID: AB_2687593
CD19	Biolegend	Cat#363034; RRID: AB_2616936
CD34	Biolegend	Cat#343510; RRID: AB_1877153
CD193	BD Biosciences	Cat#561746; RRID: AB_10896970

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CD117	BD Biosciences	Cat#563859; RRID: AB_2738453
FCER1	eBioscience	Cat#48-5899-42; RRID: AB_2574088
CD34	BD Biosciences	Cat#745070; RRID: AB_2742689
CD45RA	Biolegend	Cat#304134; RRID: AB_2563814
CD66b	Biolegend	Cat#305125; RRID: AB_2750183
CD14	BD Biosciences	Cat#564054; RRID: AB_2687593
CD10 AbSeq Oligo AHS0051	BD Biosciences	Cat# 940045; RRID: AB_2875936
CD117 AbSeq Oligo AHS0064	BD Biosciences	Cat# 940051; RRID: AB_2875942
CD11b AbSeq Oligo AHS0005	BD Biosciences	Cat# 940008; RRID: AB_2875899
CD123 AbSeq Oligo AHS0020	BD Biosciences	Cat# 940020; RRID: AB_2875911
CD14 AbSeq Oligo AHS0037	BD Biosciences	Cat# 940005; RRID: AB_2875896
CD16 AbSeq Oligo AHS0053	BD Biosciences	Cat# 940006; RRID: AB_2875897
CD184 AbSeq Oligo AHS0060	BD Biosciences	Cat# 940056; RRID: AB_2875947
CD19 AbSeq Oligo AHS0030	BD Biosciences	Cat# 940004; RRID: AB_2875895
CD196 AbSeq Oligo AHS0034	BD Biosciences	Cat# 940033; RRID: AB_2875924
CD2 AbSeq Oligo AHS0029	BD Biosciences	Cat# 940046; RRID: AB_2875937
CD20 AbSeq Oligo AHS0008	BD Biosciences	Cat# 940016; RRID: AB_2875907
CD235a_b AbSeq Oligo AHS0048	BD Biosciences	Cat# 940040; RRID: AB_2875931
CD24 AbSeq Oligo AHS0042	BD Biosciences	Cat# 940028; RRID: AB_2875919
CD244 AbSeq Oligo AHS0131	BD Biosciences	Cat# 940362; RRID: AB_2876232
CD25 AbSeq Oligo AHS0026	BD Biosciences	Cat# 940009; RRID: AB_2875900
CD3 AbSeq Oligo AHS0033	BD Biosciences	Cat# 940000; RRID: AB_2875891
CD32 AbSeq Oligo AHS0073	BD Biosciences	Cat# 940069; RRID: AB_2875960
CD33 AbSeq Oligo AHS0044	BD Biosciences	Cat# 940031; RRID: AB_2875922
CD34 AbSeq Oligo AHS0061	BD Biosciences	Cat# 940021; RRID: AB_2875912
CD38 AbSeq Oligo AHS0022	BD Biosciences	Cat# 940013; RRID: AB_2875904
CD4 AbSeq Oligo AHS0032	BD Biosciences	Cat# 940001; RRID: AB_2875892
CD44 AbSeq Oligo AHS0054	BD Biosciences	Cat# 940039; RRID: AB_2875930
CD45 AbSeq Oligo AHS0040	BD Biosciences	Cat# 940002; RRID: AB_2875893
CD45RA AbSeq Oligo AHS0009	BD Biosciences	Cat# 940011; RRID: AB_2875902
CD56 AbSeq Oligo AHS0019	BD Biosciences	Cat# 940007; RRID: AB_2875898
CD69 AbSeq Oligo AHS0010	BD Biosciences	Cat# 940019; RRID: AB_2875910
CD7 AbSeq Oligo AHS0043	BD Biosciences	Cat# 940029; RRID: AB_2875920
CD8 AbSeq Oligo AHS0027	BD Biosciences	Cat# 940003; RRID: AB_2875894
CD86 AbSeq Oligo AHS0057	BD Biosciences	Cat# 940025; RRID: AB_2875916
CD9 AbSeq Oligo AHS0082	BD Biosciences	Cat# 940078; RRID: AB_2875969
CD90 AbSeq Oligo AHS0045	BD Biosciences	Cat# 940032; RRID: AB_2875923
CD93 AbSeq Oligo AHS0124	BD Biosciences	Cat# 940215; RRID: AB_2876097
CD99 AbSeq Oligo AHS0123	BD Biosciences	Cat# 940214; RRID: AB_2876096
CLL1 AbSeq Oligo AHS0097	BD Biosciences	Cat# 940212; RRID: AB_2876094
Tim3 AbSeq Oligo AHS0016	BD Biosciences	Cat# 940066; RRID: AB_2875957
<b>Critical commercial assays</b>		
BD Rhapsody Cartridge reagent kit	BD Biosciences	Cat# 633731
<b>Deposited data</b>		
Human BMMC scRNA-seq and scADT-seq	This paper	GEO: GSE262440
<b>Software and algorithms</b>		
Seurat R package (v3.5.1 and v4.4.0)	Butler et al. <sup>89</sup> ; Stuart et al. <sup>90</sup>	<a href="https://cran.r-project.org/web/packages/Seurat/index.html">https://cran.r-project.org/web/packages/Seurat/index.html</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ArchR (v1.0.2)	Granja et al. <sup>70</sup>	<a href="https://github.com/GreenleafLab/ArchR">https://github.com/GreenleafLab/ArchR</a>
CIBERSORTx (v1.0)	Newman et al. <sup>43</sup>	<a href="https://cibersortx.stanford.edu/">https://cibersortx.stanford.edu/</a>
Workflow for benchmarking scRNA-seq annotation tools	Abdelaal et al. <sup>54</sup>	<a href="https://github.com/tabdelaal/scRNA-seq_Benchmark/tree/snakeyaml_and_docker">https://github.com/tabdelaal/scRNA-seq_Benchmark/tree/snakeyaml_and_docker</a>
BD Rhapsody™ WTA Analysis Pipeline (v1.2)		<a href="http://www.sevenbridges.com">www.sevenbridges.com</a>
STAR (v2.7.3a)	Dobin et al. <sup>91</sup>	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
NMF R package (v0.20.0)	Gaujoux and Seoighe. <sup>42</sup>	<a href="https://cran.r-project.org/web/packages/NMF/index.html">https://cran.r-project.org/web/packages/NMF/index.html</a>
factoextra R package (v1.0.7)		<a href="https://cran.r-project.org/web/packages/factoextra/index.html">https://cran.r-project.org/web/packages/factoextra/index.html</a>
ape R package (v5.4.1)		<a href="https://cran.r-project.org/web/packages/ape/index.html">https://cran.r-project.org/web/packages/ape/index.html</a>
caret R package (v6.0)	Kuhn et al. <sup>92</sup>	<a href="https://cran.r-project.org/web/packages/caret/index.html">https://cran.r-project.org/web/packages/caret/index.html</a>
DoubletFinder R package (v2.0)	McGinnis et al. <sup>93</sup>	<a href="https://github.com/chris-mcginnis-ucsf/DoubletFinder">https://github.com/chris-mcginnis-ucsf/DoubletFinder</a>
SoupX R package (v1.6.2)	Young et al. <sup>94</sup>	<a href="https://cran.r-project.org/web/packages/SoupX/index.html">https://cran.r-project.org/web/packages/SoupX/index.html</a>
mixtools R package (v2.0.0)	Benaglia et al. <sup>95</sup>	<a href="https://cran.r-project.org/web/packages/mixtools/index.html">https://cran.r-project.org/web/packages/mixtools/index.html</a>
vegan R package (v2.6.4)		<a href="https://cran.r-project.org/web/packages/vegan/index.html">https://cran.r-project.org/web/packages/vegan/index.html</a>
biomaRt		<a href="https://cran.r-project.org/web/packages/biomaRt/index.html">https://cran.r-project.org/web/packages/biomaRt/index.html</a>
clusterProfiler (v4.10.0)	Wu et al. <sup>96</sup>	<a href="https://guangchuangyu.github.io/software/clusterProfiler/">https://guangchuangyu.github.io/software/clusterProfiler/</a>
LISI (v1.0)	Korsunsky et al. <sup>46</sup>	<a href="https://github.com/immunogenomics/LISI">https://github.com/immunogenomics/LISI</a>
bluster (v1.14.0)	Lun A et al. <sup>97</sup>	<a href="https://www.bioconductor.org/packages/release/bioc/html/bluster.html">https://www.bioconductor.org/packages/release/bioc/html/bluster.html</a>
R (v3.5.1)	R Cran	<a href="https://cran.r-project.org/">https://cran.r-project.org/</a>
Scanorama (v1.7.4)	Hie et al. <sup>47</sup>	<a href="https://github.com/brianhie/scanorama">https://github.com/brianhie/scanorama</a>
Prism (v7)	GraphPad Software, Inc.	N/A
FlowJo		N/A
<b>Other</b>		
Resource website for this work	This paper	<a href="https://github.com/ediriwas/adult-human-hspsc">https://github.com/ediriwas/adult-human-hspsc</a>
Healthy scRNA-seq	Ainciburu et al. <sup>36</sup>	GEO: GSE180298
Healthy scRNA-seq	Bunis et al. <sup>55</sup>	GEO: GSE158490
Healthy scRNA-seq	Dussiau et al. <sup>63</sup>	GEO: GSE169426
Healthy scRNA-seq and ATAC-seq	Granja et al. <sup>31</sup>	GEO: GSE139369
Healthy scRNA-seq	Griffin et al. <sup>56</sup>	GEO: GSE227690
Healthy scRNA-seq	Kaufmann et al. <sup>64</sup>	GEO: GSE148884
Healthy scRNA-seq	Mende et al. <sup>57</sup>	GEO: GSE190067
Healthy scRNA-seq	Oetjen et al. <sup>58</sup>	GEO: GSE120221
Healthy scRNA-seq	Roy et al. <sup>59</sup>	GEO: GSE155259
Healthy scRNA-seq	Thongon et al. <sup>60</sup>	GEO: GSE169709
Healthy scRNA-seq	Triana et al. <sup>35</sup>	EGA: EGAS00001005593
Healthy and AML scRNA-seq	Van Galen et al. <sup>32</sup>	GEO: GSE116256
Healthy scRNA-seq	Vanuytsel et al. <sup>61</sup>	GEO: GSE160251
Healthy scRNA-seq	Zheng et al. <sup>98</sup>	GEO: GSE97104
Healthy snRNA-seq and scATAC-seq	Luecken et al. <sup>69</sup>	GEO: GSE194122
ALL scRNA-seq	Caron et al. <sup>87</sup>	GEO: GSE132509

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse scRNA-seq	Nestorowa et al. <sup>73</sup>	GEO: GSE81682
Mouse scRNA-seq	Konturek-Ciesla et al. <sup>29</sup>	GEO: GSE175702

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

### Description of healthy adult donors

Bone marrow mononuclear cells were obtained from healthy donors with informed consent and compliance with relevant ethical regulations (AllCells and STEMCELL Technologies). All healthy cells used in this study were cryopreserved and subsequently thawed by the dropwise addition of IMDM (GIBCO, cat. no. 12440-061) containing 20% fetal bovine serum (FBS; Sigma, cat. no. F1051). Characteristics of each donor, including age and sex, are available in [Table S1](#).

### Animal care

All mouse experiments were conducted in accordance with a protocol approved by the Institutional Animal Care and Use Committee (Stanford Administrative Panel on Laboratory Animal Care #22264) and in adherence with the U.S. National Institutes of Health's Guide for the Care and Use of Laboratory Animals. A mixture of male and female NOD/SCID/IL2Rγ<sup>−/−</sup> (NSG) mice was used and details on age are available in [Table S8](#).

## METHOD DETAILS

### A multi-modal framework for studying adult human HSPCs

To identify immunophenotypically distinct sub-populations within adult human HSPCs, 68,707 BM mononuclear cells from 6 donors ([Figure 2](#); [Table S1](#)) were profiled using concurrent scRNA-seq and scADT-seq ([Figure S1](#); [Table S2](#)). We performed strict quality control to identify high quality cells, and lymphocytes were subsequently removed to specifically study the HSPC and erythromyeloid compartment ([Figures S1A and S1B](#)). The subsequent workflow implements 4 key objectives: 1) computational purification of high quality HSPC and myeloid cell clusters, 2) subsequent FACS based purification, functional evaluation, and annotation of newly identified HSPCs, 3) molecular characterization of HSPCs using multi-dataset integration, 4) cross-species evaluation of human and murine HSPC homology, and 5) large-scale meta-analysis of these HSPCs in aging and cancer.

### Single cell captures for whole transcriptome and surface marker quantification

Cryopreserved adult bone marrow cells were thawed and stained with annexin V (Thermo Fisher Scientific) per protocol for 15 min. Cells were resuspended with DAPI and annexin V/DAPI negative cells were sorted for subsequent ADT staining and single cell capture per established protocols (BD Rhapsody). The BD Rhapsody workflow is a nanowell based single cell capture system with established protocols for generating single cell whole transcription and ADT cDNA libraries. Briefly, an ADT cocktail (1 μL per ADT) was prepared in a total of 100 μL of buffer. Cells were first washed and then labeled with the ADT cocktail for 30 min at 4°C, washed 3 times, resuspended in sample buffer (BD Rhapsody Cartridge reagent kit), and captured with the BD Rhapsody multi-omic single cell system using the manufacturer's instructions.<sup>99</sup> ADT and whole transcriptome libraries were generated per manufacturer's protocol, evaluated by qubit and bioanalyzer, pooled and sequenced using NextSeq500 or Illumina Novaseq S2 (Illumina). Sequencing depth was determined based on manufacturer's recommendations.

### Sample processing, integration, and quality control

Single cell datasets are inherently noisy, as multiplets and pre-apoptotic cells can significantly confound downstream analysis.<sup>100,101</sup> Further, non-specific ADT binding mediated by not only antibody cross-reactivity and poor input cell quality but also complex biological interactions<sup>100–102</sup> can further complicate analysis of multimodal single cell datasets. Therefore, we applied stringent quality control that incorporates the basic Seurat workflow for multimodal data processing and integration,<sup>89,90</sup> ambient RNA and ADT removal,<sup>94,103</sup> doublet discrimination using DoubletFinder,<sup>93</sup> and elimination of cells expressing mutually exclusive lineage markers ([Figure S1](#)).

Fastq files generated from the BD Rhapsody workflow were processed on Seven Bridges (<https://www.sevenbridges.com>) using the standard analysis pipeline (BD Rhapsody WTA Analysis Pipeline). The unique molecular identifier (UMI) count matrices were imported into R (v3.6.2) and processed using Seurat (v3.5.1 and v4.0.0).<sup>89</sup> Each domain, i.e., mRNA and ADT counts, was normalized using log-normalization or centered using log ratio normalization (CLR) respectively. A third concatenated dataset was created by merging normalized counts for specific downstream analyses. Poor quality cells were removed based on total RNA content and percent mitochondrial content using the top 1% expression level. Doublet discrimination was first performed using DoubletFinder with standard metrics.<sup>93</sup> Doublets were removed and gene and surface marker expression were corrected for ambient noise using SoupX.<sup>94</sup> DoubletFinder was used to filter doublets rather than a maximum RNA content filter due to the inherent variability of RNA



content and celltype.<sup>104</sup> Mutually exclusive lineage markers (Table S2) were subsequently used to remove potential heterotypic doublets, and poor-quality cells resulting from non-specific ADT binding.<sup>101</sup> These strict filtering thresholds removed approximately 20% of cells rather than the expected 8–9% based on loading metrics (Table S4). Of note, removing cells based on doublet discrimination would have removed only 7% of cells. We also see that cells removed based on conflicting lineage marker expression had significantly higher ambient ADT signal ( $\rho$ ) as measured by SoupX, despite having similar mitochondrial and RNA content (Figure S2B). The filtered cells were subsequently integrated by sample using canonical correlation analysis (CCA) and the resulting integrated matrix containing the top 5000 mRNA and ADT features was used for downstream computational analyses. For comparison, cells were also integrated using Scanorama<sup>47</sup> and scGen with standard parameters and labeled with cell cycle status as recommended by the developers.<sup>44</sup> Integration metrics were determined using the `compute_lisi` function from the LISI R package.<sup>46</sup>

### Clustering and diversity analysis

Cells were computationally purified using standard and updated HSPC gating strategies (Table S2) with the `CellSelector` function in Seurat. If needed, gating thresholds were determined by modeling the expression of each ADT using mixture models ( $k = 2-3$ ) and the expectation maximization algorithm with `mixtools`.<sup>95</sup> The gap statistic was calculated using the integrated matrix described above within designated sub-population using the `factoextra` R package (`nstart = 25`, 100 iterations, `d.power = 2`, `Tibs2001SEmax`, and K-means clustering).<sup>39</sup> Sub-population purity was assessed using the shannon diversity index and the `vegan` R package.<sup>105</sup> Cluster metrics were determined the `neighborPurity` and `approxSilhouette` function from the `bluster` R package.<sup>97</sup>

### NMF and Louvain clustering

Louvain clustering was performed using Seurat and the resolution was adjusted to achieve target cluster number. Clustering using the non-negative matrix factorization (NMF) approach was performed as previously described<sup>41</sup> using the NMF R package.<sup>42</sup> Briefly, the top 2000 variable features (31 ADTs; Table S2) were identified using the normalized count matrix from the integrated assay containing both mRNA and ADT expression. After nonnegative transformation, the NMF algorithm was performed for  $k = 2-20$  clusters. Optimal cluster number was estimated using the cophenetic index. Initially, the data was over-clustered and annotated using marker features, and clusters with low numbers ( $n < 10$ ) were removed. We used CIBERSORTx<sup>43</sup> to identify similar clusters by first splitting the data into a test and training set (1:2 split), and a signature matrix was created using the training dataset. As we increased the resolution parameter for Louvain clustering, we observed more sub-clusters within the differentiated myeloid and erythroid compartment rather than the HSPC compartment which limited its utility. Artificial bulk transcriptomes were created using the test dataset by subsetting cells based on their cluster label, converting counts to transcripts per million (TPM), and summing counts across all cells. We deconvolved these artificial bulk transcriptomes using the signature matrix, and poor performing clusters were collapsed with the closest mislabeled cluster. Cluster stability was assessed using silhouette scores and cluster purity. The resulting set of clusters were evaluated for established lineage or cell defining marker expression and sub-annotated accordingly (higher resolution immunophenotype; Table S2). With Seurat, differentially expressed genes and surface markers were determined using the `FindAllMarkers` and the MAST test, and cell cycle scores were calculated using the `CellCycleScoring` function.

### Trajectory analysis

Trajectories were determined using a milestone based analysis with PAGA (partition-based graph abstraction method<sup>49</sup>) and `dyn`<sup>48</sup> using established methods. Milestones were ordered based on cell annotations, and differential features associated with each milestone were determined using `FindAllMarkers` and the MAST test in Seurat.

### Surface marker identification

We performed an iterative differential surface marker analysis to identify a specific immunophenotype for each newly identified HSPC sub-population. HSCs, MPP1-3s, LMPP1-2s, GMPs, and ProBEMs were isolated from the original dataset. Differential ADTs were first determined across all cell populations. Differential markers were evaluated manually, and cells were computationally purified based on informative markers and then re-evaluated for differential markers (Table S2). Candidate features were subsequently evaluated using FACS to identify markers with sufficient dynamic range to efficiently purify each sub-population (data not shown). Due to sort complexity and low cell numbers, MPP3s and ProBEMs were purified as one population. Additionally, CD19 was included as a lineage marker, thereby excluding ProB cells in these assays.

### Fluorescence-activated cell sorting

Thawed cells were washed with FACS buffer (PBS, 2% FBS, 2 mM EDTA) and stained with antibodies for 30 min on ice in 50  $\mu$ L total volume. Cells were then stained with Zombie Red for viability assessment for 5 min and subsequently washed prior to analysis. All cell sorting steps were confirmed using post-sort purity analyses. The antibodies used for respective assays are provided in Table S4. Flow cytometry was performed on a FACSria II (Becton Dickinson). Gates were drawn by internal positive and negative controls using healthy cells and validated by back-gating on marker positive events. For CD69, gating was further evaluated using fluorescence minus one gating (Figure S5). Positive gates for CD49f expression were based on peak expression for the HSCs from a particular donor, and the same gate was used for other samples from the same donor. This allowed us to quantify relative CD49f expression compared an internal control. Gates drawn for all other markers were consistent across all samples (Figure S5).

### Single cell colony formation assay

Single HSPCs were sorted into 96 well plates containing 120  $\mu$ L of Methocult H4034 methylcellulose media (STEMCELL Technologies). Cells were allowed to differentiate for 12 days and colonies were morphologically assessed. The ratio of colonies per total cells seeded is presented for each celltype.

### Megakaryocyte expansion assay

Purified HSPCs were plated in 96 well plates containing megakaryocyte differentiation medium (StemSpan SFEM II with human low-density lipoproteins (STEMCELL Technologies) and Megakaryocyte Expansion Supplement (STEMCELL Technologies)). Megakaryocyte production was assessed at 7 days by staining cells for CD41a and CD61 expression. Megakaryocyte number was normalized to live cells and 1000 cells seeded.

### Xenotransplantation assays

Purified HSPCs were transplanted into the right femur of sub-lethally irradiated NSG mice (6–8 weeks, male and female, 200 rad 2–24 h pre-transplant).<sup>52</sup> Mice were harvested at 17–18 weeks by crushing the pelvis and bilateral femurs and tibias. Cells were filtered and stained using the engraftment panel (Table S2) to assess both engraftment and lineage. Human to mouse chimerism was calculated based on hCD45+HLA-ABC+ cells relative to mouse CD45+mTer119-cells. Chimerism >0.1% was considered positive for long term engraftment.

### OP9 and OP9-hDL4 maintenance

OP9 and OP9-hDL4 cells were gifted from the Bendall Lab at Stanford University. Cells were cultured in fresh OP9 media (MEM $\alpha$  (Gibco12561056), 15% FBS, 1% L-glutamine, and 1% Penicillin-Streptomycin (Gibco)). Cells were split 1:4 or 1:5 after reaching 90% confluency.

### OP9 co-culture lymphoid-myeloid cell differentiation assay

OP9 cells were seeded in a 96 well plate (2500 cells/well) in 37C, 5% CO<sub>2</sub> incubation. Twenty-four hours after plating, purified HSPCs were added to each well and expanded with 20 ng/mL SCF, 10 ng/mL FLT3L, 10 ng/mL G-CSF, 10 ng/mL IL-2 (all from Peprotech, London, UK), 10 ng/mL IL-15 (STEMCELL Technologies), and 0.2  $\mu$ M DUP697 (Cayman Chemicals) in OP9 media (SGF15 media). Cells were dissociated and transferred to new plates with fresh OP9 cells and SGF15 media weekly. Cells were analyzed by flow cytometry at weeks 1, 2 and 3. Lineage output was assessed by staining cells for CD45, CD34, CD15, CD14, CD19, and CD56 according to immunophenotypes described in Table S2. Cell output was normalized to hCD45 content for lineage output quantification and both hCD45 content and cells seeded for production quantification.

### OP9-hDL4 co-culture T cell differentiation assay

OP9-hDL4 cells were seeded in a 96 well plate (2500 cells/well) in 37C, 5% CO<sub>2</sub> incubation. Twenty-four hours after plating, purified HSPCs were added to each well and expanded with 10 ng/mL SCF, 5 ng/mL FLT3L and 5 ng/mL IL-7 (Peprotech, London, UK) in OP9 media (SF7 media). Cells were dissociated and transferred to new plates with fresh OP9-hDL4 cells and SF7 media weekly. Cells were analyzed by flow cytometry at weeks 3 and 4. Lineage output was assessed by staining cells for CD45, CD34, CD7, CD5, CD3, CD4, and CD8 according to immunophenotypes described in Table S2. Cell output was normalized to hCD45 content for lineage output quantification and both hCD45 content and cells seeded for production quantification.

### Basophil, eosinophil and mast cell expansion assay

Purified HSPCs were plated in 96 well plates containing fresh differentiation media (IMEM (Gibco A1048901) and 1% Penicillin-Streptomycin (Gibco)). Cells were expanded with either 20 ng/mL SCF and 20 ng/mL IL3 (mast cell differentiation media), 20 ng/mL IL3 (basophil differentiation media), or 20 ng/mL IL3 and 20 ng/mL IL5 (eosinophil differentiation media). Cytokines (Peprotech, London, UK) were replenished every other day, and output was analyzed on day 7 by flow cytometry using CD34, CD45RA, CD14, CD117, CD193, and CD66b and immunophenotypes described in Table S2.

### Integrated single cell ATAC sequencing analysis

The scRNA-seq data was integrated with published scATAC-seq and snRNA-seq data from adult donors with Analysis of Regulatory Chromatin in R (ArchR) using the developers' recommendations unless otherwise stated.<sup>69,70</sup> The integrated gene expression and chromatin accessibility analysis was performed using a non-redundant, multi-model motif reference database curated by Vierstra et al.<sup>71</sup> The integrated approach allowed us to discriminate between specific TFs within TF families, and the curated motif reference provided information across multiple models while retaining TF binding information. Briefly, arrow files were generated using published read fragment data, low quality cells and doublets were removed, a tile matrix was created, and dimensional reduction was performed using latent semantic indexing. The addGeneIntegrationMatrix function was used to perform an unconstrained integration with the snRNA assay and Latent Semantic Indexing using both the tile matrix and snRNA assay. Pseudobulk replicates were constructed using the addGroupCoverages function, peak calling was performed using the addReproduciblePeakSet, and a peak matrix was created using addPeakMatrix. Differentially opened peaks were determined using addMarkerFeatures, motif annotations

were added using the `addMotifAnnotations` function and the Vierstra database, deviations were calculated using `addDeviationMatrix` and `getGroupSE`, and footprints were identified using `getFootprints`. Correlated gene expression using the gene integration matrix and motif deviations were calculated using the `correlateMatrices` function to identify putative regulators. Trajectories were added using the `addTrajectory` function and motif deviations and gene expression were correlated across pseudotime to identify differential regulators (Table S9).

### Deriving a mixed model classifier for cell label transfer

Gene expression counts from each respective dataset (Table S12) were processed and integrated as described above. Code provided by Abdelaal et al.<sup>54</sup> was adapted to evaluate classifiers based on several models: CHETAH,<sup>106</sup> scID,<sup>107</sup> RF,<sup>108</sup> singleR,<sup>109</sup> singleCellNet,<sup>110</sup> scmapcell,<sup>111</sup> SVM (linear kernel with and without rejection),<sup>112</sup> and scPred (MDA, NNet, avNNet, GLM, SVM with radial kernel).<sup>113</sup> The intra-dataset analysis was performed as described by the authors using 10-fold CV, and default parameters for each classifier. Classifier performance was evaluated using the 15-cell type annotation (Figure 1B) and two metrics: F1 scores, which measures model accuracy, and cell annotation recovery (labeled), which quantifies the proportion of cells with confident cell annotation assignments. We first performed an intra-dataset analysis (Figure S6A) where we split our reference dataset into a test and training set and evaluated each model using 10-fold cross validation. We evaluated 13 established classifiers (Figures S6B and S6C)<sup>54</sup> using this approach, and found that high performing classifiers used support vector machines (SVMs), generalized linear models (GLMs), averaged neural networks (avNNets), or mixture discriminate analysis (MDA). We then evaluated the performance of each classifier for each cell type using the scPred R package<sup>54</sup> which revealed that classifier performance was also dependent on cell type (Figure S6D). The top models were identified based on F1 scores, and proportion of cells labeled and evaluated using a celltype specific approach with scPred with default parameters. Top performing models by celltype was used for the final mixed-model classifier. For the inter-dataset analysis, we first held out 10% of the data, and the remaining 90% of data was used to train the model. We trained our mixed model classifier using the training dataset which was used to assign labels on the external dataset, i.e., the query dataset. We then used the labeled query dataset to retrain the mixed-model classifier which then assigned labels onto the left-out test dataset from our reference. This approach allowed us to indirectly evaluate the performance of the model since the test dataset contains the true cell labels. This modified 10-fold CV approach was used to evaluate classifier performance. For final cell label assignments, we used the entire reference dataset and a probability threshold of 0.75 rather than the default of 0.5 to label the query dataset, and unassigned cells were removed from the subsequent analysis.

### Cross-species analysis

A cross-species analysis was performed using the human reference (i.e., Ref. 1) and young adult<sup>73</sup> and adult mouse cells.<sup>29</sup> Gene aliases were relabeled using gene ortholog-to-ortholog conversions in R with biomaRt<sup>114</sup> and murine single cell count matrices were subset to contain only conserved orthologs.<sup>89</sup> Non-HSPCs were removed in the mouse data to ensure that comparable cell states were being evaluated in the downstream analysis. Count tables were processed in Seurat as described above. Cell type specific gene expression signatures were determined using the `AddModuleScore` function in Seurat. Signatures were derived from published studies<sup>74–76</sup> and provided in Table S14. Murine cells were labeled using the mixed model classifier, and the data were clustered using the Louvain algorithm. Conserved features were identified using the `FindConservedMarkers` function in Seurat, and murine and human cells were integrated using only conserved markers with Seurat CCA. Co-occurrence heatmaps were generated for each integrated dataset, and average gene expression per cluster was determined using the `AverageExpression` function in Seurat to perform a gene expression correlation analysis. Trajectory analysis was performed using `dyno` and `PAGA` as previously described using conserved genes. A separate analysis was performed for human and mouse cells within each cross-species comparison, i.e., young adult and adult mice. Milestones were ordered based on known differentiation nodes, and HSC clusters were labeled as the root node. Important features associated with bifurcation nodes were determined using the `calculate_branching_point_feature_importance` function in `dyno`. Important features that were conserved between species are summarized in Table S11.

### Aging analysis

Datasets with both age and celltype annotations were included in this analysis (Table S12) and processed as described above. Samples were categorized based on age. Each dataset was labeled using the mixed model classifier using the reference profile annotation described in Figure 1B. Cells were subset by celltype and integrated as described above. To identify important genes associated with aging, gene expression was correlated with age using a spearman correlation weighted by cells per age (Table S9). Important genes were selected based on significance (adjusted  $p$ -value  $< 0.05$  based on the Benjamini-Hochberg procedure)  $< 0.05$ . GO term and Pathway (Reactome) over representation analysis using these important genes were performed using `clusterProfiler` in R.<sup>96</sup> Based on the results from the over representation analysis, gene signatures were derived to further assess these biological processes (Table S13) and evaluated in specific cell subsets using `AddModuleScore`. To derive celltype specific aging signatures, data were split into test and training sets (1:1 split) and recursive feature elimination (RFE) was performed using `caret` with linear models, bagged trees, and RF regression models on the training dataset. Informative features based on RFE was used to predict age in the test dataset and root mean squared error (RMSE) and correlation ( $R^2$ ) was calculated. The RF method demonstrated the best performance when predicting age in the test dataset and was therefore used to derive the final aging signatures. To study aging signatures in

human blood cancers, external datasets (Table S12) were processed as described above and module scores based on informative features from the RFE analysis (Table S13) using the AddModuleScore function in Seurat.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Details of statistical tests performed can be found in figure legends or materials and methods. For significance testing, a t test (2-tailed distribution and two-sample equal variance) with Mann-Whitney multiple comparison adjustment was used for comparing 2 conditions or a one or two-way ANOVA with Tukey's multiple comparison adjustment when comparing >2 conditions unless otherwise stated. Results with adjusted  $p$ -values <0.05 were considered significant. All statistical analyses were performed in R or GraphPad Prism.