New techniques for single-cell analysis have led to insights into hematopoiesis and the immune system, but the ability of these techniques to cross-validate and reproducibly identify the biological variation in diverse human samples is currently unproven. We therefore performed a comprehensive assessment of human bone marrow cells using both single-cell RNA sequencing and multiparameter flow cytometry from 20 healthy adult human donors across a broad age range. These data characterize variation between healthy donors as well as age-associated changes in cell population frequencies. Direct comparison of techniques revealed discrepancy in the quantification of T lymphocyte and natural killer cell populations. Orthogonal validation of immunophenotyping using mass cytometry demonstrated a strong correlation with flow cytometry. Technical replicates using single-cell RNA sequencing matched robustly, while biological replicates showed variation. Given the increasing use of single-cell technologies in translational research, this resource serves as an important reference data set and highlights opportunities for further refinement.
Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry

Karolyn A. Oetjen,1 Katherine E. Lindblad,1 Meghali Goswami,1 Gege Gui,1 Pradeep K. Dagur,2 Catherine Lai,1 Laura W. Dillon,1 J. Philip McCoy,2 and Christopher S. Hourigan1

1Laboratory of Myeloid Malignancies, National Heart Lung and Blood Institute, Bethesda, Maryland, USA. 2Flow Cytometry Core, National Heart Lung and Blood Institute, Bethesda, Maryland, USA.

Introduction

New technologies for characterizing cell populations are being implemented to more deeply describe the cell surface receptor phenotype and gene transcriptional signature at the single-cell level (1, 2). Benefits of single-cell approaches include examination of heterogeneity within the sample, and the most recent advances permit use of samples with very limited cell numbers for high-dimensional characterization of cell surface phenotype or transcriptome. Single-cell RNA sequencing (scRNA-Seq) has been used to elucidate hematopoietic differentiation (3–5) and immune cell subsets (6), including dendritic cells and monocytes (7), and innate lymphoid cells (8). Mass cytometry has been applied to the study of tissue-infiltrating immune cells (e.g., melanoma, ref. 9; renal cell carcinoma, ref. 10; lung cancer, ref. 11; and breast cancer, ref. 12).

Expanding these new single-cell approaches to patient samples requires a clear understanding of their correlation with established techniques, including flow cytometry. To facilitate and validate analysis of large databases of scRNA-Seq, we set out to provide a data set of human bone marrow analyzed by both scRNA-Seq and deep immunophenotyping. Our reference cohort includes a broad range of donor ages in recognition of age-related variation in the healthy population.

Results

Healthy donor characteristics. Twenty healthy volunteers were recruited for bone marrow aspiration procedures. The cohort consisted of 10 males and 10 females with ages ranging from 24 to 84 years old and median age of 57 years. A second bone marrow aspiration was performed for 2 donors (Ck, Sk) (biological replicates) either 2 or 5 months after their first aspiration, respectively. Cryopreserved cells from all 20 donors were analyzed by droplet-based scRNA-Seq and flow cytometry, and additional cryopreserved vials for 8 donors were analyzed by mass cytometry for T cell phenotyping, as well as bulk RNA sequencing, as summarized in Table 1.

Single-cell RNA sequencing. Droplet-based scRNA-Seq of bone marrow mononuclear cells (BMMCs) for all donor samples was performed with goal minimum sequencing depth of 50,000 reads/cell and detected a mean of 880 genes/cell (range 575–1,390 genes/cell, Table 1). Greater than 90,000 cells were captured; using quality filters of at least 500 genes per cell and less than 8% mitochondrial RNA content, 76,645 cells were analyzed in the final analysis.
To account for sample variations among donors, alignment of all samples was performed in Seurat using canonical correlation analysis (CCA), then visualized using t-distributed stochastic neighbor embedding (t-SNE). Cell clusters were distinguished using the Louvain clustering algorithm implemented in Seurat. Compiled analysis of all donor cells is annotated in Figure 1A, with the contribution of each individual donor displayed in Supplemental Figure 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.124928DS1. All major previously identified populations of BMMCs were present in the clustered scRNA-Seq analysis.

Single-cell trajectory analysis was performed using Monocle 3. As there were potentially multiple disjoint trajectories in this complex data set containing a large number of cells, Uniform Manifold Approximation and Projection (UMAP) was used for dimension reduction. The resulting development trajectories clearly display the major lymphoid, myeloid, and erythroid lineages of hematopoiesis with correct ordering of developmental stages (Figure 1B). Trajectories of erythroid and myeloid lineages could also be created using an earlier, well-validated, version of this software (Monocle 2, see Supplemental Figure 1) and were consistent with those observed for the full data set.

Annotation of cell cluster identities was determined using a panel of canonical gene expression, with the expression patterns for a subset of these genes displayed in Figure 1C. Analysis of each donor sample individually using principal component analysis (PCA) in Seurat revealed suboptimal quantification of frequencies of some transcriptionally similar cell subsets, including those annotated as effector T cells and NK cells. Such clusters were typically well delineated for each individual sample when using CCA in the context of the entire data set (Supplemental Figure 1).

A potential use of scRNA-Seq is to compare across 2 or more samples. To confirm the validity of Table 1. Heathy volunteer sex and age at time of bone marrow aspiration

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sex</th>
<th>Age</th>
<th>Flow cytometry</th>
<th>Mass cytometry</th>
<th>Bulk RNA</th>
<th>scRNA-Seq</th>
<th>Technical replicate</th>
<th>Cells</th>
<th>Reads/cell</th>
<th>Genes/cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>F</td>
<td>24</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>4,293</td>
<td>88,461</td>
<td>1,000</td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>F</td>
<td>28</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>3,643</td>
<td>47,132</td>
<td>649</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>M</td>
<td>30</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>3,939</td>
<td>69,061</td>
<td>575</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>M</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3,593</td>
<td>131,257</td>
<td>669</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>41</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>3,746</td>
<td>75,712</td>
<td>846</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>F</td>
<td>43</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>3,446</td>
<td>108,679</td>
<td>970</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>F</td>
<td>46</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>4,118</td>
<td>85,295</td>
<td>1,143</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>M</td>
<td>47</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>3,293</td>
<td>99,593</td>
<td>794</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>F</td>
<td>50</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>5,013</td>
<td>63,164</td>
<td>883</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>M</td>
<td>50</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>4,516</td>
<td>47,778</td>
<td>851</td>
<td></td>
</tr>
<tr>
<td>SkA</td>
<td>F</td>
<td>55</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x Sk1</td>
<td>323,589</td>
<td>823</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x Sk2</td>
<td>1,138</td>
<td>820</td>
<td></td>
</tr>
<tr>
<td>S^A</td>
<td>56</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x S1</td>
<td>113,302</td>
<td>1,089</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x S2</td>
<td>2,437</td>
<td>1,163</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>M</td>
<td>57</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>4,548</td>
<td>67,199</td>
<td>950</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>F</td>
<td>58</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>3,833</td>
<td>223,652</td>
<td>1,390</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>M</td>
<td>58</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>4,283</td>
<td>89,208</td>
<td>667</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>F</td>
<td>59</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>2,994</td>
<td>159,501</td>
<td>1,303</td>
<td></td>
</tr>
<tr>
<td>CkA</td>
<td>F</td>
<td>59</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>1,052</td>
<td>349,511</td>
<td>761</td>
<td></td>
</tr>
<tr>
<td>C^A</td>
<td>60</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>C1</td>
<td>3,556</td>
<td>692</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x C2</td>
<td>3,136</td>
<td>692</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>M</td>
<td>60</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>3,964</td>
<td>92,780</td>
<td>875</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>M</td>
<td>66</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>1,700</td>
<td>126,143</td>
<td>702</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>M</td>
<td>67</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>4,522</td>
<td>110,195</td>
<td>881</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>M</td>
<td>84</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>7,247</td>
<td>43,872</td>
<td>879</td>
<td></td>
</tr>
</tbody>
</table>

Sk and S are 2 samples taken from the same donor at different times (i.e., biological replicates). Ck and C 2 samples taken from the same donor at different times (i.e., biological replicates). Samples Sk, S, and C were split and used for technical replicates of scRNA-Seq. Biological replicate time points for a second longitudinal bone marrow aspirate from the same volunteers. Assays from matched cryopreserved BMMC vials are indicated. scRNA-Seq cell counts and sequencing depth (reads per cell and genes per cell) are listed for each donor and replicate.
Figure 1. Single-cell RNA sequencing of healthy bone marrow cells. (A) Cluster identification visualized using t-SNE. HSPCs: hematopoietic stem/progenitor cells; cDCs: conventional DCs; pDCs: plasmacytoid DCs. (B) Single-cell trajectory analysis using UMAP/Monocle 3. Color is as in (A). (C) Examples of canonical gene expression used for annotation. (D) Reproducibility of technical replicates for single-cell RNA sequencing. Linear regression line displayed in gray.
scRNA-Seq for this approach, assay reproducibility was determined by preparing duplicate, side-by-side libraries from cells thawed from the same cryopreserved vial, for a total of 3 cryopreserved samples. Cell subtype quantification for each of these technical replicate pairs matched robustly (Figure 1D). The optimum number of cells required to identify, using scRNA-Seq, subpopulations within a heterogeneous sample remains an area of interest (13). Technical replicates ranged from 1,138 to 6,692 cells from the same sample (Table 1).

**Flow cytometry.** Thirteen-color flow cytometry using 5 customized panels (T, B, and NK cells; monocytes; and DCs see Supplemental Table 1) designed to allow deep immunophenotyping of the predominant cell populations found in human bone marrow was performed on all samples. Approximately 1 million cells were stained for each panel and a median of 196,000 CD45+ events collected (25th–75th percentile: 100,000–278,000 events). The gating strategy is shown in Supplemental Table 2. Most frequent cell subtype populations observed were, in order, T cells, monocytes, B cells, NK cells, DCs, and HSPCs (see Supplemental Figure 2).

Paired analysis of the same sample by both transcriptome and cell surface phenotype offers a powerful opportunity to compare cell population frequencies determined by these methods. The proportion of major cell populations is summarized for scRNA-Seq and flow cytometry in Figure 2A. Sample-by-sample correlations for all of these populations are shown in Figure 2B. It is well established that the memory T cell population increases with increasing age in humans, likely due to response to viral infection (in particular CMV), and this trend was reproduced in our cohort using both scRNA-Seq and flow cytometry (Supplemental Figure 2 and ref. 14). Two subjects had a second bone marrow aspiration performed at either 2 or 5 months after their first aspiration. These biological replicates showed good concordance by flow cytometry but showed variation by scRNA-Seq particularly in lower frequency cell subsets, likely from sampling error (Supplemental Figure 2).

While concordance between these 2 modalities was generally good, it appeared that T cell frequency was elevated and NK cell frequency decreased in scRNA-Seq as compared with flow cytometry. This led to a more detailed examination of T cell subsets and orthogonal validation of cell surface immunophenotyping using a third single-cell modality.

**Mass cytometry.** To more deeply characterize immune populations within healthy bone marrow, and to validate our flow cytometry results, T cell phenotyping was performed by mass cytometry using a 34-marker panel for a subset of 8 donors. Using Cytobank software, CD45+ cells were visualized using viSNE across the panel of markers (Figure 3A and Supplemental Figure 3). Correlation between mass cytometry and flow cytometry for CD4+ and CD8+ T lymphocyte subsets was strong, as shown in Figure 3B.

To further compare mass cytometry and flow cytometry with scRNA-Seq of T cell populations, the frequencies of T cell subsets for this cohort of 8 donors were determined using all 3 of these methods, shown in Figure 3C, with sample correlations reported in Figure 3D. Comparing frequencies of T cell populations between mass cytometry and scRNA-Seq confirmed a small but persistent skewing in the identification of NK and T cells. Using Bland-Altman calculations, the mean difference between scRNA-Seq and mass cytometry for T cells was −6.5% (95% CI: −29% to 16%) and for NK cells was 3.2% (95% CI: −1.1% to 7.6%).

CD8+ cytotoxic T cells and NK cells are known to have substantial overlap at the transcriptome level (15). To better understand systemic bias in the frequency of NK or T cells identified, we confirmed that overlapping gene signatures were found in clusters annotated as NK or T cells in this scRNA-Seq data set (Supplemental Figure 4). The reasons for this bias are likely, however, multifactorial.

**Bulk RNA sequencing.** Analysis of bulk sample RNA expression has been used to attempt to deconvolute the proportion of each cell subtype in human tissues (16, 17). Finally, as an additional resource, stranded whole transcriptome sequencing of RNA isolated from thawed BMMCs was performed on samples from all 8 subjects for which mass and flow cytometry and scRNA-Seq were available. Initial analysis using deconvolution algorithms that attempt to predict the proportion of cell subpopulations is shown in Supplemental Table 3.

**Discussion**

Changes in the immune system (14) and hematopoiesis (18) occur during human aging. Using an unbiased approach based on unsorted human BMMCs, we describe the major cell populations of healthy human bone marrow from a cohort of donors over a wide range of adult age with multiple high-dimensional single-cell techniques. This resource serves as a complement to existing data sets that have consisted primarily of younger donors without associated paired immunophenotyping. Our data set provides
Figure 2. Comparison of single-cell RNA sequencing and flow cytometry assessment of bone marrow cell type population frequencies from 22 samples (taken from 20 donors, including 2 biological replicates). (A) Frequencies for major cell populations in human bone marrow shown for single cell scRNA-Seq and flow cytometry. Each dot represents a value from 1 sample. The thick line within each box represents median value. Box spans first to third quartile (IQR). Whiskers extend to the largest or smallest value no farther than 1.5 IQRs from the box. (B) Individual sample comparisons by scatter plot for each cell population. Each dot represents the cell subset frequency from 1 sample. Population comparisons are shown in the background in gray. Population frequencies are reported as percentage of all CD45+ cells.
Figure 3. Comparison of single-cell RNA sequencing, mass cytometry, and flow cytometry assessment of T lymphocyte frequencies in human bone marrow. (A) Mass cytometry for phenotyping of T cell populations visualized using viSNE analysis with expression of key markers shown. (B) Comparison of cell frequencies for each donor determined by mass cytometry (CyTOF) and flow cytometry. CM: central memory cells; EM: effector memory cells; TEMRA: terminally differentiated effector memory T cells; TE: effector T cells; DNT: double-negative T cells; DPT, double-positive T cells. (C) T cell frequencies for cell populations identified by mass cytometry, flow cytometry, and scRNA-Seq. Each dot represents a value from 1 sample. The thick line
a resource of scRNA-Seq, flow cytometry, and mass cytometry data for healthy control cohorts across the full range of adulthood, providing not only cell population frequencies and characteristics but also highlighting individual variation in human cohorts.

Using scRNA-Seq of over 76,000 cells from 20 healthy donors, all the major bone marrow mononuclear populations are identified, and overall population frequencies are comparable to flow cytometry of the matched samples. A primary limitation is distinguishing cell populations such as NK cells and CD8+ effector T cells, which have overlapping transcriptional programs with a small number of distinguishing genes captured by droplet-based scRNA-Seq. To overcome this limitation and provide additional reference data beyond previous reports of major healthy bone marrow populations by flow cytometry (19) and mass cytometry (20, 21), we used the strength of mass cytometry to resolve T cell subpopulations (22), both to validate our flow cytometry results and provide quantification of rare T cell subpopulations within healthy human bone marrow.

As a data resource, these high-dimensional approaches to bone marrow characterization add valuable information on transcriptional and cell surface marker coexpression. The growing number of bioinformatics tools for mass cytometry (23) and scRNA-Seq (24, 25) will benefit from reference data sets for validation and integrated comparison across techniques. Future opportunities for integrating these data sets include droplet-based sequencing with oligonucleotide-tagged antibodies, including Cellular Indexing of Transcriptome and Epitopes sequencing (26), RNA expression and protein sequencing (27), and AbSeq (28), which can be compared to this reference set of cell surface protein and transcriptome expression. As techniques (29) and repositories (30) of high-dimensional single-cell human data sets are expanded, validating the observed cell identities will be a critical aspect of interpreting large–data set analysis.

Additional aliquots of bone marrow aspirate from this cohort together with paired blood samples, that were not yet analyzed, have been stored. Should transformative technologies emerge over the next few years, we would be willing, subject to relevant technology transfer and clinical regulatory approvals, to share remaining samples with academic investigators for additional benchmarking and validation. In summary, this resource provides a reference data set for cell populations in healthy human bone marrow across a wide age range as assessed by multiple single-cell approaches. We show that scRNA-Seq quantification of marrow-resident cell populations has good concordance with immunophenotyping by flow and mass cytometry with some discrepancies in T and NK subsets. We hope this unique combined data set will prove useful both to those seeking to refine or innovate bioinformatic algorithms for scRNA-Seq data and to those investigators hoping to apply these powerful single-cell technologies in their own research.

Methods

**Bone marrow aspirate collection.** Using standard operating procedures, mononuclear cells from bone marrow aspirates were isolated using Ficoll density gradient separation and cryopreserved in 90% FBS/10% DMSO for storage in liquid nitrogen. Assays were performed as listed in Table 1 using matched cryopreserved vials from each donor.

**Single-cell RNA sequencing.** scRNA-Seq was performed using 10X Genomics Single Cell 3′ Solution, version 2, according to the manufacturer’s instructions (protocol rev A). Libraries were sequenced on HiSeq3000 (Illumina) and analyzed using Cell Ranger version 2.0.0 (10X Genomics). Quality control metrics were used to select cells with mitochondrial gene percentage less than 8% and at least 500 genes detected. Samples were analyzed using Seurat (https://satijalab.org/seurat/) using CCA with Louvain clustering and visualized by t-SNE (31). Developmental trajectories were created using Monocle versions 2 and 3 (32–34), the latter using UMAP for dimension reduction (35).

**Flow cytometry.** BMMCs were thawed in RPMI 1640 (Gibco) with 10% FBS and resuspended in cell staining buffer. Benzonase nuclease (Sigma-Aldrich, catalog E1014-25KU) was added for some samples during thawing to minimize cell clumping. Cells were blocked with Human TruStain FcX Fc receptor blocking solution (BioLegend, catalog 422302) and stained with antibodies listed in Supplemental Table 1 followed by LIVE/DEAD Fixable Yellow stain (Life Technologies Corporation) and fixation with 1%
formaldehyde. Data were acquired with a BD Biosciences (BD) LSRFortessa equipped with 5 lasers (355-, 407-, 488-, 532-, and 633-nm wavelengths) and 22 PMT detectors using DIVA 8 software using the High Throughput Sampler system (BD) at a flow rate of 2.5 μl/second in a 96-well U-bottom tissue culture plate (see Supplemental Table 4). Compensation controls were performed using single-color staining of compensation beads (BD), and daily quality assurance was performed using Cytometer Setup and Tracking beads (BD) as per manufacturer’s recommendation along with 1-peak Rainbow beads (BD) and 8-peak beads (SPHEROTEC) (36, 37). Analysis after acquisition was performed using FlowJo version 9.9.6 (Tree Star Inc). Analysis excluded debris and doublets using light scatter measurements and dead cells by live/dead stain. Gating strategies used to identify immune cell subsets are provided in Supplemental Table 2.

**Mass cytometry.** Thawed BMMCs were stained for 34 markers using the Maxpar Complete Human T Cell Immuno-Oncology Panel Set (Fluidigm), according to manufacturer instructions (see also Supplemental Table 5). Briefly, cells were thawed, washed, incubated with cisplatin cocktail for viability, fixed in 1.6% formaldehyde, and permeabilized. Cells were then stained with the antibody cocktail, incubated with intercalation solution, mixed with EQ Four Element Calibration Beads, and acquired with a Helios mass cytometer (Fluidigm). Gating and viSNE analysis (38) were performed using Cytobank (https://www.cytobank.org/). Initial analysis excluded doublets using DNA content and nonviable cells using cisplatin. CD45+ cells were gated for viSNE analysis of 100,000 total events from all analyzed samples.

**Bulk RNA sequencing.** RNA was harvested from thawed cell vials of BMCCs using AllPrep kits (QIAGEN). Libraries were prepared using TruSeq Stranded Total RNA Sample Preparation Kit (Illumina) with 1 μg of RNA input. Sequencing was performed by paired-end 75 nt on Illumina HiSeq 3000. FASTQ files were mapped to using kallisto, and gene counts were tabulated using tximport. Deconvolution was performed using xCell version 1.1 (http://xcell.ucsf.edu/, ref. 16) or CIBERSORT using LM22 gene signature and 100 permutations (https://cibersort.stanford.edu/index.php, ref. 17).

**Sequencing data availability.** FCS files for flow cytometry and mass cytometry data sets have been deposited in FlowRepository (FR-FCM-ZYQ9, FR-FCM-ZYQB). scRNA-Seq and bulk RNA sequencing data sets have been deposited in the National Center for Biotechnology Information’s Gene Expression Omnibus (GSE120221, GSE120446).

**Statistics.** Data were analyzed and visualized, and statistical comparisons were performed with R (cran.r-project.org). Bland-Altman analysis (39) was implemented in the BlandAltmanLeh package version 0.3.1.

**Study approval.** Healthy volunteers were recruited for bone marrow aspirate collection at the NIH. This research was approved by the National Heart, Lung, and Blood Institute Institutional Review Board, and all participants provided oral and written informed consent.

**Author contributions**

KAO performed experiments, analyzed data, and wrote the manuscript. KEL and MG performed flow cytometry experiments and analyzed data. GG and LWD analyzed data. PKD and JPM designed, supervised, and analyzed flow cytometry experiments. CL coordinated donor recruitment. CSH designed experiments, analyzed data, and wrote the manuscript. All authors reviewed the final manuscript.

**Acknowledgments**

This work was supported by the Intramural Research Program of the National Heart, Lung, and Blood Institute (NHLBI) of the NIH. We appreciate the technical expertise of Brian Sellers and the NIH Center for Human Immunology, Yan Luo, Yuesheng Li and the NHLBI DNA Sequencing Core, and Alan Hoofring of NIH Medical Arts and NIH High Performance Computing. We thank Sheenu Sheela, Blair DeStefano, Janet Valdez, and NHLBI research nurses for bone marrow aspirate procedures. The authors would like to thank Neal Young, Cindy Dunbar (both of NIH), and Elizabeth Jaffee (of Johns Hopkins) for reading and comments and Efthymia Papalexii and Rahul Satija (both of New York Genome Center) for help with scRNA-Seq analysis and comments.

Address correspondence to: Christopher S. Hourigan, Laboratory of Myeloid Malignancies, Hematology Branch, National Heart, Lung, and Blood Institute, Room 10CRC 5-5130, 10 Center Drive, Bethesda, Maryland 20814-1476, USA. Phone: 301.451.0257; Email: hourigan@nih.gov.