

Title: Identification of two distinct pathways of human myelopoiesis

Authors: Roy Drissen¹, Supat Thongjuea^{2,3}, Kim Theilgaard-Mönch^{4,5} and Claus Nerlov^{1*}.

Affiliations:

¹MRC Molecular Hematology Unit, MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, United Kingdom

²MRC WIMM Centre for computational biology, MRC Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, United Kingdom.

³The NIHR Oxford Biomedical Research Center, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, United Kingdom.

⁴Department of Hematology and Finsen Laboratory, National University Hospital.

⁵Biotech Research and Innovation Centre, University of Copenhagen, Copenhagen, Denmark.

*Correspondence to:

Prof. Claus Nerlov

MRC Molecular Hematology Unit

MRC Weatherall Institute of Molecular Medicine

Radcliffe Department of Medicine, University of Oxford

John Radcliffe Hospital, Headington

Oxford OX3 9DS, United Kingdom

Phone: +44 1865 222 324

Fax: +44 1865 222 500

Email: claus.nerlov@imm.ox.ac.uk

Abstract: Human myelopoiesis has been proposed to occur through oligo-potent common myeloid progenitor (CMP) and lymphoid-primed multi-potent progenitor (LMPP) populations (1, 2). However, other studies have proposed direct commitment of multi-potent cells to uni-
5 lineage fates, without specific intermediary lineage co-segregation patterns (3, 4). We here show that distinct human myeloid progenitor populations generate the neutrophil/monocyte and mast cell/basophil/eosinophil lineages as previously shown in mouse (5). Moreover, we find that neutrophil/monocyte potential selectively co-segregates with lymphoid lineage potential and mast cell/basophil/eosinophil potential with megakaryocyte/erythroid potential early during
10 lineage commitment. Furthermore, after this initial commitment step mast cell/basophil/eosinophil and megakaryocyte/erythroid potentials co-localize at the single cell level in restricted oligo-potent progenitors. These results show that human myeloid lineages are generated through two distinct cellular pathways defined by complementary oligo-potent cell populations.

One Sentence Summary: We here identify distinct myeloid differentiation pathways giving rise to the neutrophil-monocyte and basophil-mast cell-eosinophil lineages, respectively.

Introduction: Myeloid cell types carry out essential innate immune functions, with macrophages and neutrophils critical to antibacterial defense, eosinophils and basophils providing anti-
20 parasitic immunity, and mast cells acting as sentinels in the skin and gut, and playing a key role in regulating allergy (6). Defects in myeloid cells lead to immune deficiencies, whereas excessive production can cause severe inflammation-induced tissue degeneration (7, 8), and

myeloid cell types are critical regulators of inflammation (9) and the tumor microenvironment (10). Finally, malignancies affecting the myeloid lineages include acute myeloid leukemia, myeloproliferative neoplasms and myelodysplastic syndromes, all increasing in prevalence with population ageing. Consequently, the ontogeny of myeloid cells has been intensively studied, with particular emphasis on the identification of progenitor populations committed to specific myeloid fates, and the transcription factors (11) and cytokines (12) that regulate myeloid lineage commitment and differentiation. Initially, granulocyte and monocyte progenitors (GMPs) were identified in both murine (13) and human hematopoiesis (1), and proposed to constitute a single progenitor capable of forming all myeloid cell types. Subsequently, murine GMPs, were found to contain eosinophil, basophil and mast cell potential (14, 15). In contrast, in humans eosinophil potential was located to the CMP population (16), whereas the origin of human mast cell/basophil progenitors remains to be determined (17). We recently found that murine pre-GMs and GMPs could be sub-divided using a *Gata1*-EGFP reporter, and that the *Gata1*-EGFP⁺ and *Gata1*-EGFP⁻ sub-populations contained mast cell/eosinophil and monocyte/neutrophil potentials, respectively (5). Importantly, the separation of the two sets of myeloid lineage potentials preceded their separation from megakaryocyte/erythroid and lymphoid potentials, respectively. As also human myeloid cell types can be classified into *GATA1*-expressing (mast cells, eosinophils, basophils) and *GATA1*-non-expressing cell types (neutrophils, monocytes/macrophages) (18, 19), we here investigate whether human myelopoiesis is similarly organised. We find that basophil/mast cell potential resides within the human CMP, and using single cell RNA sequencing that CMPs can be prospectively sub-divided into CD114⁺ and CD131⁺ sub-populations that contain neutrophil/monocyte and eosinophil/mast cell/basophil potential, respectively. Furthermore, neutrophil/monocyte potential co-segregates with lymphoid

potential, whereas eosinophil/mast cell/basophil potential co-segregates with erythroid/megakaryocytic potential, with CD131+ CMPs as oligo-potent cells with combined eosinophil/mast cell/basophil and erythroid/megakaryocytic lineage potential. These findings lead to a revised model for human myelopoiesis, and show that initial lineage commitment in human hematopoiesis involves the generation of lineage-restricted oligo-potent progenitor populations containing defined subsets of lineage potentials.

Results: To localize human mast cell/basophil potentials we cultured single CMPs (defined as LIN-CD34+CD38+CD123+CD45RA-) and GMPs (LIN-CD34+CD38+CD123+CD45RA+) (Fig. 1A) in the presence of cytokines promoting the differentiation of all myeloid lineages (hSCF, hFLT3L, hIL-3, hIL-5; hIL-6, hGM-CSF, hG-CSF; M-conditions) and analysed the morphology of the cells generated. This showed that individual CMPs generated cultures containing eosinophils and basophils/mast cells (Fig. 1, C and E), or neutrophils and monocytes (Fig. 1, C and D), but only very rarely cells from both of these subsets (2/181 cultures; Fig. 1C). The majority of basophil/eosinophil-containing cultures (24/39 or 62%) were bi-lineage. In contrast, GMPs only generated neutrophils and monocytes under these conditions (Fig. 1, C and F). In the mouse we found that LMPPs generate neutrophils and monocytes, but not eosinophils or mast cells (5). We therefore analysed the myeloid lineages generated by LMPPs (defined as LIN-CD34+CD38-CD45RA+ (20)) (Fig. 1B), and observed only neutrophils and monocytes (Fig. 1, C and G).

Human mast cells/basophil progenitors therefore reside within the phenotypically defined CMP compartment, as previously reported for eosinophil progenitors (16). In addition, the observed

combined eosinophil and mast cell/basophil output indicated the presence of a common progenitor for these lineages, distinct from progenitors with monocyte/neutrophil potential, within the CMP compartment. To de-convolute these functionally distinct progenitor subsets we performed Smart-seq2-based single cell RNA sequencing (21) of index sorted CMPs (237 cells after final QC), including GMPs (32 cells) and MEPs (27 cells) as reference populations (Fig. 2A). T-distributed stochastic neighbour embedding (t-SNE) analysis (22) followed by kernel density estimation and *k*-mean clustering analysis identified seven cellular clusters (Fig. 2B). MEPs and GMPs primarily mapped to cluster (C)1 and C6, respectively, indicating that these represent megakaryocyte/erythroid- and neutrophil/monocyte-committed CMP sub-populations. Consistent with this, *GATA1* and *KLF1* were highly expressed in C1-C3 and C7, but not C4-C6 (Fig. 2, C and D), with *CEBPA* expressed in C4-C6, and C7 (Fig. 2E). We previously observed co-expression of *Cebpa*, *Klf1* and *Gata1* in murine progenitors with mast cell/eosinophil potential, whereas neutrophil/monocyte progenitors expressed neither *Gata1* nor *Klf1*. Furthermore, *Cebpa* expression was lost in committed erythroid progenitors, where *Klf1* and *Gata1* expression were maintained (5). We therefore hypothesized that C4-C6 were candidate neutrophil/monocyte progenitors, whereas C7 was a candidate eosinophil/basophil/mast cell progenitor population.

To allow further characterization of these sub-populations we mined the RNA sequencing data for differentially expressed genes encoding surface markers (Table S2). 10 candidates were identified (Fig. 3A and Fig. S1A), and the encoded surface markers tested for their ability to subdivide the CMP population (Fig. S1, B and C). We observed that FcεR1α, CD114 and CD131 each labeled a sub-fraction of CMPs. Of these CD131 and CD114 were of particular interest, as

genes that encode them (*CSF2RB* and *CSF3R*, respectively) were expressed in *GATA1* expressing (C1-C3, C7) and non-expressing (C4-C6) clusters, respectively (Fig. 3A). The CD133 marker (encoded by *PROM1*) previously used to separate myeloid progenitor subsets (23) showed an expression pattern similar to CD114 by both gene expression (Fig. S1A) and flow cytometry (Fig. S2A), but was less specific for C4-C6. We depicted CD131 and CD114 expression on MEP, GMP and more immature populations. CD131 expression is not observed in HSC/MPP or LMPP, but is sustained on MEPs. CD114 is expressed at low levels in both the HSC/MPP and LMPP subsets, but up-regulated in CD114⁺ CMPs and sustained in GMPs (Fig S2B).

In order to test the ability of these markers to prospectively separate the identified clusters, signature genes selectively expressed in each cluster were identified (Fig. S3A and Table S2). We found that C4 and C5 were separated only by distinct cell cycle status (Fig. S3, B and C), and common signature genes were therefore identified for these two clusters. C2 was too weak to generate signature genes. To test if CD114 and CD131 surface expression identified the predicted clusters single CMPs were index sorted on CD114 and CD131 (Fig. 3B) and expression of the above identified cluster signature genes was analysed by microfluidics-based qRT-PCR. As predicted, C1 and C7 signature genes were expressed in CD114⁻CD131⁺ CMPs (henceforth CD131⁺ CMPs), and C4/5 and C6 signatures in CD114⁺CD131⁻ CMPs (henceforth CD114⁺ CMPs) (Fig. 3C), validating the ability of these markers to prospectively separate clusters with molecular characteristics of neutrophil/monocyte progenitors (C4/5+C6) and eosinophil/mast cell progenitors (C7). The CD114⁻CD131⁻ CMPs were a mixed population containing predominantly cells expressing C4/5 signature genes, as well as some C1/3 and C7

signature gene-expressing cells. The CD114⁻CD131⁻ double negative CMP fraction therefore did not contain distinct cell types, but rather cells with functional characteristics of CD114⁺ CMPs or CD131⁺ CMPs that could not be separated using these markers.

To efficiently separate the GATA1-expressing and non-expressing clusters we therefore focused on CD114⁺ and CD131⁺ CMPs. These two populations were sorted (Fig. 3D), bulk cells cultured under M-conditions and analysed for the generation of myeloid cell types by flow cytometry (Fig. 4A). We observed that CD114⁺ CMPs generated predominantly neutrophils and monocytes (>98%; Fig. 4B), whereas CD131⁺ CMPs generated predominantly eosinophils, mast cells and basophils (>92%; Fig. 4B), demonstrating that CD114 and CD131 expression can be used for the prospective separation of these two sets of myeloid lineage potentials. Also in this assay GMPs generated essentially only neutrophils and monocytes (Fig. 4B), and these cultures were less proliferative than the CD114⁺ CMPs (Fig. 4C).

To co-localize lineage potentials at the single cell level we index sorted CMPs followed by single cell culture under conditions supporting megakaryocyte and erythroid lineage output (MME conditions; see methods). In order to stringently define lineage readout all cultures were subjected to both morphological analysis and microfluidics-based expression analysis of key lineage-specific genes, and only readouts supported by both criteria considered positive, except for megakaryocytes, which could be reliably identified by morphology alone. At least one positive lineage-readout was obtained from 243 cultures (Fig. 5A). We observed that megakaryocyte/erythroid lineage output was highly dissociated from neutrophil/monocyte output ($P=2 \times 10^{-14}$, χ^2 test), but not from eosinophil/basophil/mast cell lineage output ($P=0.22$, χ^2 test)

(Fig. 5B). Mapping single cells to CD114 and CD131 expression confirmed that CD131⁺ CMPs were restricted to mast cell/basophil/eosinophil output, whereas CD114⁺ CMPs generated neutrophils and monocytes (Fig. 5C). Separate analysis of all *GATA1*⁺ lineages showed that the majority of cultures were bi- or oligo-lineage under MME conditions (129/160 or 81%; Fig. 5D), including megakaryocyte/erythroid (Fig. 5E) and megakaryocyte/erythroid/basophil (Fig. 5F) readouts. However, compared to the data obtained under M-conditions (Fig. 1) we noticed that eosinophil lineage output was reduced. We therefore repeated the analysis, using the same donor bone marrows, under M-conditions. Here neutrophil and eosinophil readouts, and consequently bi-lineage myeloid readouts, were improved in M-condition compared to MME-condition cultures (Fig. S4A,B). Importantly, under both conditions a significant dissociation of neutrophil/monocyte and mast cell/basophil/eosinophil readouts was observed (M- conditions: $P=0.0014$; MME conditions: $P=1.2 \times 10^{-7}$, χ^2 test). While this dissociation was absolute under MME conditions, under M-conditions we observed rare cultures where mast cell/basophil/eosinophil potential was combined with neutrophil potential (but not monocyte potential) (Fig. S4B-D), an observation similar to that previously made in murine preGMs (5). Therefore, optimal measurement of myeloid and myeloid-erythroid co-potentials may require distinct culture conditions, and sub-optimal conditions lead to an underestimate of the potency of the cells analysed. Finally, to determine the relationship of the two myeloid lineage subsets to lymphoid lineages we measured lymphoid lineage output of LMPPs, GMPs, CD114⁺ CMPs and CD131⁺ CMPs in MS5 (for B- and NK cell readout) and OP9-hDL1 co-cultures (for T-cell readout). We observed B-, T- and NK cell output from populations with neutrophil/monocyte potential (LMPP, CD114⁺ CMP and GMP), but not from CD131⁺ CMPs (Fig. 5G; Fig. S4E).

To molecularly characterize the identified CMP subpopulations we performed Smart-seq2-based RNA sequencing of bulk cell populations from 4 independent bone marrow donors. In addition to CD114+ CMPs and CD131+ CMPs we also profiled GMPs, MEPs, LMPPs and HSCs/MPPs (defined as LIN-CD34+CD38-CD45RA-) (Fig. S5). We validated that the purified CMP subpopulations expressed *CSF3R* and *CSF2RB*, respectively (Fig. 6A). In principal component analysis CD131+ CMPs clustered with MEPs, whereas CD114+ CMPs clustered with GMPs and adjacent to LMPPs, consistent with these populations defining distinct developmental pathways (Fig. 6B). Similarly, expression of mast cell/basophil (Fig. 6C), eosinophil (Fig. 6D) and megakaryocyte/erythroid (Fig. 6E) genes was observed in CD131+ CMPs and MEPs, whereas neutrophil/monocyte genes were expressed in CD114+ CMPs and GMPs (Fig. 6F). Comparison of CD114+ and CD131+ CMPs using Gene Set Enrichment Analysis (GSEA (24)) showed that the former were enriched for neutrophil, monocyte and lymphoid (B cell, T cell and NK cell) gene expression (Fig. S6A,B), whereas basophil and eosinophil genes were enriched in the latter (Fig. S6C). The gene expression signatures of the CMP sub-fractions therefore correspond to their lineage potentials. Hematopoietic lineage specification is controlled by the combinatorial action of key transcription factors (TF) (11). CD131+ CMPs express TFs associated with both erythroid/megakaryocytic differentiation (*KLF1*, *GATA1*; Fig. 6G) and with specification of *GATA1*+ myeloid cell types (*CEBPA*, *SPI1*; Fig. 6I). Expression of myeloid TFs (*SPI1*, *CEBPA*) was decreased upon MEP specification, concomitant with expression of *ZFPM1* and its target gene *TRIB2* (25), a negative regulator of C/EBPs (26, 27) (Fig. 6H), and paralleled by down-regulation of basophil and eosinophil programs (Fig. S6D) consistent with the loss of myeloid lineage potentials in this population (1). In contrast, CD114+ CMPs and GMPs expressed genes involved in monocyte/neutrophil specification (*GFIL*, *IRF8*; Fig. 6J) as well as general myeloid

lineage specification (*SP11*, *CEBPA*), but lack expression of *GATA1* (Fig. 6G). GSEA analysis showed that neutrophil and monocyte/macrophage genes were expressed at higher levels in GMPs compared to CD114⁺ CMPs (Fig. S6E), consistent with their lower expression of the transcription factors driving these differentiation programs (*IRF8*, *GF11*, *CEBPA* and *SP11*), and indicating that CD114⁺ CMPs represent an earlier, committed stage of neutrophil/monocyte development than GMPs. In line with the lymphoid potentials found in the populations, T-cell (*ZAP70*), B-cell (*IGLL1*), NK cell (*NKG7*) and pan-lymphoid (*CD96*) representative genes are all preferentially expressed in the lymphoid-competent progenitor subsets (LMPPs, CD114⁺ CMPs and GMPs), when compared to CD131⁺ CMPs and MEPs (Fig 6K).

Discussion: We have here used a combination of single cell RNA sequencing and single cell functional readouts to assess the lineage potentials of individual human bone marrow progenitor cells. Importantly, we have assessed all myeloid cell fates, and combined morphological, flow cytometric and gene expression analysis to obtain coherent and reliable identification of the cell types generated. We observe that when assessed *in vitro* at the single cell level CD34⁺CD38⁺ human progenitor cells possess either neutrophil/monocyte or eosinophil/basophil/mast cell potential, but virtually never a combination of these. We also find that restriction to neutrophil/monocyte myeloid fate occurs in LMPPs, where lymphoid lineage readout was previously found to co-occur with neutrophil and/or monocyte readout in the vast majority of single cells analysed (2). Conversely, the majority of CD131⁺ CMPs that generate basophil/mast cell/eosinophils also produce megakaryocyte or erythroid lineage cells, but not neutrophils, monocytes or lymphoid cells (this study). Together with previous findings that GMPs (1) and LMPPs (20) lack megakaryocyte/erythroid lineage potential these results support a hierarchical

model of adult human bone marrow hematopoiesis where *GATAI*⁻ (neutrophils and monocytes) and *GATAI*⁺ myeloid lineages (mast cells/basophils/eosinophils) separate before they segregate from lymphoid and megakaryocyte/erythroid lineages, respectively (Fig. S7).

5 This model has several key features in common with our current understanding of murine hematopoiesis (5). In particular, in neither mouse nor human has a myeloid lineage-committed progenitor containing all myeloid lineage potentials at the single cell level been identified. Instead, progenitors restricted to neutrophil/monocyte differentiation can be prospectively isolated in both species (human CD114⁺ CMP and GMP, respectively, corresponding to murine 10 *Gata1*⁻ preGM and GMP (5)), as can progenitors restricted to mast cell/basophil and eosinophil myeloid differentiation (CD131⁺ CMP, *Gata1*⁺ GMP). Single cell RNA-sequencing of human myeloid progenitor cells has been used to computationally generate trajectories leading to eosinophil and mast cell/basophil progenitor formation, variably predicting that these cell types co-segregate with neutrophil/monocyte (4) or erythroid lineage progenitors (28). Our data would 15 be largely compatible with the latter model. Our model is not readily compatible with the previously observed segregation of neutrophil (CD133⁺) and basophil/eosinophil (CD133⁻/lo) progenitors, where monocytes were generated from both populations (23). The use of two markers (CD114 and CD131) and the greater selectivity of CD114 expression for NM-restricted progenitors likely explain why we obtain strict segregation of neutrophil-monocyte from 20 basophils-mast cell-eosinophil, as well as megakaryocyte-erythroid potential.

Recent studies have proposed direct commitment of multi-potent HSCs/MPPs to uni-lineage fate in adult bone hematopoiesis, based on the inability to detect oligo-potent cells with combined

myeloid and erythroid/megakaryocytic lineage potential functionally (3), or computationally based on co-expression of lineage-affiliated genes in single cells (4). However, we here observe combined megakaryocyte/erythroid and myeloid readout from >15% of human bone marrow CMPs. One possible explanation for this discrepancy is that our assay conditions and analysis were optimized for the detection of basophils/mast cells and eosinophils, which are not readily detected by the CD11b antibody used by Notta et al. to identify myeloid cells (Fig. S8). While this remains to be clarified it underscores the importance of evaluating the output of myeloid lineages individually rather than as a whole, since they derive from two distinct progenitor pathways, and therefore cannot be used as proxies for each other.

Our results establish the existence of two complementary oligo-potent progenitors in human hematopoiesis: one containing neutrophil/monocyte as well as lymphoid lineage potentials (the previously described LMPP (20, 29)), and one capable of generating megakaryocytes, erythroid cells and the GATA-1-expressing myeloid cell types (basophils/mast cells and eosinophils), here designated EMPP (erythroid-megakaryocyte primed multipotent progenitor; corresponding to the CD131+ CMP). Importantly, as discussed above, in both these populations myeloid potentials co-localize with other potentials at the single cell level, demonstrating true oligo-potentiality. These observations are therefore consistent with an early lineage bifurcation that generates *GATA1*⁺ and *GATA1*[−] progenitor domains Fig S7, with EMPPs and LMPPs, respectively, defining the entry points, similarly to what has been proposed for murine hematopoiesis (5). Importantly, however, our data do not exclude that direct commitment to individual lineages also occurs, a notion supported by the identification of murine HSCs that are fate-restricted to platelet lineage output (30-32). Finally, LMPPs and GMPs have been identified as target cells for

transformation of the neutrophil/monocyte lineages (20), and it will now be relevant to investigate the role of the herein identified eosinophil and basophil/mast cell lineage-restricted progenitor populations in sustaining malignancies of these lineages.

5 **Materials and methods:**

Study Design

The aim of this study was to determine heterogeneity within the CMP population of healthy adult human bone marrow cells and to prospectively isolate subpopulations for investigating their lineage potentials and thus providing evidence for independent pathways towards different myeloid cell types. Progenitor cells were analysed by gene expression and FACS, and cell potentials were tested by in vitro cultures. Sample size, replicates, number of experiments, statistical analysis and donor sample information is specified in figure legends and in this section.

15 **Human bone marrow cells**

Bone marrow samples were from AllCells® or taken from healthy male volunteers at the age of 21-29 years who provided written informed consent in accordance with local guidelines established by and with the approval of the local Ethics Committee of the Cities of Copenhagen and Frederiksberg. Mononuclear cells were isolated using Ficoll density gradient. Cryopreserved mononuclear cells were thawed and processed for flow cytometry as previously described (33). Where possible and relevant, experiments were repeated with cells from at least two donors (Table S3).

Flow cytometry

For flow cytometry and cell sorting a BD LSRFortessa, LSRII, a FACS Aria II, FACS Aria III and FACS Aria Fusion (BD Biosciences) were used. FlowJo analysis software was used for subsequent data analysis. All antibody stainings were preceded by incubation with human FcR Blocking Reagent (Miltenyi Biotec, 130-059-901) at a 1/10 dilution. For antibody stainings that included purified CD131, cells were first stained with purified CD131, followed by BV421™ anti-mouse IgG1. The cells were then resuspended in 500 µg/ml Purified Mouse IgG (PMP01X, Bio-Rad) and after 5-10 minutes an equal volume containing relevant antibodies was added. In all flow experiments, 7-AAD (40037, Biotium) was used at a final concentration of 1 µg/ml to exclude dead cells. Antibodies, suppliers and dilutions used are listed in Table S4. Populations were defined as follows: Lineage cocktail: CD2, CD3, CD4, CD7, CD8, CD10, CD11b, CD14, CD19, CD20, CD56, CD235a. HSC/MPP: LIN⁻CD34⁺CD38⁻CD45RA⁻, LMPP: LIN⁻CD34⁺CD38⁻CD45RA⁺, CMP: LIN⁻CD34⁺CD38⁺CD45RA⁻CD123⁺, MEP: LIN⁻CD34⁺CD38⁺CD45RA⁻CD123⁻, GMP: LIN⁻CD34⁺CD38⁺CD45RA⁺CD123⁺. Cultured cells: monocytes: CD14⁺CD15⁻, neutrophils: CD14^{+/lo}CD15⁺Siglec8⁻CCR3⁻CD117⁻FcεR1α⁻, mast cells: CD14⁻CD15^{-/lo}Siglec8⁻CD117⁺FcεR1α⁺, basophils: CD14⁻CD15^{-/lo}Siglec8⁻CD117⁻FcεR1α⁺, eosinophils: CD14⁻CD15^{-/lo}Siglec8⁺CCR3⁺, B cells: CD45⁺CD14⁻CD15⁻CD19⁺CD56⁻, NK cells: CD45⁺CD14⁻CD15⁻CD19⁻CD56⁺, T cells: CD45⁺CD7^{hi}.

Generation cDNA libraries using Smart-seq2 protocol

Single cells or bulk cells (100 cells) were isolated by FACS into 96-well plates (Thermo, single cells) or Eppendorf tubes (bulk) containing 4µl of a lysis mix, consisting of 0.2% triton, 4U RNase inhibitor (Takara), 2.5µM oligo-dT₃₀VN (Biomers) and 2.5mM dNTP mix (Fermentas).

This was stored at -80°C for up to one week. For the reverse transcriptase, 6µl of the following mix was added: 2µl Superscript II first strand buffer, 0.5µl DTT (100mM), 2µl Betaine (5M), 0.1µl MgCl₂ (1M), 0.25µl RNase Inhibitor (40U/µl), 0.1µl TSO (100 µM), 0.25µl Superscript II RT (200U/µl) and 0.8µl (single cells) or 0.4µl (bulk) water. After reverse transcriptase, 15µl was added containing 12.5µl Kapa Hifi HS Ready Mix (2x) and 0.125µl ISPCR primers (10µM). The thermal conditions for RT and pre-amplification were according to the original Smart-seq2 protocol (21). The number of cycles used for PCR amplification was 22 for single cells and 16 for bulk samples. After PCR amplification, cDNA libraries were purified using Ampure XP magnetic beads according to the manufacturer's instructions. After purification, the libraries were resuspended in 17.5µl of buffer EB (Qiagen) and stored at -20°C. Quality and concentration of the cDNA libraries generated was assessed using High-Sensitivity Bioanalyzer (Agilent).

Illumina library preparation and sequencing

From single cells, 1.25µl of cDNA was used and from bulk samples, 0.7ng of cDNA was tagged using the Nextera XT DNA Sample Preparation kit (Illumina) according to the manufacturer's instructions, except that one-fourth of the volumes indicated were used. Purification of the product was done with a 1:1 ratio of AMPure XP beads, with a final elution in 17.5µl in resuspension buffer provided from the Nextera kit. Samples were loaded on a High-Sensitivity DNA chip (Agilent Technologies) to check the size and quality of the indexed library, and the concentration was measured with Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) on a CLARIOstar® (BMG LABTECH) or with Qubit High-Sensitivity DNA kit (Invitrogen). Libraries were pooled to a final concentration of 4nM and were sequenced with Illumina

NextSeq 500 (76bp single-end read) after preparation according to manufacturer's instructions. For the bulk samples, gene expression data was accumulated from two sequence runs.

Single-cell RNA sequencing analysis

Short reads were aligned to the human (GRCh37/hg19) genome using Tophat (v2.0.13) (34). The mapping parameter '-g 1' was used to allow one alignment to the reference for a given read. Cells with <500,000 mapped reads, with the percentage of mapping to the mitochondrial chromosome >10% or <2,000 detected genes were excluded from further analysis. 296 cells (237 CMP, 32 GMP and 27 MEP) fulfilled these criteria. The featureCounts (35) software was used to count reads on the basis of the RefSeq gene model. Counts Per Million (CPM) values were calculated using a script in R. We normalized the CPM values into \log_2 (CPM) scale and set up the limit of detection at 1 CPM. Log2 scale of genes expressed at <1 CPM was set to 0. We selected 4,731 predicted variable genes on the basis of a simple noise model using the Lowess model of the average gene expression and the coefficient of variation (CV) (36). We then performed PCA analysis. Top 100 genes, without the cell cycle related genes, with the highest absolute correlation coefficient (PCA component loadings, one of the first three components) were used for the T-distributed stochastic neighbor embedding (tSNE) analysis. The Rtsne package, a barnes-hut implementation, was used to perform the tSNE analysis. We used the kde function in the ks package to perform the kernel smoothing density estimation using the basis of tSNE analysis results (from dimensions 1 and 2) to visualise high density regions of data points. We estimated 7 high-density regions. We then used the centers of each high-density region as the input for the *K*-means clustering analysis ($k = 7$) to assign the cell clusters. Differentially expressed genes analysis was performed using nonparametric Wilcoxon test for the expression

level and Fisher's exact test for the expressing cell frequency. *P* values generated from both tests were then combined using Fisher's method and were adjusted using Benjamini–Hochberg. Differentially expressed genes were selected on the basis of the absolute log₂ fold change of >2 and the adjusted *P* value of <0.05. Core cell-cycle genes from S and G2M phases were previously described (37).

Bulk RNA sequencing analysis

The same mapping and gene expression quantification procedures were performed as described in the single-cell analysis. RPKM values were calculated using a script in R. 1,868 genes were selected for PCA analysis based on the high variation of gene expression across populations (analysis of variance with adjusted *P* values of <0.05 and coefficient of variation equal to or larger than 0.3). Differentially expressed gene analysis was performed on the read count using DESeq2 (38).

Gene set enrichment analysis

Gene set enrichment was performed using GSEA v3.0 (<http://software.broadinstitute.org/gsea/index.jsp>) and human cell type-specific gene sets for neutrophils, monocytes, macrophages, basophils and eosinophils (<http://nano.princeton.edu/standard/>). Human gene sets for NK cells, B cells and T cells were generated by taking the top 100 genes identified as specific for these cell types by Novershtern et al. (39) (Table S5). The monocyte gene set was >500 genes, and was therefore divided into two set and tested individually with similar results. A representative analysis is shown.

Myeloid cell cultures

To test myeloid potential, single cells were cultured in Terasaki plates in 20µl of IMDM with L-glutamine (Gibco), 20% HyClone™ Defined Fetal Bovine Serum (SH30070.03, GE Healthcare), penicillin/streptomycin (Invitrogen) and 0.1µM β-mercaptoethanol (Sigma), supplemented with hSCF (20ng/ml), hFlt3L (20ng/ml), hIL-3 (20ng/ml), hIL-5 (50ng/ml), hIL-6 (20ng/ml), hGM-CSF (50ng/ml), hG-CSF (20ng/ml). For bulk cultures, 50-150 cells were cultured in 400µl of the same culture medium in 48 well plates. For combined erythroid, megakaryocyte and myeloid readout, single cells were cultured in round bottom 96 well plates in 50µl of Stemspan (Stemcell Technologies) with hSCF (20ng/ml), hFlt3L (20ng/ml), hIL-3 (20ng/ml), hIL-5 (50ng/ml), hIL-6 (20ng/ml), hGM-CSF (50ng/ml), hG-CSF (20ng/ml), hLDL (Sigma-Aldrich) (40µg/ml), EPO (0.5U/ml) and TPO (100ng/ml). Cells were cultured at 37°C, 5% CO₂. Cytokines, suppliers and concentrations used are found in Table S6. Cytospins were prepared with a Shandon cytospin at 1000 RPM with low acceleration, followed by May-Grünwald-Giemsa stain (VWR).

Lymphoid cultures

For NK and B cell potential, cultures of 30 or 5 cells were grown in gelatin coated 24 well plates seeded with MS-5 feeder cells in MEM α, GlutaMAX™ Supplement (32561, Gibco/Life Technologies) supplemented with 10% HyClone™ Defined Fetal Bovine Serum (SH30070.03, GE Healthcare), 1% penicillin-streptomycin, 1% L-glutamine, 20ng/ml hSCF, 10ng/ml hFLT3L, 10ng/ml hIL-2 and 10ng/ml hIL-15. Half of the medium was changed every week. Cultures were analysed by flow cytometry after three weeks. Cultures with more than 8 cells in the respective gates were considered positive for NK or B cells. For T cell potential, cultures of 30 or 5 sorted cells were grown in gelatin-coated 24 well plates seeded with OP9hΔ1 feeder cells in freshly

prepared α -MEM medium (Gibco/Thermo Fisher Scientific, 12000-063) with 20% HyClone™ Defined Fetal Bovine Serum, 1% penicillin-streptomycin, 1% L-glutamine, 10 ng/ml hSCF, 5ng/ml hFLT3L and 5ng/ml hIL-7. Every week cells were transferred to new plates with feeder cells and fresh medium. Cultures were analysed by flow cytometry after five weeks. Cultures with more than 50 cells in the respective gates were considered positive T cells.

Quantitative PCR

For single cells, the CellsDirect One-Step qRT-PCR kit (Life technologies, 11753-100) was used according to manufacturer's protocol for preparation and amplification of cDNA. The BioMark 192.24 Dynamic Array platform (Fluidigm) and Taqman assays (Table S7) were used according to the manufacturers' instructions. Δ Ct values were zero centered for each gene by subtraction of the mean value of all positive cells for the gene. These normalized values were used to generate a heat map using the web-based tool Morpheus (<https://software.broadinstitute.org/morpheus/>). For gene expression of cultured cells, medium of the cultures was removed and the cells were re-suspended in 15 μ l CellDirect 2x Reaction Mix containing 0.2U/ μ l SUPERase In™ RNase Inhibitor (AM2694) and placed at -80°C. Preparation of cDNA and preamplification was done with 2.5 μ l of the lysed cells, using 22 pre-amplification cycles in a total volume of 5 μ l. This was diluted 50 times for further gene expression using the BioMark 192.24 Dynamic Array platform (Fluidigm) and Taqman assays according to the manufacturers' instructions. Δ Ct values relative to *HPRT* ($Ct(HPRT) - Ct(Gene)$) were zero-centred for each gene by subtraction of the mean value of all positive cells for the gene. Single cell-derived cultures were divided with one half used for cytopsin and one half for quantitative PCR. A positive lineage readout was defined as both morphologically differentiated cells and signature gene expression being observed, with

positive readout defined as $Ct(IRF8) \leq Ct(HPRT)$ for monocytes, $Ct(CSF3R) \leq Ct(HPRT)$ for neutrophils, $Ct(HDC) \leq Ct(HPRT)$ for mast cells/basophils, $Ct(IL5RA) \leq Ct(HPRT)$ and $Ct(EPX) \leq Ct(HPRT)$ for eosinophils, and $Ct(KLF1) \leq Ct(HPRT)+4$, $Ct(AHSP) \leq Ct(HPRT)+4$ and $Ct(CA) \leq Ct(HPRT)+4$ for erythroid cells.

5

Statistical analysis

Significance of differences in distribution of lineage potentials in single cells was analysed using the χ^2 test. For the identification of differentially expressed genes from single cell RNA sequencing data a combination of the Wilcoxon and Fisher's exact test were used, combined
10 significance calculated using Fisher's method and P-values adjusted using Benjamini–Hochberg.

List of Supplementary Materials:

Fig. S1. Candidate cell surface markers for separation of myeloid progenitor subsets.

Fig. S2. Expression of CD131, CD114 and CD133 on hematopoietic stem- and progenitor
15 subsets.

Fig. S3. Single cell heatmap of clustered CMPs.

Fig. S4. Influence of culture conditions on lineage readout.

Fig. S5. Gating strategy for isolation of human stem- and progenitor cell subsets.

Fig. S6. Myeloid and lymphoid potential of progenitor populations is reflected in their gene
20 expression.

Fig. S7. Proposed model of the human hematopoietic hierarchy.

Fig. S8. Expression of CD11b on myeloid cell types generated *in vitro*.

Table S1. Cloning efficiencies of myeloid progenitors.

Table S2. Genes differentially expressed between CMP clusters.

Table S3. Donor samples.

Table S4. Antibodies used for flow cytometry and FACS.

Table S5. Gene sets used for gene set enrichment analysis.

5 Table S6. Cytokines used for progenitor assays.

Table S7. TaqMan probes for qRT-PCR.

References and notes.

1. M. G. Manz, T. Miyamoto, K. Akashi, I. L. Weissman, Prospective isolation of human
10 clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A* **99**, 11872-11877 (2002).
2. D. Karamitros, B. Stoilova, Z. Aboukhalil, F. Hamey, A. Reinisch, M. Samitsch, L.
Quek, G. Otto, E. Repapi, J. Doondeea, B. Usukhbayar, J. Calvo, S. Taylor, N. Goardon, E. Six,
F. Pflumio, C. Porcher, R. Majeti, B. Gottgens, P. Vyas, Single-cell analysis reveals the
continuum of human lympho-myeloid progenitor cells. *Nat Immunol* **19**, 85-97 (2018).
- 15 3. F. Notta, S. Zandi, N. Takayama, S. Dobson, O. I. Gan, G. Wilson, K. B. Kaufmann, J.
McLeod, E. Laurenti, C. F. Dunant, J. D. McPherson, L. D. Stein, Y. Dror, J. E. Dick, Distinct
routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**,
aab2116 (2016).
- 20 4. L. Velten, S. F. Haas, S. Raffel, S. Blaszkiewicz, S. Islam, B. P. Hennig, C. Hirche, C.
Lutz, E. C. Buss, D. Nowak, T. Boch, W. K. Hofmann, A. D. Ho, W. Huber, A. Trumpp, M. A.
Essers, L. M. Steinmetz, Human haematopoietic stem cell lineage commitment is a continuous
process. *Nat Cell Biol* **19**, 271-281 (2017).

5. R. Drissen, N. Buza-Vidas, P. Woll, S. Thongjuea, A. Gambardella, A. Giustacchini, E. Mancini, A. Zriwil, M. Lutteropp, A. Grover, A. Mead, E. Sitnicka, S. E. W. Jacobsen, C. Nerlov, Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol* **17**, 666-676 (2016).

6. L. Koenderman, W. Buurman, M. R. Daha, The innate immune response. *Immunol Lett* **162**, 95-102 (2014).

7. M. G. Manz, S. Boettcher, Emergency granulopoiesis. *Nat Rev Immunol* **14**, 302-314 (2014).

8. H. Sadrzadeh, O. Abdel-Wahab, A. T. Fathi, Molecular alterations underlying eosinophilic and mast cell malignancies. *Discov Med* **12**, 481-493 (2011).

9. A. Ortega-Gomez, M. Perretti, O. Soehnlein, Resolution of inflammation: an integrated view. *EMBO Mol Med* **5**, 661-674 (2013).

10. D. I. Gabrilovich, S. Ostrand-Rosenberg, V. Bronte, Coordinated regulation of myeloid cells by tumours. *Nat Rev Immunol* **12**, 253-268 (2012).

11. S. H. Orkin, L. I. Zon, Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644 (2008).

12. D. Metcalf, Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation. *Blood* **92**, 345-347; discussion 352 (1998).

13. K. Akashi, D. Traver, T. Miyamoto, I. L. Weissman, A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-197 (2000).

14. H. Iwasaki, S. Mizuno, R. Mayfield, H. Shigematsu, Y. Arinobu, B. Seed, M. F. Gurish, K. Takatsu, K. Akashi, Identification of eosinophil lineage-committed progenitors in the murine bone marrow. *J Exp Med* **201**, 1891-1897 (2005).

15. Y. Arinobu, H. Iwasaki, M. F. Gurish, S. Mizuno, H. Shigematsu, H. Ozawa, D. G.

5 Tenen, K. F. Austen, K. Akashi, Developmental checkpoints of the basophil/mast cell lineages in adult murine hematopoiesis. *Proc Natl Acad Sci U S A* **102**, 18105-18110 (2005).

16. Y. Mori, H. Iwasaki, K. Kohno, G. Yoshimoto, Y. Kikushige, A. Okeda, N. Uike, H.

10 Niiro, K. Takenaka, K. Nagafuji, T. Miyamoto, M. Harada, K. Takatsu, K. Akashi, Identification of the human eosinophil lineage-committed progenitor: revision of phenotypic definition of the human common myeloid progenitor. *J Exp Med* **206**, 183-193 (2009).

17. O. Schmetzer, P. Valentin, M. K. Church, M. Maurer, F. Siebenhaar, Murine and human mast cell progenitors. *Eur J Pharmacol* **778**, 2-10 (2016).

18. L. I. Zon, Y. Yamaguchi, K. Yee, E. A. Albee, A. Kimura, J. C. Bennett, S. H. Orkin, S. J. Ackerman, Expression of mRNA for the GATA-binding proteins in human eosinophils and
15 basophils: potential role in gene transcription. *Blood* **81**, 3234-3241 (1993).

19. M. Babina, Y. Schulke, L. Kirchhof, S. Guhl, R. Franke, S. Bohm, T. Zuberbier, B. M. Henz, A. F. Gombart, The transcription factor profile of human mast cells in comparison with monocytes and granulocytes. *Cell Mol Life Sci* **62**, 214-226 (2005).

20. N. Goardon, E. Marchi, A. Atzberger, L. Quek, A. Schuh, S. Soneji, P. Woll, A. Mead,

20 K. A. Alford, R. Rout, S. Chaudhury, A. Gilkes, S. Knapper, K. Beldjord, S. Begum, S. Rose, N. Geddes, M. Griffiths, G. Standen, A. Sternberg, J. Cavenagh, H. Hunter, D. Bowen, S. Killick, L. Robinson, A. Price, E. Macintyre, P. Virgo, A. Burnett, C. Craddock, T. Enver, S. E.

Jacobsen, C. Porcher, P. Vyas, Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138-152 (2011).

21. S. Picelli, A. K. Bjorklund, O. R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10**, 1096-1098 (2013).

22. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

23. A. Gorgens, S. Radtke, M. Mollmann, M. Cross, J. Durig, P. A. Horn, B. Giebel, Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. *Cell Rep* **3**, 1539-1552 (2013).

24. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

25. E. Mancini, A. Sanjuan-Pla, L. Luciani, S. Moore, A. Grover, A. Zay, K. D. Rasmussen, S. Luc, D. Bilbao, D. O'Carroll, S. E. Jacobsen, C. Nerlov, FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. *Embo J* **31**, 351-365 (2012).

26. K. Keeshan, Y. He, B. J. Wouters, O. Shestova, L. Xu, H. Sai, C. G. Rodriguez, I. Maillard, J. W. Tobias, P. Valk, M. Carroll, J. C. Aster, R. Delwel, W. S. Pear, Tribbles homolog 2 inactivates C/EBPalpha and causes acute myelogenous leukemia. *Cancer Cell* **10**, 401-411 (2006).

27. T. Naiki, E. Saijou, Y. Miyaoka, K. Sekine, A. Miyajima, TRB2, a mouse Tribbles ortholog, suppresses adipocyte differentiation by inhibiting AKT and C/EBPbeta. *J Biol Chem* **282**, 24075-24082 (2007).
28. S. Zheng, E. Papalexi, A. Butler, W. Stephenson, R. Satija, Molecular transitions in early progenitors during human cord blood hematopoiesis. *Mol Syst Biol* **14**, e8041 (2018).
29. J. Adolfsson, R. Mansson, N. Buza-Vidas, A. Hultquist, K. Liuba, C. T. Jensen, D. Bryder, L. Yang, O. J. Borge, L. A. Thoren, K. Anderson, E. Sitnicka, Y. Sasaki, M. Sigvardsson, S. E. Jacobsen, Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295-306 (2005).
30. A. Grover, A. Sanjuan-Pla, S. Thongjuea, J. Carrelha, A. Giustacchini, A. Gambardella, I. Macaulay, E. Mancini, T. C. Luis, A. Mead, S. E. Jacobsen, C. Nerlov, Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun* **7**, 11075 (2016).
31. J. Carrelha, Y. Meng, L. M. Kettyl, T. C. Luis, R. Norfo, V. Alcolea, H. Boukarabila, F. Grasso, A. Gambardella, A. Grover, K. Hogstrand, A. M. Lord, A. Sanjuan-Pla, P. S. Woll, C. Nerlov, S. E. W. Jacobsen, Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106-111 (2018).
32. A. E. Rodriguez-Fraticelli, S. L. Wolock, C. S. Weinreb, R. Panero, S. H. Patel, M. Jankovic, J. Sun, R. A. Calogero, A. M. Klein, F. D. Camargo, Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216 (2018).

33. P. S. Woll, U. Kjallquist, O. Chowdhury, H. Doolittle, D. C. Wedge, S. Thongjuea, R. Erlandsson, M. Ngara, K. Anderson, Q. Deng, A. J. Mead, L. Stenson, A. Giustacchini, S. Duarte, E. Giannoulatou, S. Taylor, M. Karimi, C. Scharenberg, T. Mortera-Blanco, I. C. Macaulay, S. A. Clark, I. Dybedal, D. Josefsen, P. Fenaux, P. Hokland, M. S. Holm, M. Cazzola, L. Malcovati, S. Tauro, D. Bowen, J. Boulwood, A. Pellagatti, J. E. Pimanda, A. Unnikrishnan, P. Vyas, G. Gohring, B. Schlegelberger, M. Tobiasson, G. Kvalheim, S. N. Constantinescu, C. Nerlov, L. Nilsson, P. J. Campbell, R. Sandberg, E. Papaemmanuil, E. Hellstrom-Lindberg, S. Linnarsson, S. E. Jacobsen, Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell* **25**, 794-808 (2014).
34. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
35. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
36. A. Giustacchini, S. Thongjuea, N. Barkas, P. S. Woll, B. J. Povinelli, C. A. G. Booth, P. Sopp, R. Norfo, A. Rodriguez-Meira, N. Ashley, L. Jamieson, P. Vyas, K. Anderson, A. Segerstolpe, H. Qian, U. Olsson-Stromberg, S. Mustjoki, R. Sandberg, S. E. W. Jacobsen, A. J. Mead, Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* **23**, 692-702 (2017).
37. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, 2nd, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P.

K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O.

Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016).

38. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

39. N. Novershtern, A. Subramanian, L. N. Lawton, R. H. Mak, W. N. Haining, M. E. McConkey, N. Habib, N. Yosef, C. Y. Chang, T. Shay, G. M. Frampton, A. C. Drake, I. Leskov, B. Nilsson, F. Preffer, D. Dombkowski, J. W. Evans, T. Liefeld, J. S. Smutko, J. Chen, N. Friedman, R. A. Young, T. R. Golub, A. Regev, B. L. Ebert, Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309 (2011).

Acknowledgements:

We thank Francoise Pflumio and Emmanuelle Six for providing OP9-hDL1 stromal cells, Bilyana Stoilova and Dimitris Karamitros for advice on lymphoid culture experiments, and Paresh Vyas and Irene Roberts and the Nerlov laboratory for helpful discussions. The assistance of the WIMM flow cytometry facility, WIMM single cell facility and WIMM sequencing facility is gratefully acknowledged. **Funding:** This work was supported by the MRC (G0701761, G0900892 and MC_UU_12009/7 to C.N.) and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The WIMM Single Cell Core Facility is supported by the MRC MHU (MC_UU_12009), the Oxford Single Cell Biology Consortium (MR/M00919X/1) and the WT-ISSF (097813/Z/11/B#) funding, and the WIMM Strategic Alliance awards G0902418 and MC_UU_12025]. The WIMM FACS Core Facility is supported by the MRC HIU, MRC MHU (MC_UU_12009), NIHR Oxford BRC and the John Fell Fund

(131/030 and 101/517), the EPA fund (CF182 and CF170) and by WIMM Strategic Alliance awards (G0902418 and MC_UU_12025). **Author contributions:** R.D. and C.N. designed the experiments; R.D. performed the experiments; S.T. performed bioinformatics analysis; K.T. provided bone marrow samples; R.D., S.T. and C.N. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** RNA sequencing data are available from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE113046.

Figure legends:

Fig. 1. Myeloid potential in predefined hematopoietic progenitor populations. (A, B) Gating strategy of populations in adult human bone marrow samples. Displayed are live, lineage negative singlets, positive or negative for CD34 and CD38 as indicated above the plots. CD10 was included in the lineage cocktail. CMP: common myeloid progenitor, GMP: granulocyte and monocyte progenitor, LMPP: lymphoid primed multi-potent progenitor. Data are representative of 7 donors in 16 experiments, and the numbers show gated cells as percentage \pm SD of the parental gate. **(C)** Histogram showing cell type determined by morphology of cytopins from single cell cultures of CMP (N=181), LMPP (N=75) and GMP (N=77) as percentage of total number of cultures analysed. Mo: monocyte, Ne: neutrophil, Eo: eosinophil, Ma/Ba: mast cell/basophil, Mixed: neutrophil/monocyte in combination with eosinophil/basophil/mast cell morphology. Data are summary of 6 independent experiments (Table S1A). **(D-G)** Morphology of single cell cultures from CMP **(D, E)**, GMP **(F)** or LMPP **(G)**, representative of the data in **(C)**. Scale bars: 25 μ m.

Fig. 2. Heterogeneity of human CMPs revealed by single cell RNA sequencing. (A) Plot indicating the CD45RA and CD123 expression measures by flow cytometry of the CMP, GMP and MEP cells used for single cell RNA sequencing. (B) Left: TSNE plot of 237 CMP, 32 GMP and 27 MEP cells showing 7 clusters as indicated by coloring. The contour plot (red gradient color in the background) indicates the kernel smoothing density of cells. Right: The same TSNE plot showing the individual cell types: CMP (grey), MEP (light blue) and GMP (dark blue). (C) *GATA1*, (D) *KLF1* and (E) *CEBPA* expression (as log₂(CPM)) superimposed on the TSNE plot from (B).

Fig. 3. Characterization of progenitor populations defined by CD114 and CD131. (A) Gene expression level of *CSF2RB* and *CSF3R* in single CMP cells in identified clusters. Y-axis represents the expression level in the log₂(CPM) scale. X-axis represents the cluster ID. Numbers below each plot show the number of expressing cells and the total number of cells within the respective cluster. Box plot shows median and quartile values and whiskers show outlier values within 1.5 times of the interquartile range. (B) Plot indicating the CD114 and CD131 flow cytometry signals of index sorted single CMP cells used for gene expression analysis. (C) Heatmap showing gene expression of index sorted CD131+ and CD114+ CMPs (indicated on top) as shown in (B). The genes were chosen as markers for the indicated clusters (left). Cells were clustered manually and the gene expression was used to annotate the cells to one of the clusters (bottom). (D) Flow cytometry plot showing CD114 and CD131 signals of CMP cells and gates used for sorting CMP subpopulations. Numbers are cells within each gate as average percentage±SD of parental gate. Data are representative of 11 independent experiments using 6 different donors.

Fig 4. Measurement of lineage output from CD114+ and CD131+ CMPs. (A) Gating strategy for flow cytometry analysis of bulk cultured progenitors. Examples of cultures of CD114+ CMPs (top) and CD131+ CMPs (bottom) are shown. Arrows in top panel indicate sequential gating. (B) Quantification of cell types identified as in (A) after culture of indicated progenitor populations, presented as frequency of total defined cells. Mo: monocyte, Ne: neutrophil, Ba: basophil, Eo: eosinophil, Ma: mast cell. Data is average from 5 (GMP) or 6 (CMP CD114+ and CMP CD131+) donors, and 2-10 cultures were grown for each population. (C) Number of live cells

retrieved from cultures in (B) measured by flow cytometry. P-values indicate significance of differences in cell numbers (Mann-Whitney U-test).

Fig. 5. Lineage affiliation of prospectively isolated myeloid progenitors. (A) CMP cells were index sorted and cultured under MME-conditions. The CD131 and CD114 expression of cells with at least one positive lineage readout indicated. The data are cumulative from 2 independent experiments (Table S1B). (B) Venn diagram of lineage output from cells in (A) with positive lineage readout defined by presence of both appropriate morphology and of lineage-specific gene expression. (C) CD114 and CD131 intensity of isolated CMPs that gave rise to cultures containing mast cells, basophils or eosinophils (orange), or monocytes or neutrophils (blue). (D) Venn diagram as in (B) showing the overlap between positive eosinophil, mast cell/basophil, megakaryocyte and erythrocyte lineage readouts. (E, F) cytopins of single CMP cultures showing (E) erythroid (Er) and megakaryocyte (Mk) morphology, and (F) erythroid, megakaryocyte and basophil (Ba) morphology. (G) Lymphoid potential of indicated progenitor cell populations (30 cells per culture) grown in B/NK-cell conditions (MS5 stromal cells) or T-cell conditions (OP9-hDL1 stromal cells) shown as frequency of cultures producing B, NK or T cells. Data are from 2 donors in 2 independent experiments. Numbers above bars indicate number of cultures with a positive readout/number of cultures.

Fig. 6. Gene expression in prospectively isolated progenitor populations. (A) The indicated cell populations were purified from 4 independent bone marrow donors and RNA sequenced. The box-and-whisker plots show the expression of *CSF2RB* and *CSF3R*. Boxes show mean and central quartiles; whiskers show data range. Individual data points are overlaid the plots. (B) The 3D-plot shows principal component analysis using the first three components derived from 1,868 genes (P -value<0.05 (ANOVA) and $CV \geq 0.3$); encircled areas indicate clusters containing myelo-erythroid progenitor cells with or without *GATA1* expression as indicated. (C-K) Gene expression levels measured by RNA sequencing of the indicated genes, shown as in (A).

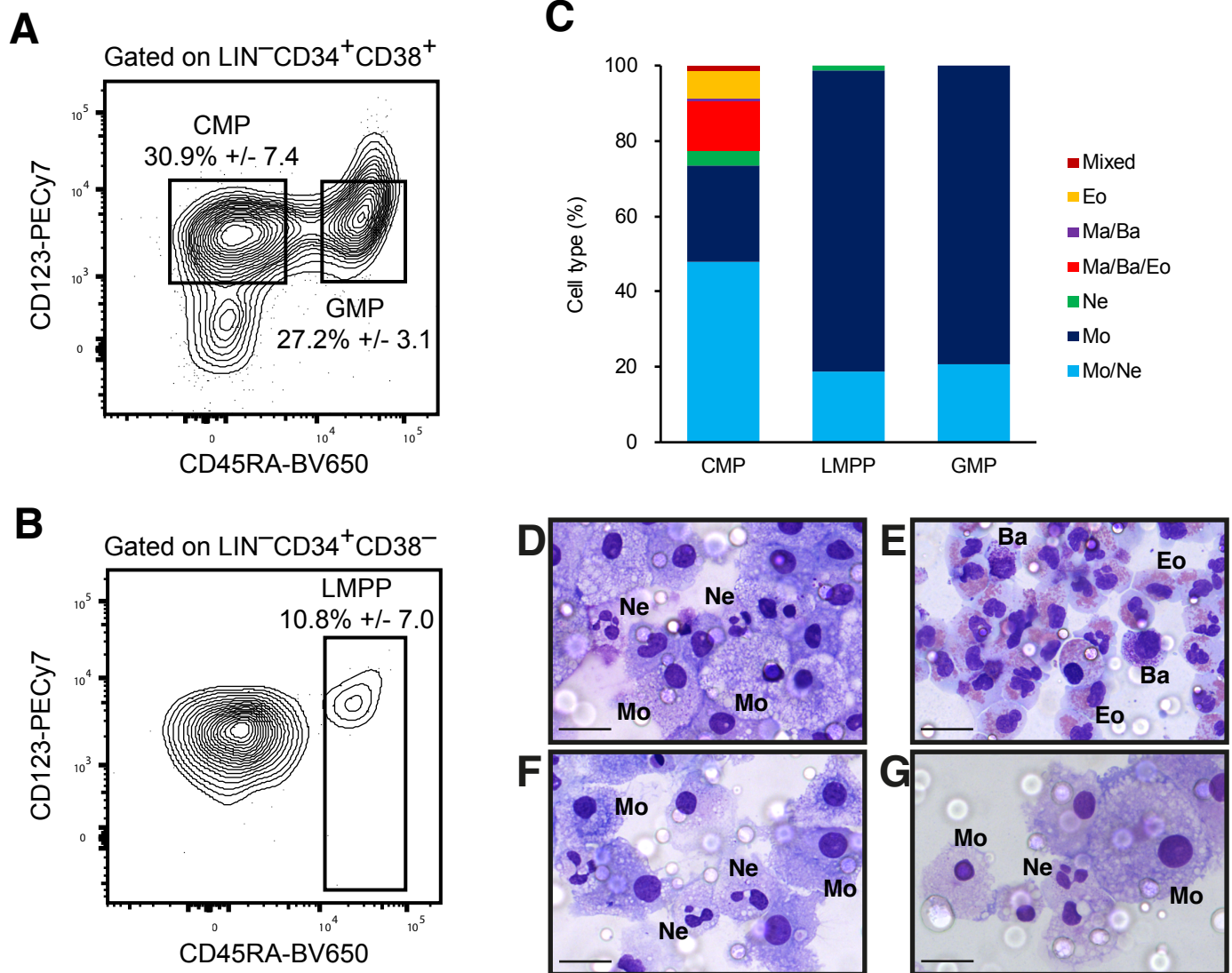


Figure 1

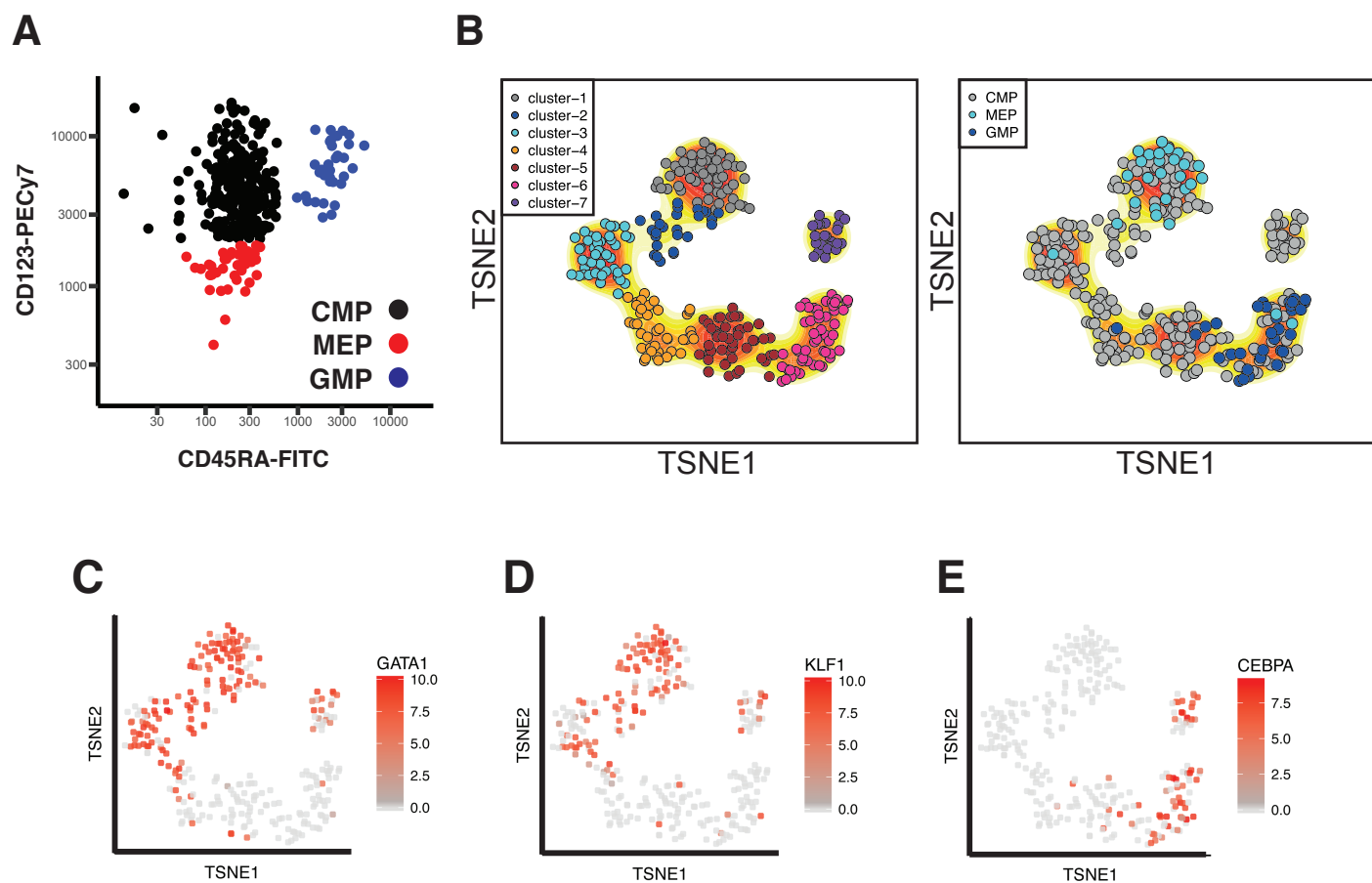


Figure 2

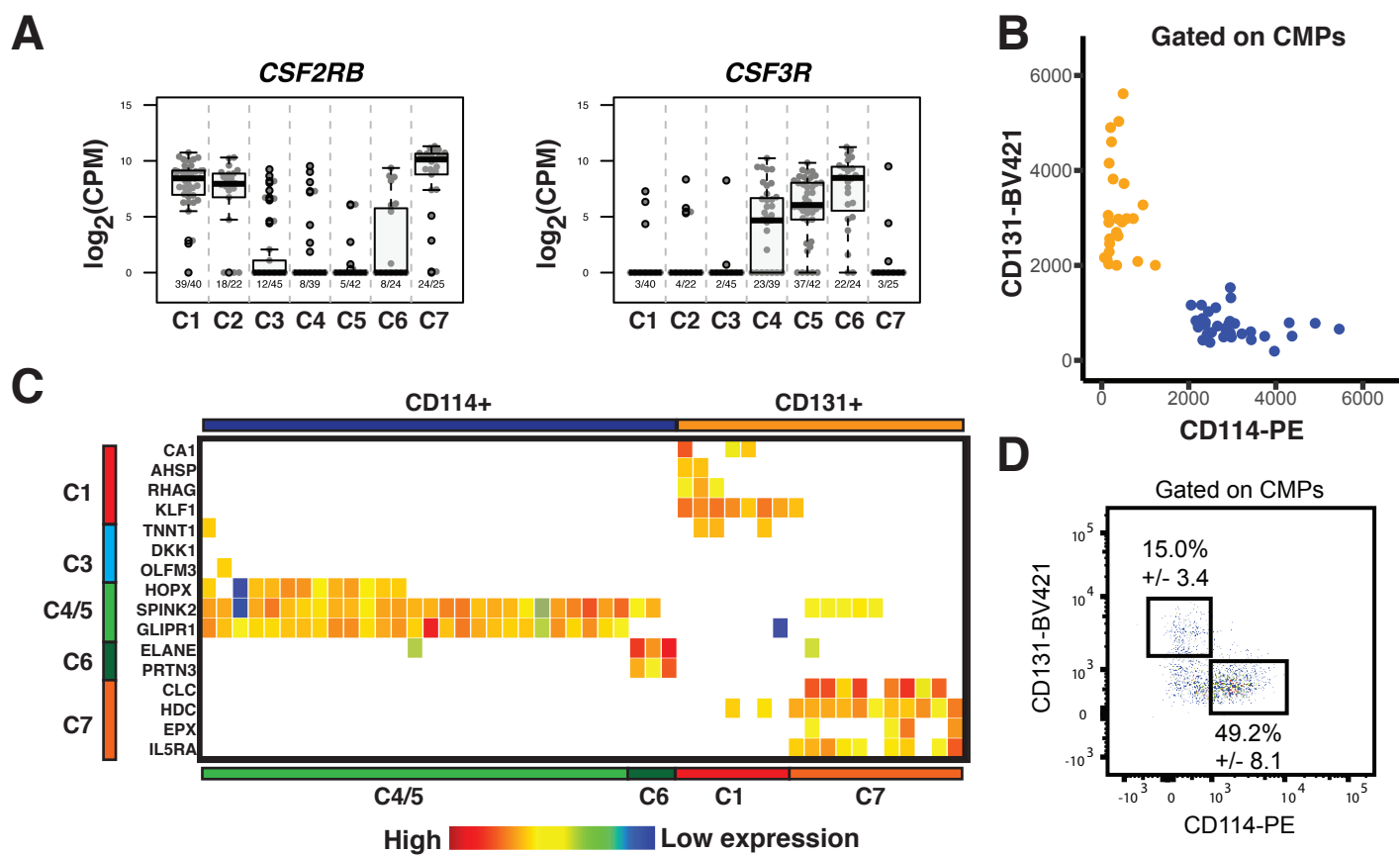


Figure 3

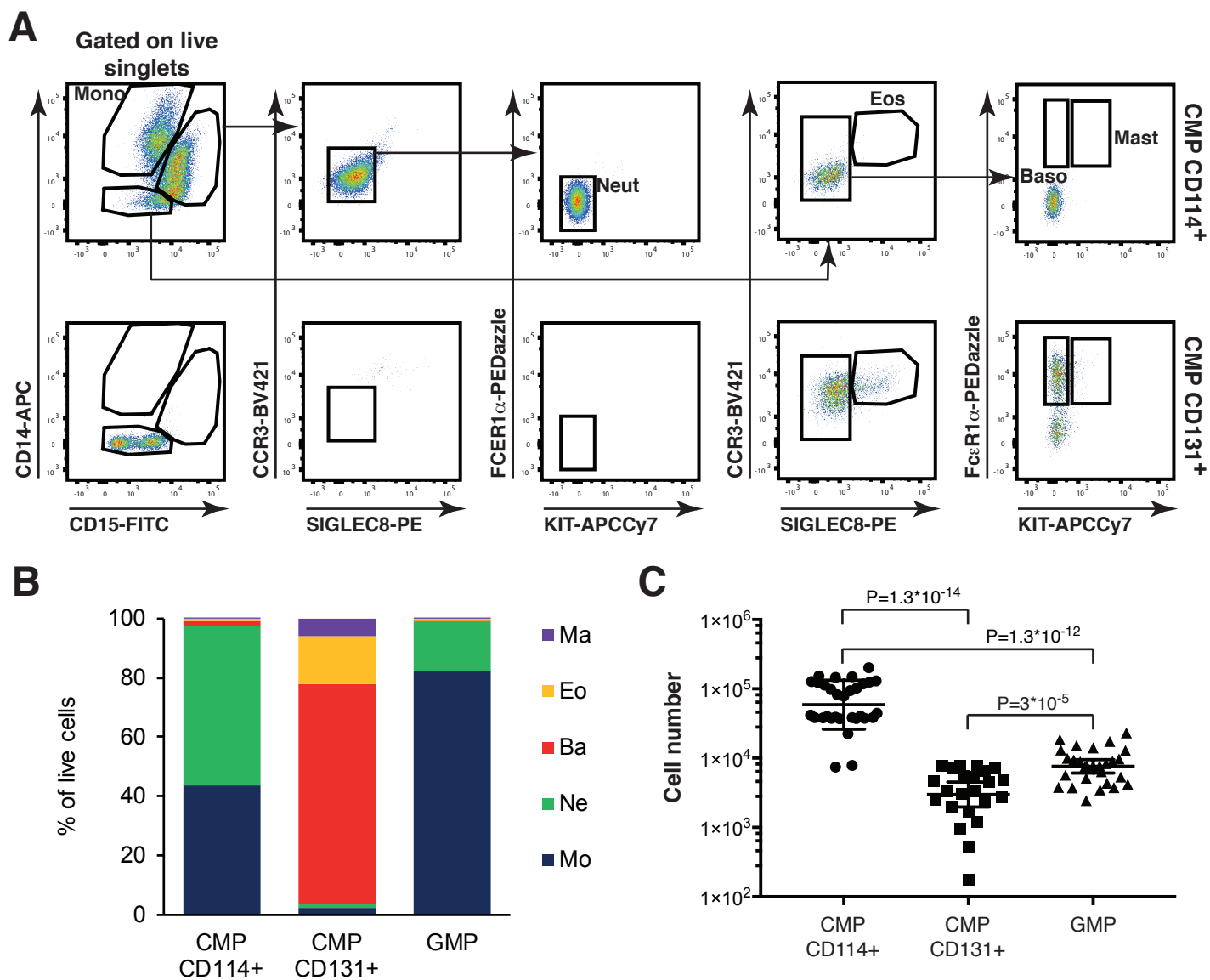


Figure 4

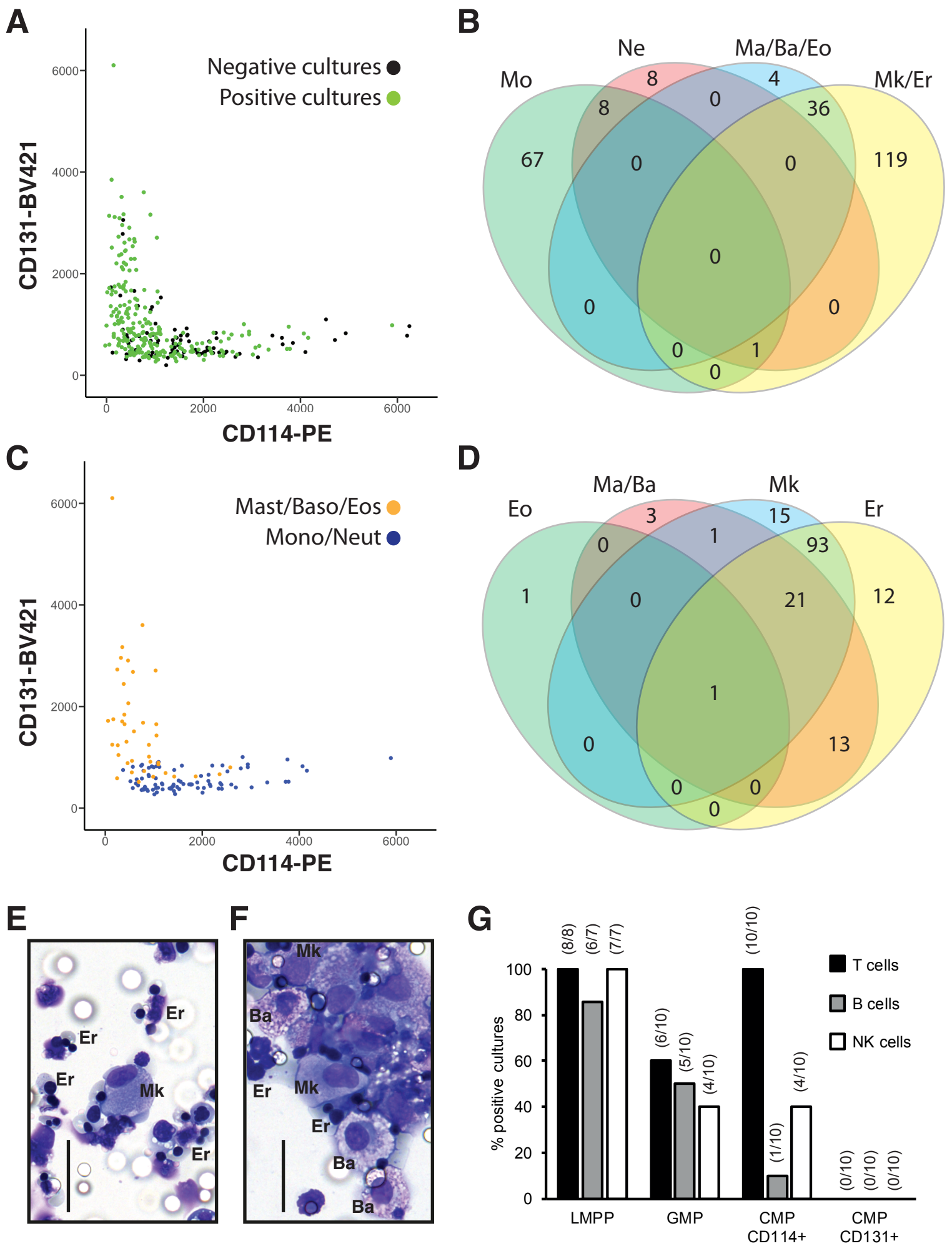


Figure 5

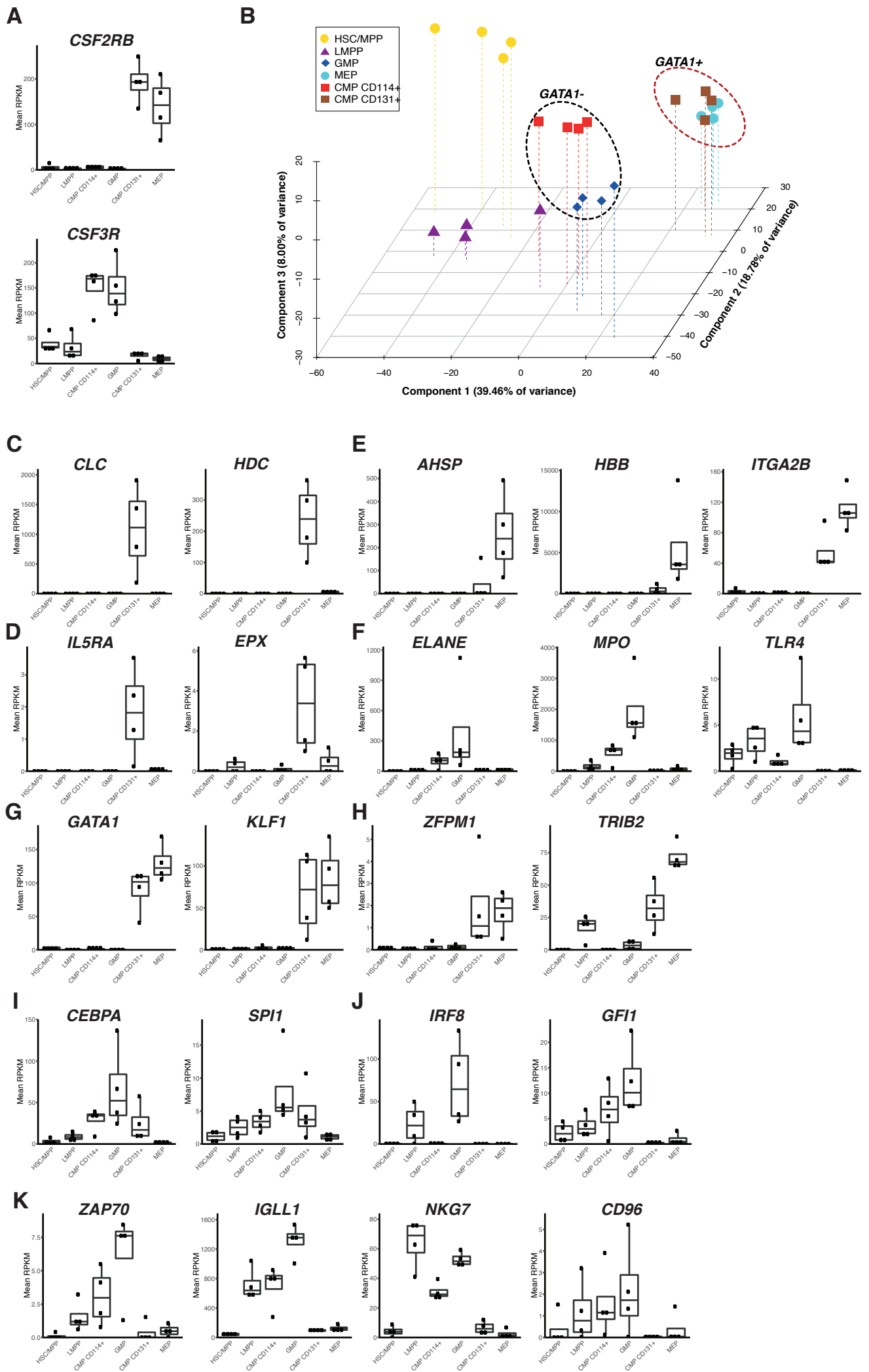
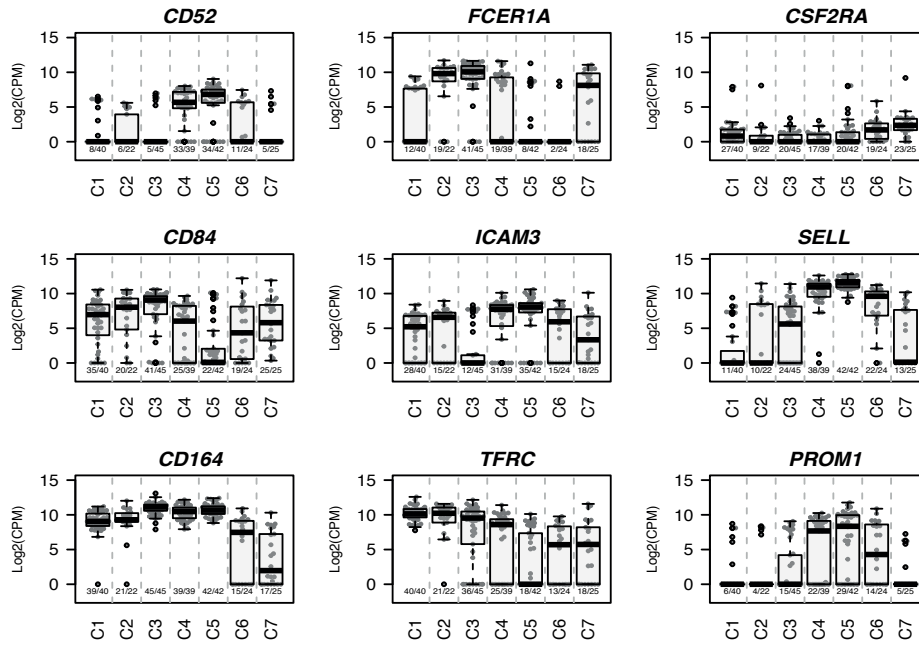
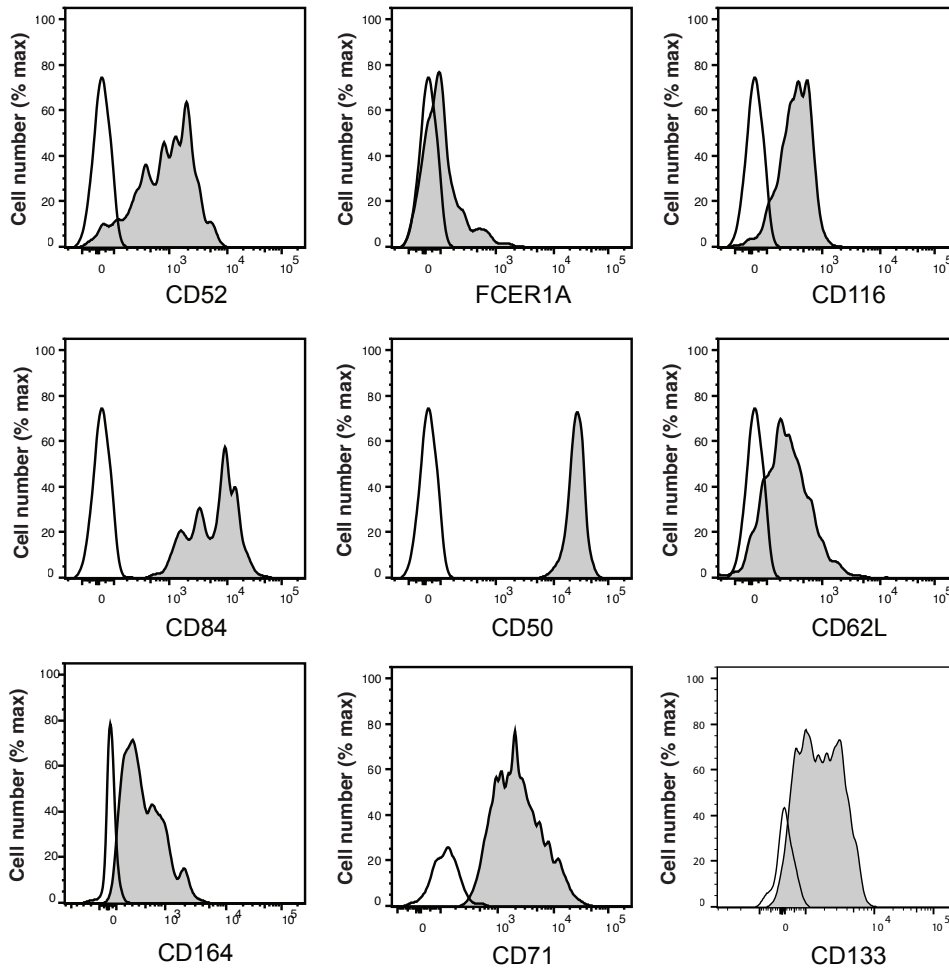
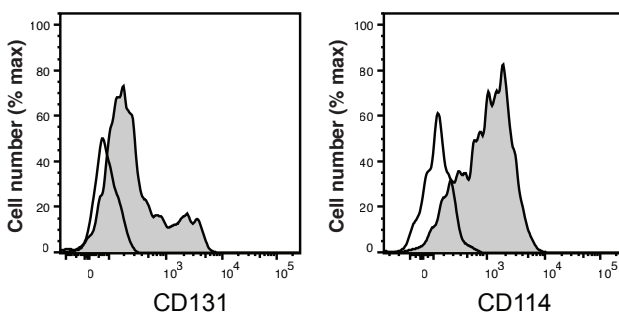
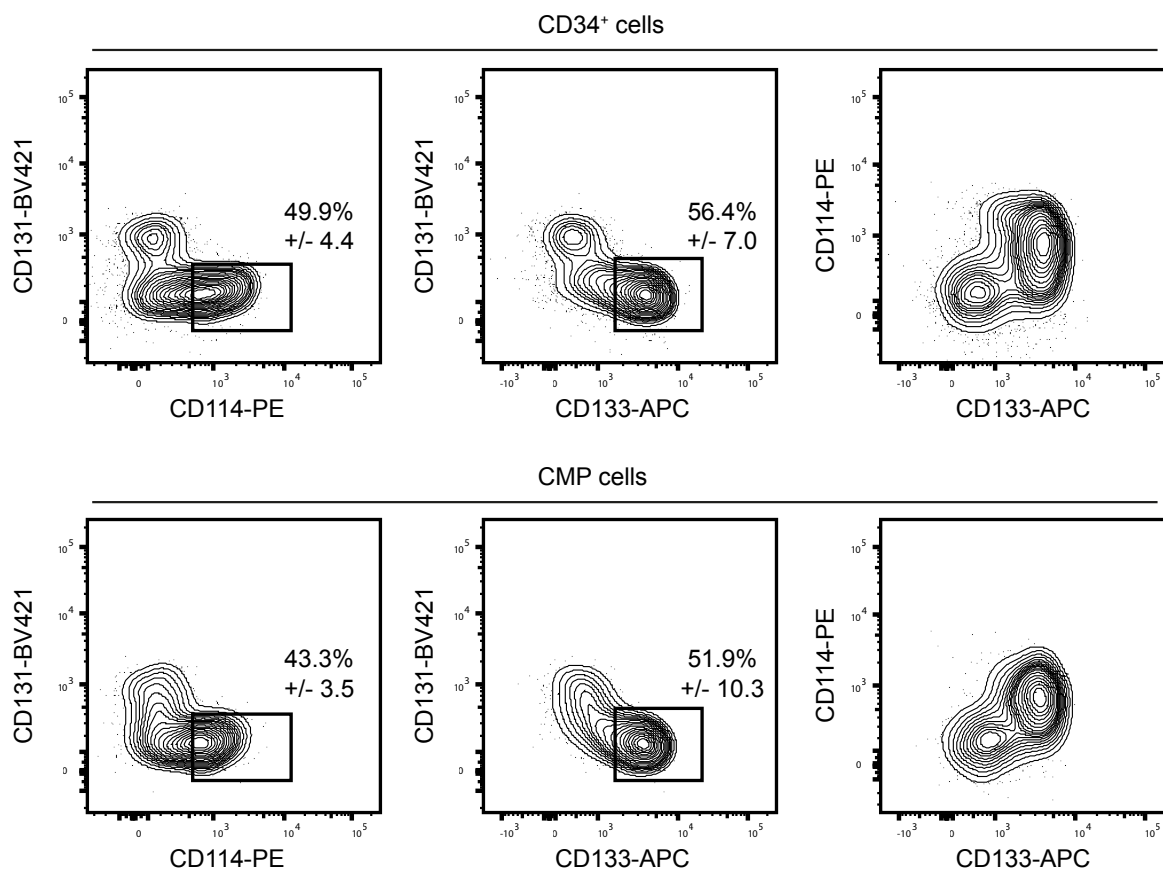
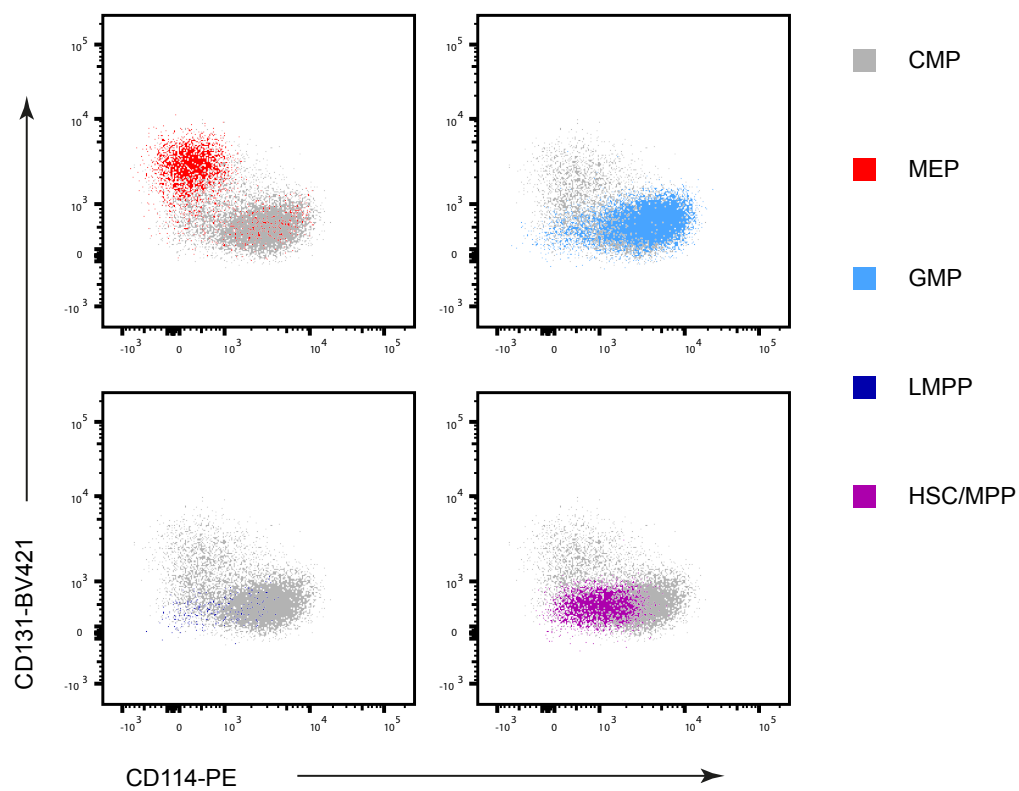


Figure 6

A**B****C****Figure S1**

A**B****Figure S2**

A

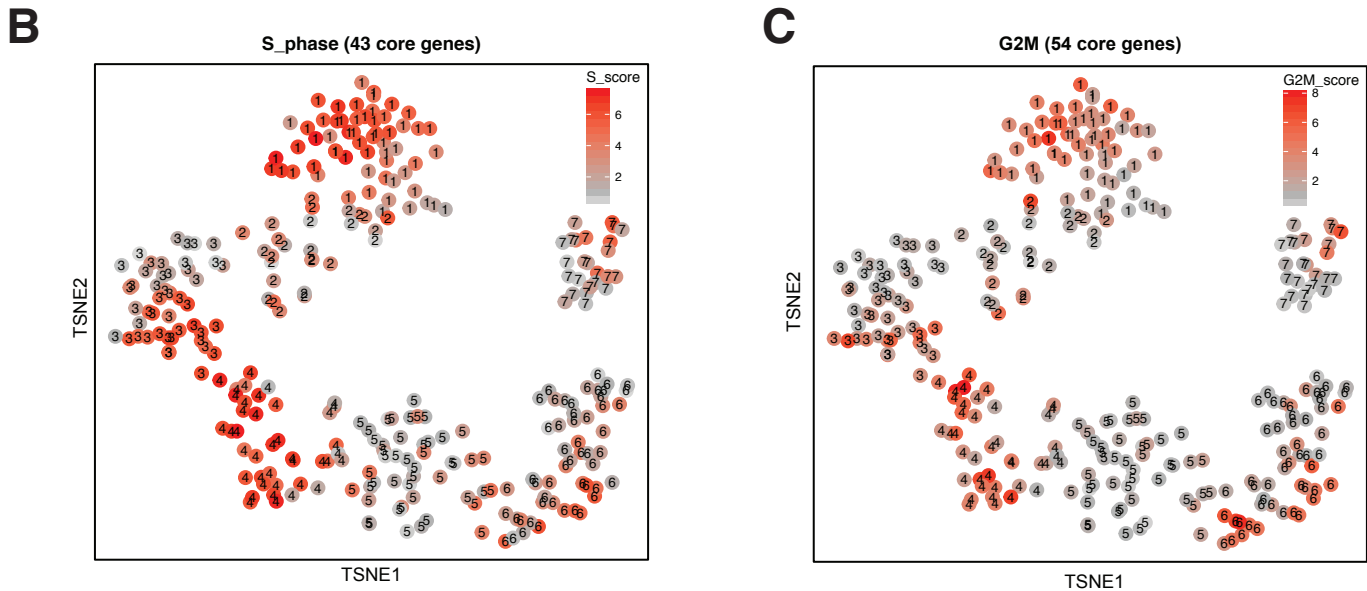
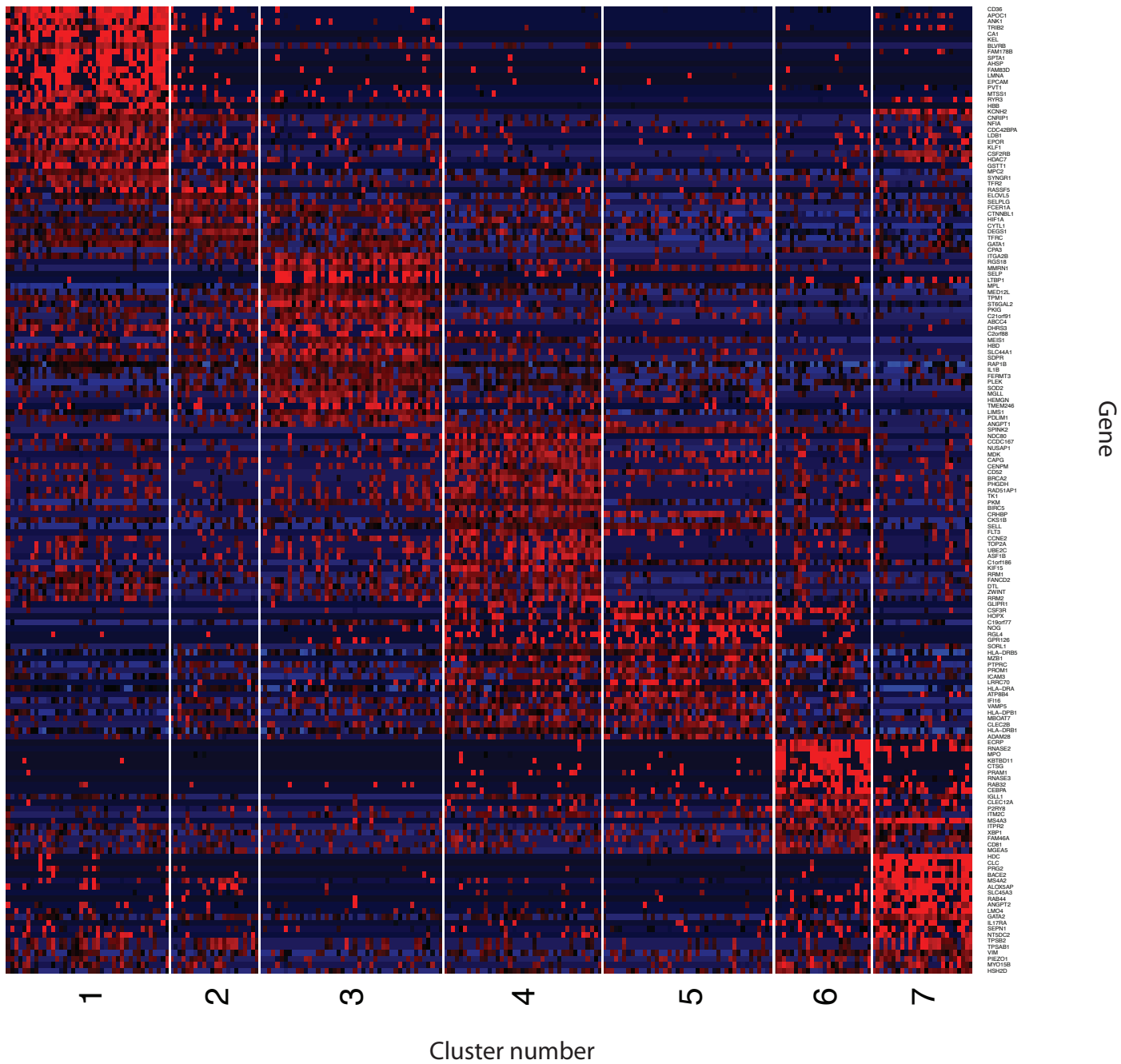


Figure S3

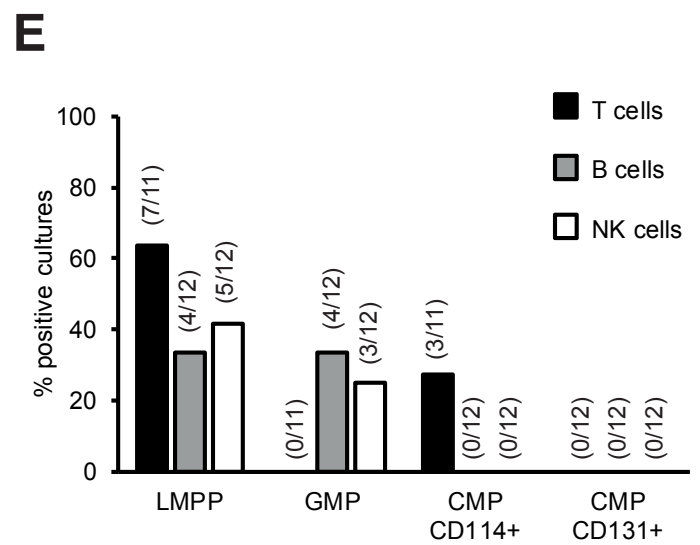
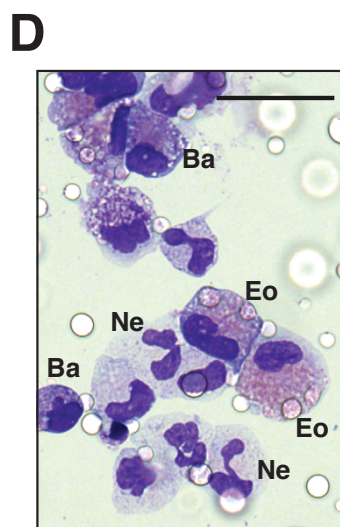
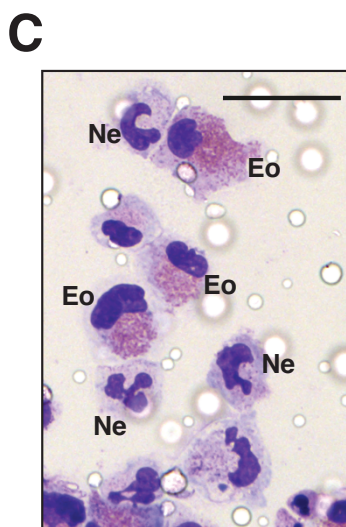
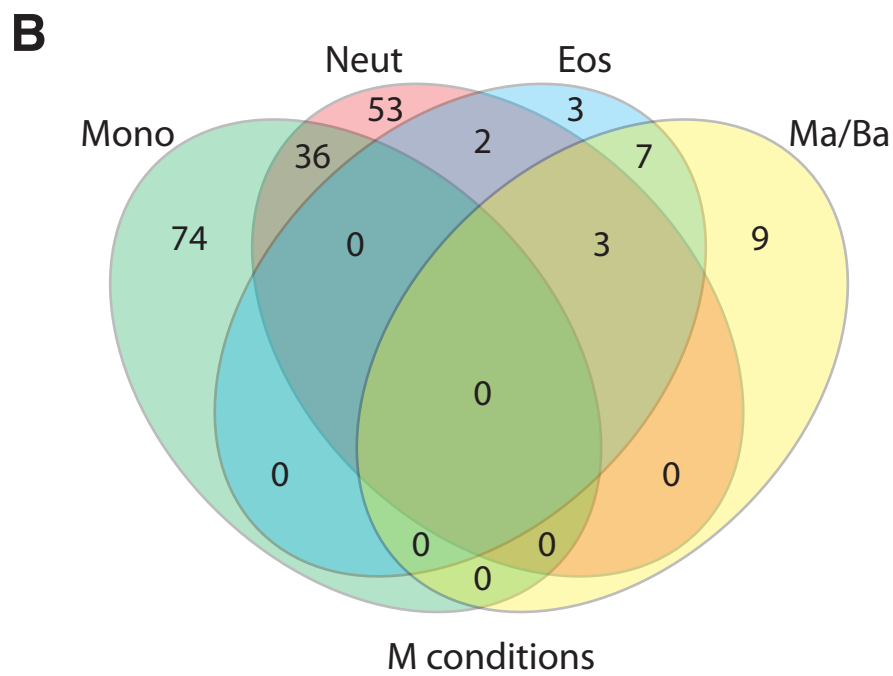
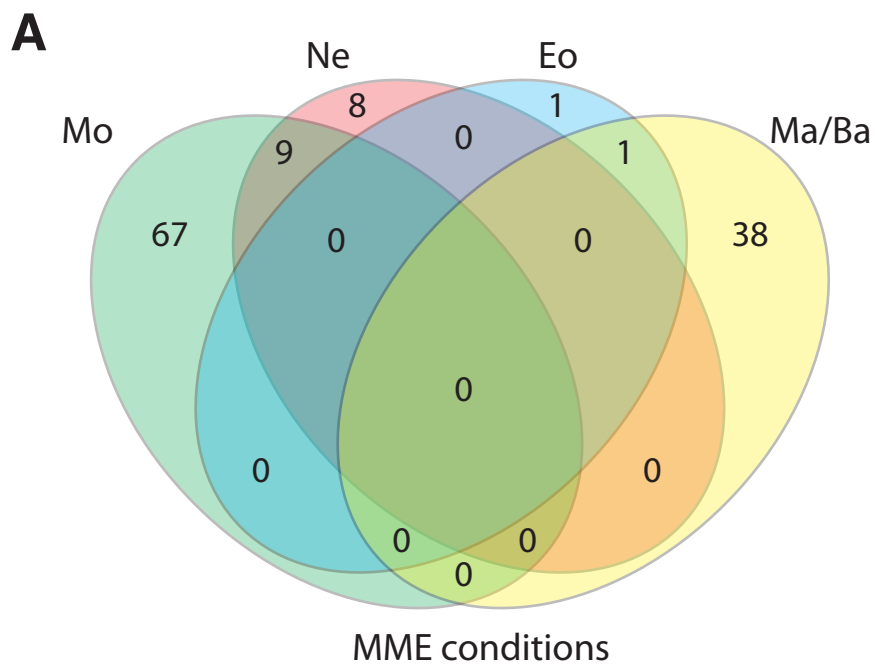


Figure S4

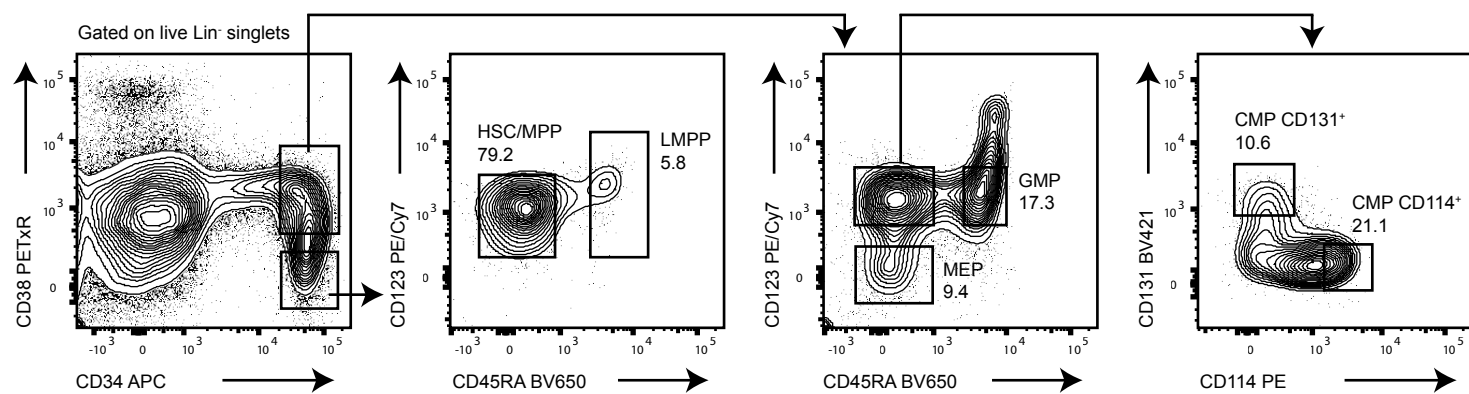
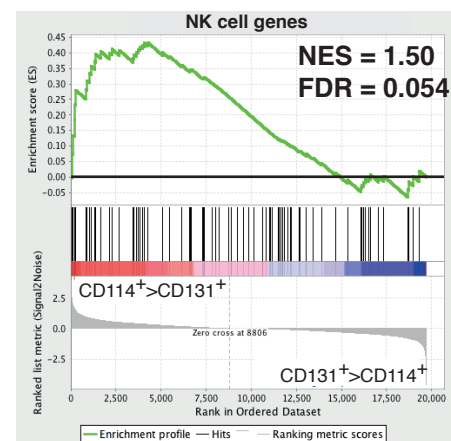
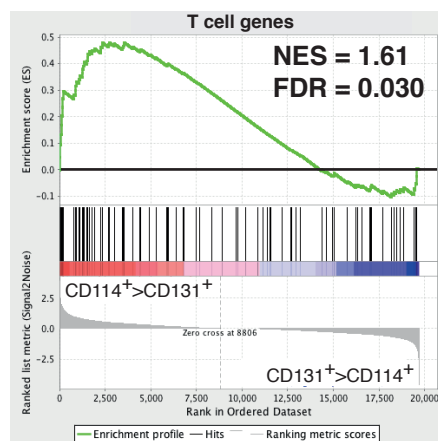
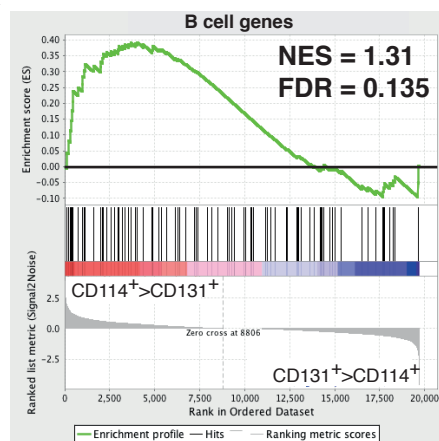
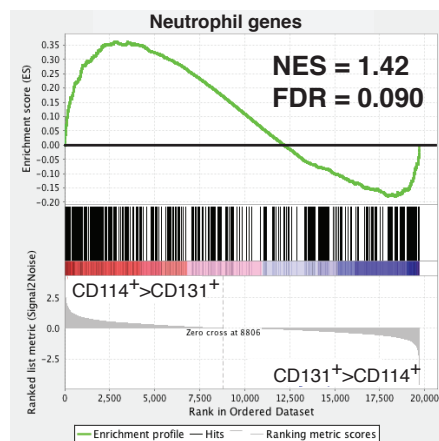


Figure S5

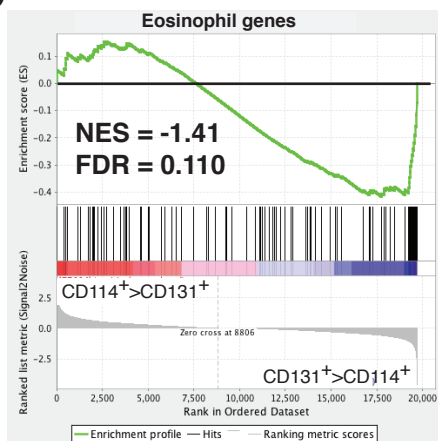
A



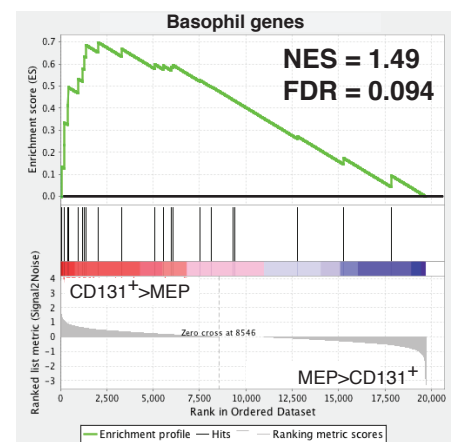
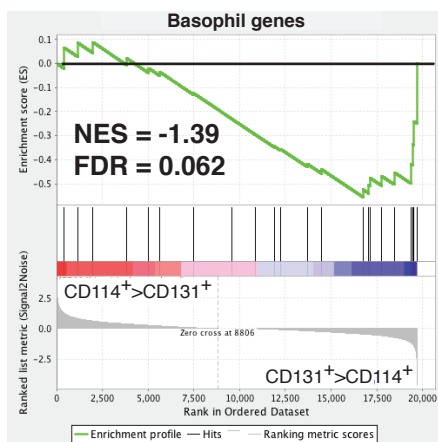
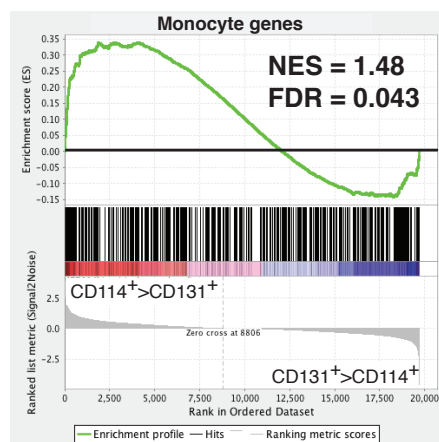
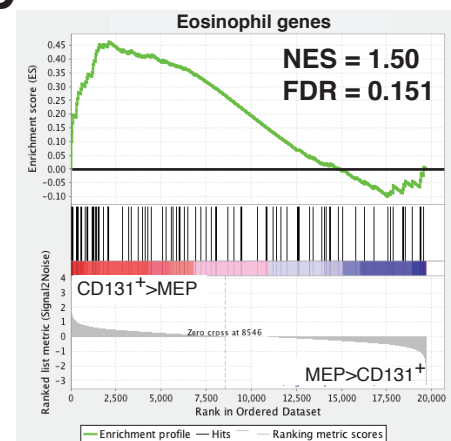
B



C



D



E

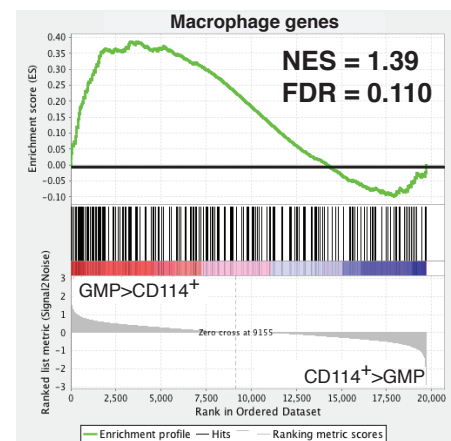
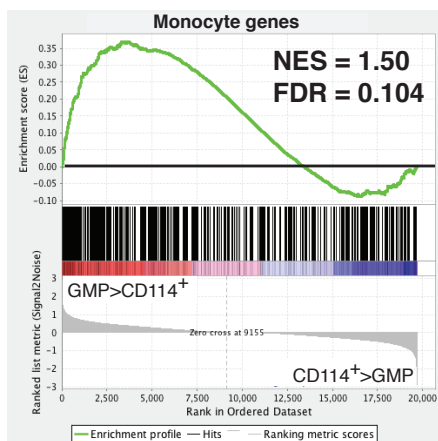
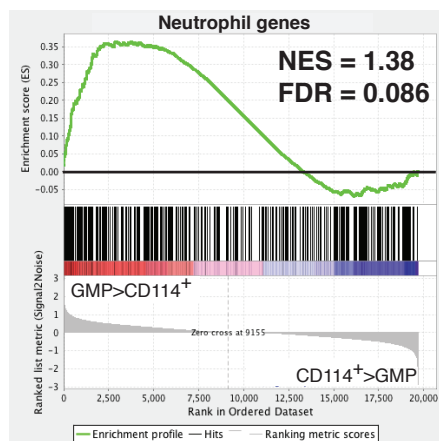


Figure S6

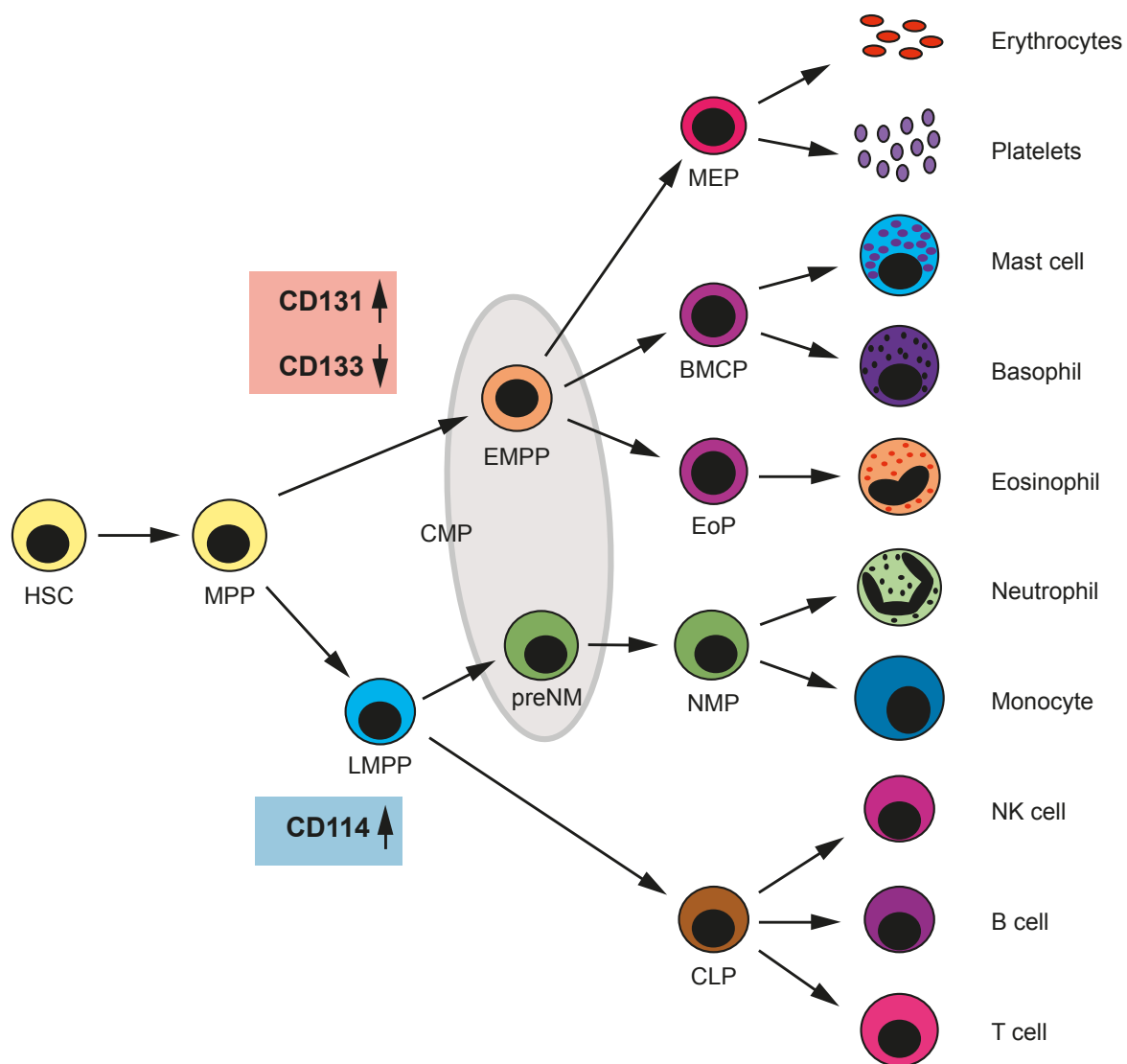


Figure S7

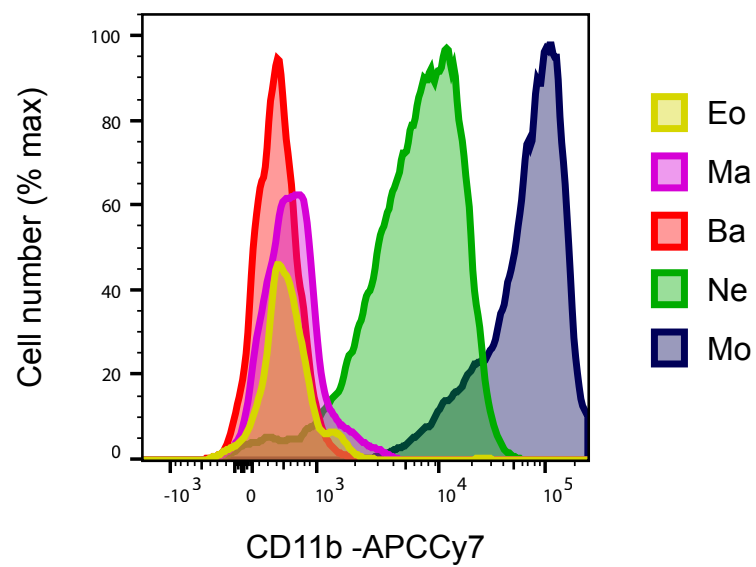


Figure S8