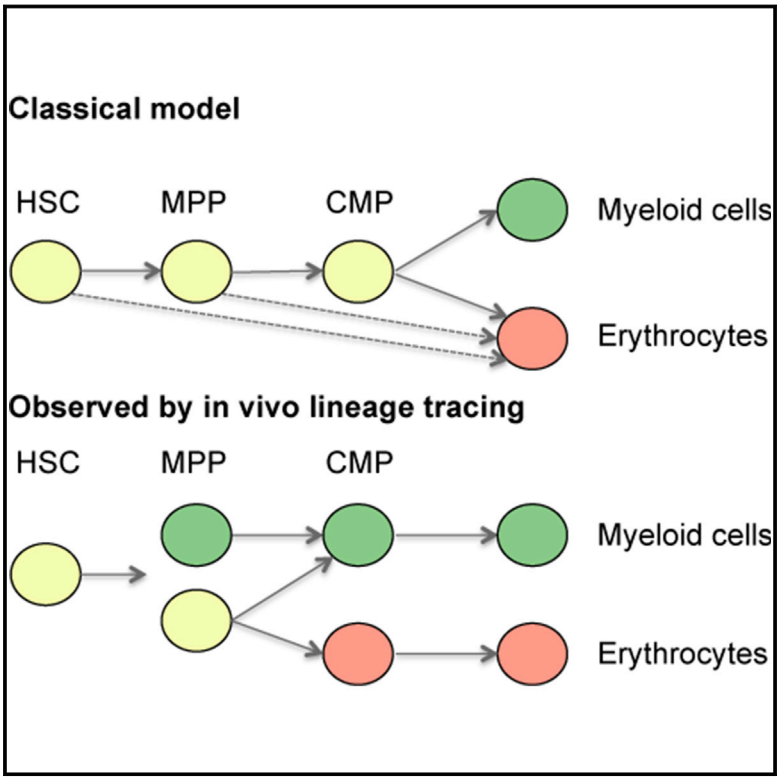
Article

Cell

The Branching Point in Erythro-Myeloid Differentiation

Graphical Abstract



Highlights

- Classically defined common myeloid progenitors (CMPs) are highly heterogeneous
- Most individual CMPs yield either erythrocytes or myeloid cells
- A true common progenitor exists within the multipotent progenitor (MPPs) pool
- The divergence between myeloid cells and erythrocytes starts within the MPPs

Authors

Leïla Perié, Ken R. Duffy, Lianne Kok, Rob J. de Boer, Ton N. Schumacher

Correspondence

leila.perie@curie.fr (L.P.), t.schumacher@nki.nl (T.N.S.)

In Brief

The so-called common myeloid progenitors (CMPs) are highly heterogeneous, with most individual CMPs yielding either only erythrocytes or only myeloid cells after transplantation. This provides in vivo evidence for a model of early parallel lineage branching within the multipotent progenitor pool that is in sharp contrast with the standard model of hematopoiesis.

Perié et al., 2015, Cell *163*, 1655–1662
December 17, 2015
http://dx.doi.org/10.1016/j.cell.2015.11.059

CellPress

Cell

The Branching Point in Erythro-Myeloid Differentiation

Leïla Perié,[1,2,4,5,*] Ken R. Duffy,[3] Lianne Kok,[1] Rob J. de Boer,[2] and Ton N. Schumacher[1,*]
[1]Division of Immunology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands
[2]Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands
[3]Hamilton Institute, Maynooth University, Maynooth, Co Kildare, Ireland
[4]Institut Curie, PSL Research University, CNRS UMR168, 26 rue d'Ulm, 75005 Paris, France
[5]Sorbonne Universités, UPMC Univ Paris 06, 4 place Jussieu, 75005 Paris, France
*Correspondence: leila.perie@curie.fr (L.P.), t.schumacher@nki.nl (T.N.S.)
http://dx.doi.org/10.1016/j.cell.2015.11.059

SUMMARY

Development of mature blood cell progenies from hematopoietic stem cells involves the transition through lineage-restricted progenitors. The first branching point along this developmental process is thought to separate the erythro-myeloid and lymphoid lineage fate by yielding two intermediate progenitors, the common myeloid and the common lymphoid progenitors (CMPs and CLPs). Here, we use single-cell lineage tracing to demonstrate that so-called CMPs are highly heterogeneous with respect to cellular output, with most individual CMPs yielding either only erythrocytes or only myeloid cells after transplantation. Furthermore, based on the labeling of earlier progenitors, we show that the divergence between the myeloid and erythroid lineage develops within multipotent progenitors (MPP). These data provide evidence for a model of hematopoietic branching in which multiple distinct lineage commitments occur in parallel within the MPP pool.

INTRODUCTION

During hematopoiesis, hematopoietic stem cells (HSCs) self-renew or differentiate into all blood cell types through successive stages of lineage commitment, a process that has become a prototype of multi-lineage diversification from a stem cell pool. The prevailing model of hematopoiesis predicts a step-by-step process of lineage commitment in which HSCs give rise to multipotent progenitors (MPPs) that subsequently generate intermediate lineage restricted progenitors. The strongest support for this model has come from the identification of the common lymphoid progenitor (CLP) (Kondo et al., 1997) that produces lymphoid cells (i.e., T and B lymphocytes and NK cells) and the common myeloid progenitor (CMP) (Akashi et al., 2000) that gives rise to granulocyte-macrophage (GM) progenitors and to the megakaryocyte-erythroid (MkE) progenitors. In this model, the lymphoid-myeloid split forms the first step in lineage commitment downstream of MPPs (Reya et al., 2001). However, a number of studies have questioned the CMP-CLP split as the first step of commitment. Reports of unequal output of GM or MkE progenitors by subsets of CMPs (Iwasaki et al., 2005; Nutt et al., 2005; Terszowski et al., 2005) suggest a separation between the erythroid and myeloid lineages earlier than the CMP stage. Additionally, earlier bypasses toward lymphoid, myeloid (GM), or MkE commitment have been reported in HSCs (Benz et al., 2012; Dykstra et al., 2007; Sanjuan-Pla et al., 2013; Yamamoto et al., 2013) and MPPs (Adolfsson et al., 2005; Cabezas-Wallscheid et al., 2014; Igarashi et al., 2002; Lai and Kondo, 2006; Månsson et al., 2007; Miyawaki et al., 2015; Naik et al., 2013; Pietras et al., 2015; Takano et al., 2004; Yoshida et al., 2006). Based on some of these data, a myelo-lymphoid versus erythro-megakaryocytic separation has been proposed as the first commitment step (Adolfsson et al., 2005), but other results have also contradicted this model (Boyer et al., 2011; Forsberg et al., 2006). Overall, which lineage commitment occurs first between the lymphoid, myeloid, or erythroid lineages remains to be determined. Importantly, most of the results that bear on this question have been obtained using in vitro clonal assays or population-based tracking approaches that can miss in vivo cellular heterogeneity and, thereby, influence our interpretation of lineage commitment. To address this issue, here we have utilized cellular barcoding technology that allows the in vivo tracking of single cell fate (Gerlach et al., 2013; Gerrits et al., 2010; Lu et al., 2011; Naik et al., 2013, 2014; Schepers et al., 2008) to describe the steps of myeloid and erythroid commitment during successive stages of hematopoietic development.

RESULTS

To understand the clonal output of individual CMPs, we obtained mouse CMPs based on their original description ($ckit^+Sca1^-CD16/32^{low}CD34^+$) (Akashi et al., 2000), labeled these with unique genetic barcodes in a 6 hr in vitro process, and assessed their ability for myelopoiesis (for the GM lineage) and erythropoiesis (for the MkE lineage). At different time points after transplantation of 2×10^3 barcode-labeled CMPs into irradiated recipients, CMP-derived myeloid cells (GFP^+CD11b^+) and erythroblasts ($GFP^+Ter119^+$) were sorted and assessed for their barcode identity using PCR and deep-sequencing (Figures 1A, 1C, and 1D; Table S1). Erythroblasts were measured as a proxy for erythropoiesis, as they retain the nucleus required for barcode

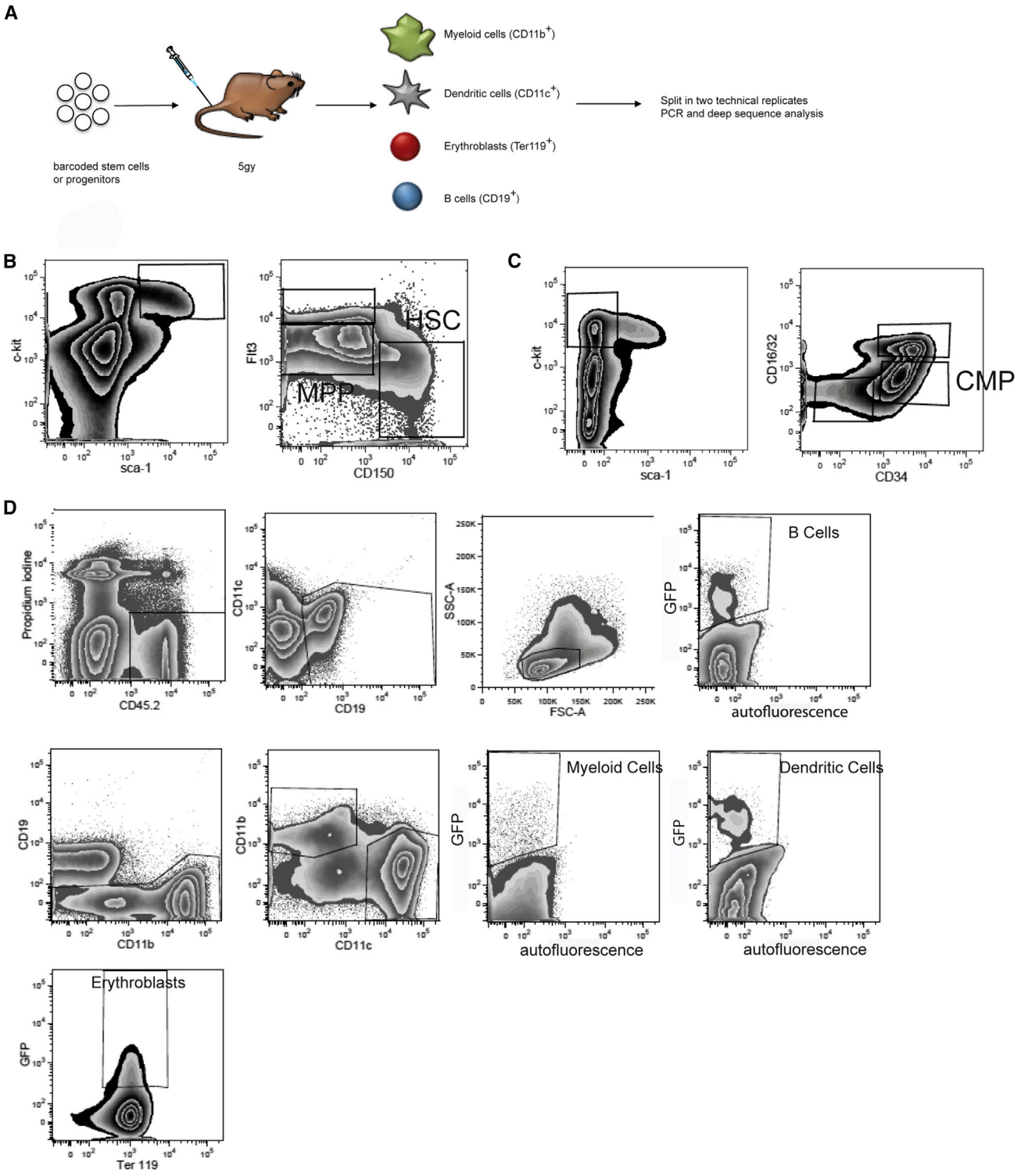


Figure 1. Summary of Experimental Setup and Sort Profiles of Progenitors and Output Cell Types, Related to Table S1

(A) Progenitors (HSC, MPP, or CMP) were transduced with low efficiency (5%–15%) to minimize the chance of two different barcode virus particles integrating into the same progenitor. Progenitors from a single transduction batch were then transplanted into 2–8 5gy irradiated mice (to allow later comparison of barcodes between mice as a control for random barcode sharing). At the indicated time points after transplantation, myeloid cells ($CD11b^+$), erythroblasts ($Ter119^+$), DCs ($CD11c^+$) and B cells ($CD19^+$) were isolated from the spleen of recipient mice. Each sample was split into two replicates. Barcodes were amplified by PCR and analyzed by deep sequencing.

(legend continued on next page)

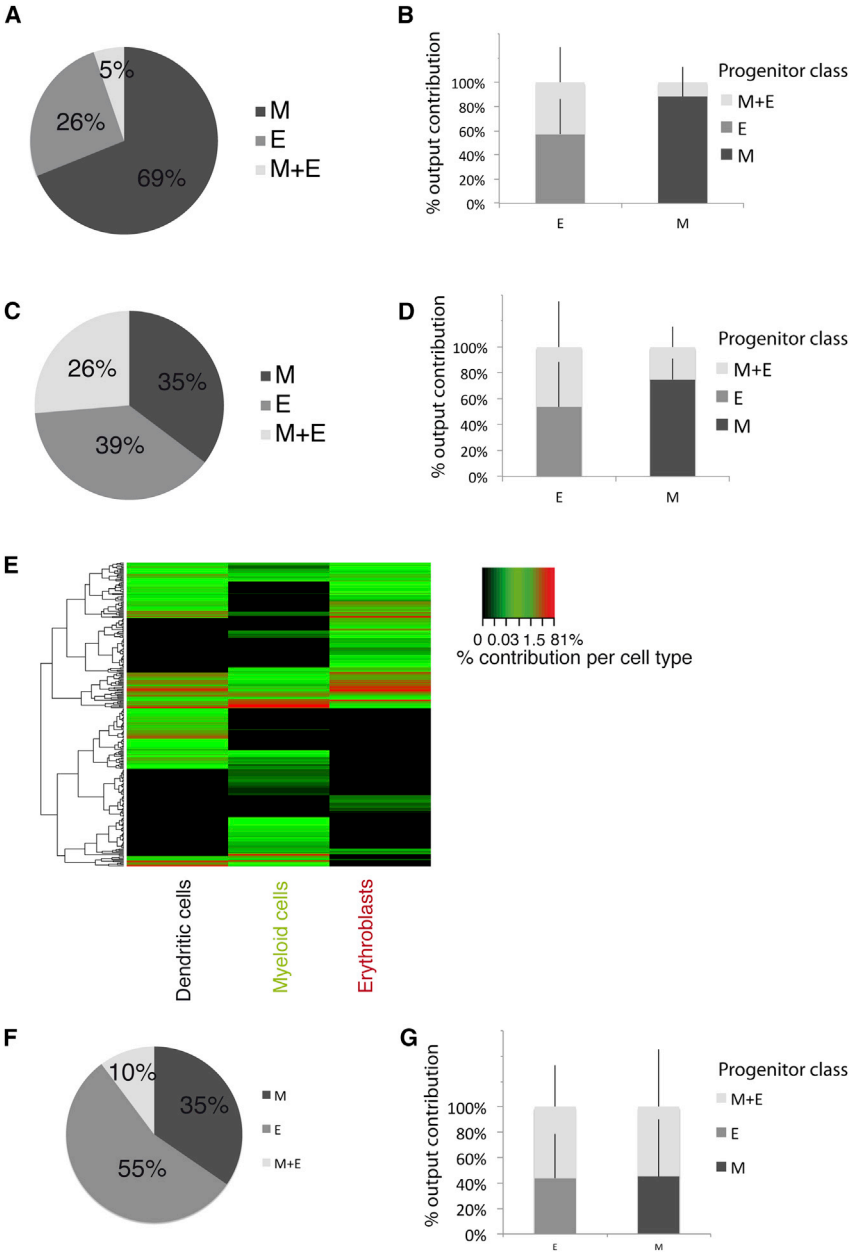


Figure 2. Individual CMPs Yield Either Erythrocytes or Myeloid Cells, Related to Figures S1, S2, and S3

(A–G) Barcode-labeled CMPs were transplanted into 5gy-irradiated mice. At day 6 (A and B) or day 14 (C–E) post-transplantation, myeloid cells (CD11b$^+$), erythroblasts (Ter119$^+$), and DCs (CD11c$^+$) were recovered from spleen and barcode content was analyzed. (A and C) Proportion of CMPs classified as producing either M (myeloid cells), E (erythroblasts), or both, using a 1% classification threshold (other thresholds in Figure S1). (B and D) Quantitative contribution of the three classes to each lineage (mean + SD). (E) Heatmap representation of the output of individual CMPs (rows) to different cell types (columns) at day 14 (248 barcodes, pooled data of four mice), and arcsine transformed data clustered by complete linkage using Euclidean distances. For day 6, 636 single CMPs were analyzed (58 ± 46 CMPs/mouse), data shown in (A) and (B) displays average + SD over three experiments, four mice per experiment; for day 14, 248 single CMPs were analyzed (62 ± 18 CMPs/mouse), data shown in (C) and (D) displays average + SD over two experiments, two mice per experiment. Quality controls in Figure S1. (G) Proportion of CMPs classified as producing output as defined in (A) in bone marrow at day 14 post-transplantation.

(H) Quantitative contribution of the three classes to each lineage (mean + SD). For (F) and (G), 89 barcodes were analyzed (29 ± 6 CMPs/mouse) over two experiments, two mice per experiment), other thresholds than 1% displayed in Figures S2A and S2B.

analysis. Subsequently, barcode-labeled CMPs were assigned to a class of (lineage-biased) progenitor using a hand-tailored classifier that is based on cellular output of each individual progenitor toward the examined lineages (Naik et al., 2013). Six days after transplantation, we found that 95% of the engrafted CMPs showed highly biased (>99% of cellular output) production of myeloid cells (69% ± 17% of presumed CMPs) or highly biased (>99% of cellular output) production of erythroblasts (26% ± 13% of presumed CMPs) (Figure 2A; quality controls shown in Figures S1A and S1B; effect of classification threshold in Figures S1D and S1F). Furthermore, this large pool of myeloid-biased or erythroid-biased CMPs accounted for 82% ± 13% and 53% ± 27% of total myeloid cell and erythroblast production, respectively (Figures 2B and S1D). Only 5% ± 4% of CMPs produced appreciable numbers of both myeloid cells and erythroblasts, overall accounting for 11% ± 12% of total myeloid cell reads and 40% ± 27% of total erythroblast reads (Figures 2B, S1D, and S1F). The observation of this high frequency of uni-outcome CMPs could not be explained by limits of detection or sampling issues, as determined by the following detection controls: (1) the per progenitor output (as measured in read counts) of uni-outcome CMPs toward one specific lineage (i.e., myeloid cells or erythroblasts) was as high

(B) Bone marrow progenitors were enriched using anti-CD117 (c-kit) magnetic beads and subsequently stained with antibodies against CD16/32, CD117, Sca-1, CD135 (Flt3), CD34, and CD150. MPP and HSC were sorted using the gates shown.

(C) Gating strategy used to isolate CMP.

(D) Splenic cells were separated into a Ter119$^+$ and Ter119$^-$ fraction by magnetic-bead-based selection. The Ter119$^-$ fraction was used to sort B cells, myeloid cells, and DCs as shown. The Ter119$^+$ fraction was used to sort erythroblasts as shown.

as the per progenitor output of bi-outcome CMPs toward the same lineage (Figure S1H), and (2) the absence of output of uni-outcome CMPs toward the other lineage was reproducible in technical replicates before applying the replicate filter (data not shown). In other words, lack of detectable output of a given CMP toward either the myeloid cell or erythroblast compartment was a reproducible feature.

Similarly, when assessed 14 days after transplantation, 74% of the engrafted CMP produced essentially only myeloid cells (35% ± 3% of presumed CMPs) or only erythroblasts (39% ± 12% of presumed CMPs) (Figures 2C and S1E), accounting for 74% ± 16% and 53% ± 35% of total myeloid cell and erythroblast production, respectively (Figure 2D; effect of the thresholds shown in Figure S1G; detection control shown in Figure S1I). Also at this point in time, only a minority of CMPs showed substantial production of both myeloid cells and erythroblasts. Finally, heterogeneity in output toward either the myeloid and erythrocyte lineage is not specific to the spleen but was also observed in the bone marrow, where 90% of the engrafted CMP also produced essentially only myeloid cells (35% ± 9% of presumed CMPs) or essentially only erythroblasts (55% ± 12% of presumed CMPs) (Figures 2F, 2G, S2A, and S2B). The observation of the somewhat higher fraction of CMPs with detectable output in both the myeloid and erythrocyte lineage over time (day 6: 5% ± 4% of presumed CMPs; day 14: 26% ± 15% of presumed CMPs) may either be interpreted as a more prolonged engraftment potential of the small subset of CMPs producing both erythroblasts and myeloid cells or as contamination of CMPs by $Lin^-Sca1^+ckit^+$ (LSK) cells (Figure S1C), whose contribution may also be expected to increase over time after transplantation. A preponderance of uni-output was also observed at two CMP doses tested (Figures S2C and S2D).

In addition to heterogeneity in output toward the myeloid and erythrocyte lineage, further disparity in cellular output between individual CMPs was observed at day 14 after transplantation (Figures 2E and S3A), with 16% ± 4% of CMPs producing detectable output only in the dendritic cell (DC) lineage (Figure S3B, note that DC production could not be assessed at day 6). These DC uni-outcome CMPs accounted for 24% ± 25% of total DC production (Figure S3B), and their presence could not be explained by limits of detection (Figure S3C). While these data do not exclude the possibility that some DC uni-outcome CMPs might have produced myeloid cells or erythroblasts at an earlier point in time, this would then imply that individual CMPs differ in the kinetics with which they contribute to DC versus myeloid/erythroid lineage. Collectively, these data show a striking heterogeneity within the CMP pool, greater than previously suggested by in vitro or in vivo experiments that examined output at the population level (Iwasaki et al., 2005; Nakorn et al., 2003; Nutt et al., 2005; Terszowski et al., 2005). Importantly, this heterogeneity could not be explained by detection issues, and it was relatively insensitive to the threshold used for output classification (Figures S1D–S1I, S2A, S2B, S3C, and S3D).

To determine at which developmental point the observed divergence between the myeloid and erythroid lineage occurs, we subsequently generated barcode-labeled HSCs (LSK^+CD150^+) and MPPs ($LSK^+CD150^-Flt3^+$, excluding lymphoid-primed multipotent progenitors) (Figure 1B) and, after transplantation, analyzed their production of myeloid cells and erythroblasts together with their lymphoid cell production (measured by analysis of $CD19^+$ B cells) and $CD11c^+$ DC cell production (Figure 1D). As previously described using cellular barcoding (Gerrits et al., 2010; Naik et al., 2013), a small number of HSCs were the major contributors to total cellular output at day 27 after transplantation, and these HSCs were multi-outcome (Figures 3A, 3B, and S4A). With respect to contribution toward the erythroid and myeloid lineage, >99% of the total cell output of myeloid cells and erythroblasts was derived from HSCs (45% ± 7%) that produced both cell types at this time point (Figures 3C and 3D; quality controls and effect of classification threshold shown in Figures S5A–S5C). Some HSCs produced only myeloid cells (10% ± 4%) or erythroblasts (12% ± 7%), without detectable production of any of the other cell types (Figure 3B). However, these HSCs collectively contributed to <0.1% of the total cell output of erythroblasts and myeloid cells (Figures 3D and S5C). Furthermore, the occurrence of these biased HSCs may well be explained by insufficient sensitivity of detection, as these HSCs have a per progenitor output (as measured in read counts) toward the erythroid or the myeloid lineage that is several-fold lower than that of multi-outcome HSCs for the same lineage (Figure S5D). Also, when assessed at day 42 after transplantation (Figure S4B), the vast majority of myeloid cells and erythroblasts (≥90%) were derived from HSCs that contributed to both lineages. In conclusion, the large majority of erythroid and myeloid cells are produced by multi-outcome HSCs.

In contrast to this, substantial variability in the ability to contribute to erythroid and myeloid cell production was apparent among MPPs (Figures 4A and 4B). At day 14 after transplantation, only 20% ± 9% of the MPPs produced both erythroblasts and myeloid cells, irrespective of their capacity to produce DCs or B cells (Figures 4C and S6; quality controls and effect of classification threshold shown in Figures S7A and S7B). While these cells still contributed to a large fraction of erythroblast output (84% ± 23%), ~50% of total myeloid output was derived from the large pool of MPPs (51% ± 9%) that produced myeloid cells without producing erythroblasts (Figures 4D and S7C). Furthermore, contrary to the uni-outcome biased HSCs, the observation of these myeloid-not-erythroid-biased MPPs could not be explained by limits of detection, as: (1) these cells consistently had a per progenitor output (as measured in read counts in myeloid cells) as high as the per progenitor output (read counts in myeloid cells) of MPP producing both myeloid cells and erythroblasts (Figure S7D), and (2) their lack of output in erythroblasts was reproducible in technical replicates before applying the replicate filter (data not shown). Note that we also detected a significant fraction of MPPs that only show DC output (Figure 4B), as has also been observed for lymphoid-primed multipotent progenitors (Naik et al., 2013). These data show the existence of a subpopulation of true common progenitors to myeloid and erythroid lineages within the MPP pool, potentially within the recently described $CD41^+$ cell population (Miyawaki et al., 2015). However, within the MPP population, divergence of a myeloid biased subpopulation is already apparent.

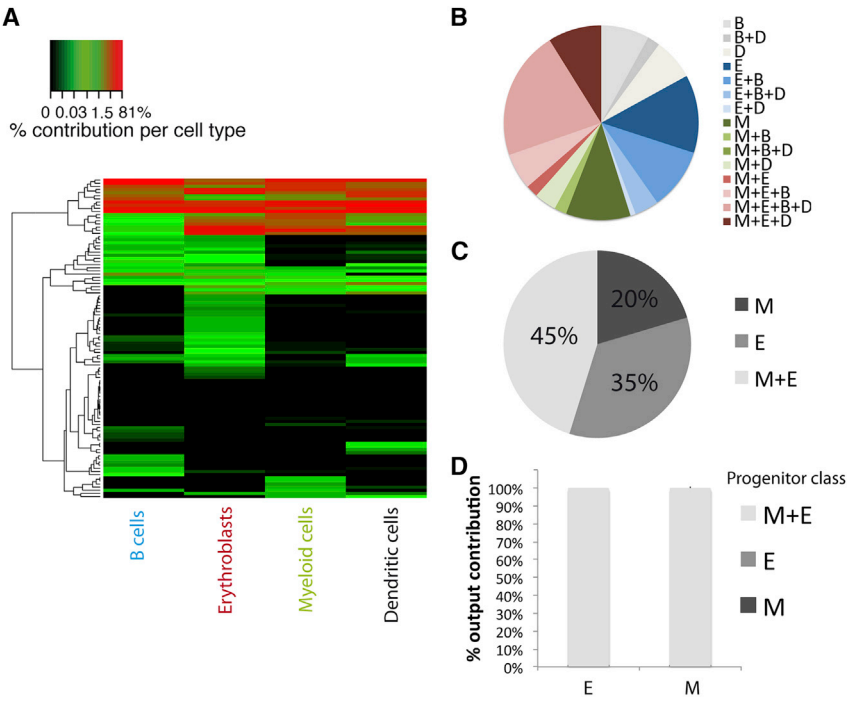


Figure 3. Common Origin of Erythrocytes and Myeloid Cells within HSC, Related to Figures S4 and S5

Barcode-labeled CD150$^+$ HSCs were transplanted into 5gy-irradiated mice. At day 27 post-transplantation, myeloid cells (CD11b$^+$), erythroblasts (Ter119$^+$), B cells (CD19$^+$), and DCs (CD11c$^+$) were recovered from spleen and barcode content was analyzed.

(A) Heatmap representation of the output of individual CD150$^+$ HSC (rows) to different cell types (columns), arcsine transformed data.

(B) Proportion of HSCs classified using a 1% classification threshold in the different possible categories (averaged over mice).

(C) Same analysis as in (B) but examining erythroblast and myeloid cell output only.

(D) Quantitative contribution of the different classes to myeloid cell and erythroblast production (mean + SD).

All of the results come from four mice per experiment. Quality controls and other thresholds in Figure S5.

DISCUSSION

The main finding of this work is that the population classically defined as common myeloid progenitors is highly heterogeneous, with most individual CMPs only yielding erythrocytes or only yielding myeloid cells after transplantation and with some CMPs contributing toward the DC lineage only. A minor subpopulation is observed that could represent a rare but true subset of bipotent CMP but may also reflect a contamination of LSK+ cells.

As the existence of a "common" erythro-myeloid progenitor has been the basis for the model in which the first step of commitment is formed by the separation between CMP and CLP, our findings suggest a revised model of hematopoiesis. Miyawaki et al. (2015) have recently identified a more immature CMP within the MPP population. Existence of such a common erythro-myeloid progenitor is still consistent with a model in which the first step of commitment is formed by the myeloid-lymphoid split. Likewise, our data also show that a true common erythro-myeloid progenitor exists within the MPP population. Importantly though, 40% of the total output of myeloid cells was derived from progenitors that produced neither erythrocytes nor lymphoid cells (Figure S7C). These data indicate that, next to the classical erythro-myeloid versus lymphoid split, an additional myeloid-only commitment already manifests itself at the MPP stage.

Alternative models for hematopoiesis have been proposed in which the megakaryocyte-erythroid lineage branches first, followed by a myeloid-lymphoid split (Adolfsson et al., 2005; Arinobu et al., 2007). In this model, direct erythropoiesis from HSC or MPP is predicted to occur independently of myelopoiesis and lymphopoiesis. We did not observe evidence for a significant role of such an erythroid bypass from either HSC or MPP. Specifically, of total erythroblast production from HSC or MPP, respectively, only 0.1% ± 0.1% and 9% ± 0.01% was derived from cells that did not yield myeloid or B cells. Elegant recent in vivo lineage tracing analysis suggests that transplantation can skew output of HSCs toward a few dominant clones (Sun et al., 2014). However, at least under transplantation conditions, an early erythroid commitment (Yamamoto et al., 2013) is at best a minor pathway.

The experimental measurements that we make in our work are the in vivo output of single cells. As such, the strong bias in single-cell output that we observe could result from the commitment pre-injection of the cells that are used for reconstitution or from a dominant effect of the niche in which an individual progenitor cell finds itself after transplantation. We strongly believe that the biased output that we observe for individual progenitor cells originates from pre-commitment for two reasons. First, a pre-commitment of, for instance, individual CMP to produce either erythrocytes or myeloid cells implies an "encoding" of this commitment at the genetic level, and the observation of substantial transcriptional heterogeneity between individual cells within the murine CMP pool by Amit and coworkers (Franziska et al., 2015 [in this issue of *Cell*]) provides strong support for such a model. Second, if the commitment toward the production of a single type of output is pre-existing rather than induced by the niche that is encountered, one would expect to also observe heterogeneous behavior under in vitro conditions in which individual cells all experience identical conditions. In support of this, Notta et al. (2015) have provided strong evidence that human bone marrow CMPs and MPPs are highly heterogeneous and composed of subpopulations with uni-outcome myeloid, erythroid, and megakaryocyte potential in in vitro assay systems.

Collectively with these data, our single-cell-tracing data provide evidence for a model in which multiple combinations of

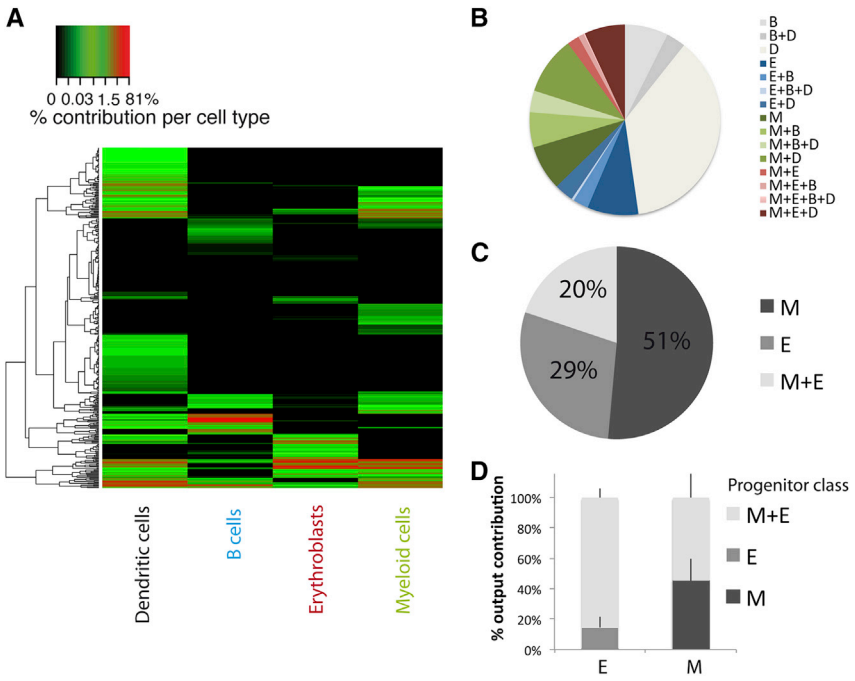


Figure 4. Divergence between Erythrocyte and Myeloid Output within the MPP Pool, Related to Figures S6 and S7

Barcode-labeled MPPs were transplanted into 5gy-irradiated mice. At day 14 post-transplantation, myeloid cells (CD11b$^+$), erythroblasts (Ter119$^+$), B cells (CD19$^+$), and DCs (CD11c$^+$) were recovered from spleen and barcode content was analyzed.

(A) Heatmap representation of the output of individual MPPs (rows) to different cell types (columns), arcsine transformed data.

(B) Proportion of MPPs classified using a 1% classification threshold in the different possible categories (average over mice).

(C) Same analysis as (B) but examining erythroblasts and myeloid cell output only.

(D) Quantitative contribution of the classes to myeloid cell and erythroblast production from barcode-labeled MPPs (mean + SD).

All results come from four mice per experiment. Quality controls and other thresholds in Figure S7.

lineage commitments (e.g., erythroid-myeloid, myelo-lymphoid) are already manifest during the transition of the MPP stage of hematopoietic development. The combination of single-cell transcriptomics and single-cell-lineage tracing, preferably under conditions of unperturbed hematopoiesis (Sun et al., 2014; Busch et al., 2015) should help reveal how sibling cells commit to different types of output at this stage.

EXPERIMENTAL PROCEDURES

Mice, Progenitor Isolation, and Barcode Labeling

C57BL/6J (CD45.2) donor mice and C57BL/6 Pep3b (CD45.1) recipient mice irradiated with 5 gy were used in all experiments. Labeling of cells with the lentiviral barcode library was performed as described in Naik et al. (2013). In brief, bone marrow was harvested from femurs, tibias, and ilia and enriched using anti-CD117 magnetic beads (Miltenyi). The c-kit$^+$ fraction was stained with antibodies against CD16/32 (FITC, clone 24G2, BD PharMingen or PerCPCy5.5, clone 24G2, BD Biosciences), CD117 (c-kit APC, clone 2B8, Biolegend), Sca-1 (Pacific Blue, clone D7, Biolegend), CD135 (Flt3 PE, clone A2F10, ebioscience), CD34 (alexa700, clone RAM34, ebioscience), and CD150 (Slam PEcy7, clone TC15-12F12.2, Biolegend). MPPs, CMPs, and HSCs were sorted using a strategy similar to that described previously (Akashi et al., 2000) (Figures 1B and 1C). For the experiment, up to 10^5 progenitors were transduced for 6 hr in stem cell medium (stempan SFEM, Stem Cell Technologies) with 50 ng/ml stem cell factor (SCF). For the first 90 min of transduction, progenitors were centrifuged at 200 rpm (low brake) and cells were subsequently incubated at 37°C for 4.5 hr. Following transduction, cells were washed, resuspended in saline solution, and injected intravenously into at least two recipient mice. Efficiency of transduction ranged between 10%–40%.

Cell Isolation

Cell suspensions were derived from spleen or bone marrow of killed mice and positively selected for Ter119 expression using biotinylated anti-Ter119 antibody (BD PharMingen) and anti-biotin beads (Miltenyi). The Ter119$^-$ fraction was stained at 4°C using a combination of antibodies against CD45.2 (Pacific blue, clone 104, Biolegend), CD11C (APC, clone HC3, BD biosciences), CD11b (PercPCy5.5, clone M1/70, ebioscience), CD19 (APC-Cy7, clone 1D3, BD PharMingen), and Ter119 (PEcy7, BD PharMingen) to sort myeloid cells, B cells, and DCs (Figure 1D). The Ter119$^+$ fraction was stained at 4°C with antibody against Ter119 (Pecy7, BD PharMingen) to sort the erythroblast fraction (Figure 1D). Output cell types were sorted following the gating strategy in Figure 1D, and cell numbers are depicted in Table S1.

PCR and Sequencing

Sorted cell samples were lysed in Viagen buffer (direct PCR, Viagen) and split into technical replicates (Schumacher et al., 2010). Barcodes were amplified by nested PCR as previously described in Naik et al. (2013). In brief, barcode sequences were first amplified using top-LIB (5′TGCTGCCGTCAACTAG AACA-3′) and bot-LIB (5′GATCTCGAATCAGGCGCTTA-3′) primers (30 cycles: 15 s at 95°C; 15 s at 57.2°C; 15 s at 72°C). PCR products were then subjected to a second PCR (30 cycles: 5 s at 94°C; 5 s at 57.2°C; 5 s at 72°C) in which the P5 and P7 (5′-CAAGCAGAAGACGGCATACGAGAT-3′) adaptors required for sequencing on Illumina platforms as well as a unique 8 bp index and a sequencing primer annealing site were attached. Primers used were as follows: forward (5′-P5-seq_prim-index-CAGGCGCTTAGGATCC-3′) and reverse (5′-P7-TGCTGCCGTCAACTAGAACA-3′). Up to 192 tagged PCR products were pooled and sequenced by next-generation sequencing using a HiSeq2000 or HiSeq2500 platform (Illumina). Total read lengths were 50–60 bp, of which the first 22 bp consisted of the index and constant regions. The following 15 bp were used to distinguish between different barcodes.

Filtering Procedure

Sequence reads were first screened for quality by the software packages that are part of the Illumina pipeline, and only high-quality reads were selected (~10–70 × 10^6 per run). Subsequently, only reads showing a 100% match to the expected scheme (index-CAGGCGCTTAGGATCC-random_sequence) were selected (typically >90% of reads), whereby a 100% match to one of the 192 index sequences was required. Of the random sequence (the barcode), only the first 15 bp were considered. To filter out reads that reflect PCR or sequencing errors within the 15 bp barcode region, all filtered reads were compared to a barcode reference list (Gerlach et al., 2013), and only sequences that showed a 100% match to one of the sequences within the barcode reference file were maintained (typically 80%–93%).

Using a customized script in R (R Development Core Team, 2014), we subsequently applied a filtering procedure composed of three steps, similar to the procedure used in Naik et al. (2013). Specifically, samples for which

insufficient read counts were obtained during deep sequencing (average of the two technical replicates $<10^3$, compared to the expected 10^5) were first excluded. Samples having passed this step were then normalized to 10^5 reads for each sample. To control the quality of the analysis of the barcode content of a given sample, we additionally excluded samples in which the two technical replicates displayed a Pearson correlation coefficient of <0.8. These technical replicate controls are essential to give confidence in whether barcode expansion and cell type recovery were sufficient for informative lineage relationship analysis. Third, after passing these technical replicate quality controls, all reads of barcodes present in only one of the two technical replicates of a given sample—an indication that there was a low confidence for inclusion of that barcode—were changed to zero (0) reads for that sample. Resulting data were collated in a table of normalized barcode read counts in each cell type.

Barcode Analysis and Barcode Classification

For heatmap analysis, read counts were transformed using the hyperbolic arcsine function that is similar to a logarithmic function but can accommodate barcode reads with a value of 0. After confirming that technical replicates yielded similar data, as described above and by visual inspection (Figures S1A, S5A, and S7A), their average was taken for further analysis.

To classify individual progenitors by their lineage bias, we used a previously published hand-tailored classifier (Naik et al., 2013). In summary, an additional normalization per progenitor was applied in each mouse, thereby enabling categorization into classes of biased output toward the analyzed lineages (myeloid cells, erythroblasts, DCs, or B cells, depending on the experiment). Potency to contribute to a given lineage was assigned if >1% of the output of a given CMP, MPP, or HSC was toward that lineage (Figures 2–4). Other thresholds were also analyzed (Figures S2, S3, S5, and S7).

Quality Controls

To illustrate the quality of the samples, the barcodes found in both duplicates of the same sample are always compared and representative plots of each experiment are provided in the supplemental figures (Figures S2, S3, S5, S7). Furthermore, by cross-comparing two mice that received barcoded progenitors from the same transduction batch, the extent of "repeat use"—instances in which two progenitors are transduced with virus particles that harbor the same barcode—was always compared and was found to be minimal (Figures S2, S3, S5, S7).

Uni-outcome progenitors were further analyzed for potential detection issues by: (1) determining whether the distribution of clonal outputs of uni-outcome progenitors toward one specific lineage was similar to the clonal output toward the same lineage of multi-outcome progenitors, and (2) determining whether the lack of output of uni-outcome progenitors toward the other lineages was reproducible in technical replicates before applying the replicate filter.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.11.059.

AUTHOR CONTRIBUTIONS

L.P. designed experiments, performed experiments, and analyzed data; K.D. performed data analysis; L.K. performed experiments; R.J.d.B. supervised data analysis; T.S. supervised the study; L.P. and T.S. wrote the paper, incorporating feedback from all coauthors.

ACKNOWLEDGMENTS

The authors thank A. Pfauth and F. van Diepen for cell sorting, J. Urbanus for lentivirus production, A. Velds and R. Kluin for computational assistance, and S. Naik for discussion. The data and the code presented in this manuscript are available upon request. This work was supported by Human Frontier Science Program grant RGP0060/2012 (to T.S. and K.D.), Netherlands Organization for Scientific Research grant TOPGO.L.10.042 (to T.S. and R.J.d.B.), ERC grant Life-His-T (to T.S.), Science Foundation Ireland grant 12 IP 1263 (to K.D.), an ATIP-Avenir grant from CNRS and Bettencourt-Schueller Foundation (to L.P.), and grants from the Labex CelTisPhyBio (No. ANR-10-LBX-0038) and Idex Paris-Science-Lettres Program (ANR-10-IDEX-0001-02 PSL) (to L.P.).

Received: July 20, 2015
Revised: November 18, 2015
Accepted: November 28, 2015
Published: December 17, 2015

REFERENCES

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A., 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.

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

Arinobu, Y., Mizuno, S., Chong, Y., Shigematsu, H., Iino, T., Iwasaki, H., Graf, T., Mayfield, R., Chan, S., Kastner, P., and Akashi, K. (2007). Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell Stem Cell *1*, 416–427.

Benz, C., Copley, M.R., Kent, D.G., Wohrer, S., Cortes, A., Aghaeepour, N., Ma, E., Mader, H., Rowe, K., Day, C., et al. (2012). Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. Cell Stem Cell *10*, 273–283.

Boyer, S.W., Schroeder, A.V., Smith-Berdan, S., and Forsberg, E.C. (2011). All hematopoietic cells develop from hematopoietic stem cells through Flk2/Flt3-positive progenitor cells. Cell Stem Cell *9*, 64–73.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S.M., Reth, M., Höfer, T., and Rodewald, H.R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature *518*, 542–546.

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.

Dykstra, B., Kent, D., Bowie, M., McCaffrey, L., Hamilton, M., Lyons, K., Lee, S.J., Brinkman, R., and Eaves, C. (2007). Long-term propagation of distinct hematopoietic differentiation programs in vivo. Cell Stem Cell *1*, 218–229.

Forsberg, E.C., Serwold, T., Kogan, S., Weissman, I.L., and Passegué, E. (2006). New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. Cell *126*, 415–426.

Franziska, P., Arkin, Y., Giladi, A., Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Meital, G., Weiner, A., et al. (2015). Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell *163*, this issue, 1663–1677.

Gerlach, C., Rohr, J.C., Perié, L., van Rooij, N., van Heijst, J.W., Velds, A., Urbanus, J., Naik, S.H., Jacobs, H., Beltman, J.B., et al. (2013). Heterogeneous differentiation patterns of individual CD8+ T cells. Science *340*, 635–639.

Gerrits, A., Dykstra, B., Kalmykowa, O.J., Klauke, K., Verovskaya, E., Broekhuis, M.J., de Haan, G., and Bystrykh, L.V. (2010). Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood *115*, 2610–2618.

Igarashi, H., Gregory, S.C., Yokota, T., Sakaguchi, N., and Kincade, P.W. (2002). Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow. Immunity *17*, 117–130.

Iwasaki, H., Somoza, C., Shigematsu, H., Duprez, E.A., Iwasaki-Arai, J., Mizuno, S., Arinobu, Y., Geary, K., Zhang, P., Dayaram, T., et al. (2005). Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. Blood *106*, 1590–1600.

Kondo, M., Weissman, I.L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell *91*, 661–672.

Lai, A.Y., and Kondo, M. (2006). Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. J. Exp. Med. *203*, 1867–1873.

Lu, R., Neff, N.F., Quake, S.R., and Weissman, I.L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat. Biotechnol. *29*, 928–933.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., et al. (2007). Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. Immunity *26*, 407–419.

Miyawaki, K., Arinobu, Y., Iwasaki, H., Kohno, K., Tsuzuki, H., Iino, T., Shima, T., Kikushige, Y., Takenaka, K., Miyamoto, T., and Akashi, K. (2015). CD41 marks the initial myelo-erythroid lineage specification in adult mouse hematopoiesis: redefinition of murine common myeloid progenitor. Stem Cells *33*, 976–987.

Naik, S.H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature *496*, 229–232.

Naik, S.H., Schumacher, T.N., and Perié, L. (2014). Cellular barcoding: a technical appraisal. Exp. Hematol. *42*, 598–608.

Nakorn, T.N., Miyamoto, T., and Weissman, I.L. (2003). Characterization of mouse clonogenic megakaryocyte progenitors. Proc. Natl. Acad. Sci. USA *100*, 205–210.

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. (2015). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science. Published online November 5, 2015. http://dx.doi.org/10.1126/science.aab2116.

Nutt, S.L., Metcalf, D., D'Amico, A., Polli, M., and Wu, L. (2005). Dynamic regulation of PU.1 expression in multipotent hematopoietic progenitors. J. Exp. Med. *201*, 221–231.

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.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105–111.

Sanjuan-Pla, A., Macaulay, I.C., Jensen, C.T., Woll, P.S., Luis, T.C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., et al. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nature *502*, 232–236.

Schepers, K., Swart, E., van Heijst, J.W., Gerlach, C., Castrucci, M., Sie, D., Heimerikx, M., Velds, A., Kerkhoven, R.M., Arens, R., and Schumacher, T.N. (2008). Dissecting T cell lineage relationships by cellular barcoding. J. Exp. Med. *205*, 2309–2318.

Schumacher, T.N., Gerlach, C., and van Heijst, J.W. (2010). Mapping the life histories of T cells. Nat. Rev. Immunol. *10*, 621–631.

Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. Nature *514*, 322–327.

Takano, H., Ema, H., Sudo, K., and Nakauchi, H. (2004). Asymmetric division and lineage commitment at the level of hematopoietic stem cells: inference from differentiation in daughter cell and granddaughter cell pairs. J. Exp. Med. *199*, 295–302.

R Development Core Team (2014). R: A language and environment for statistical computing (R Foundation for Statistical Computing). http://www.R-project.org/.

Terszowski, G., Waskow, C., Conradt, P., Lenze, D., Koenigsmann, J., Carstanjen, D., Horak, I., and Rodewald, H.R. (2005). Prospective isolation and global gene expression analysis of the erythrocyte colony-forming unit (CFU-E). Blood *105*, 1937–1945.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K.L., Ema, H., and Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell *154*, 1112–1126.

Yoshida, T., Ng, S.Y., Zuniga-Pflucker, J.C., and Georgopoulos, K. (2006). Early hematopoietic lineage restrictions directed by Ikaros. Nat. Immunol. *7*, 382–391.