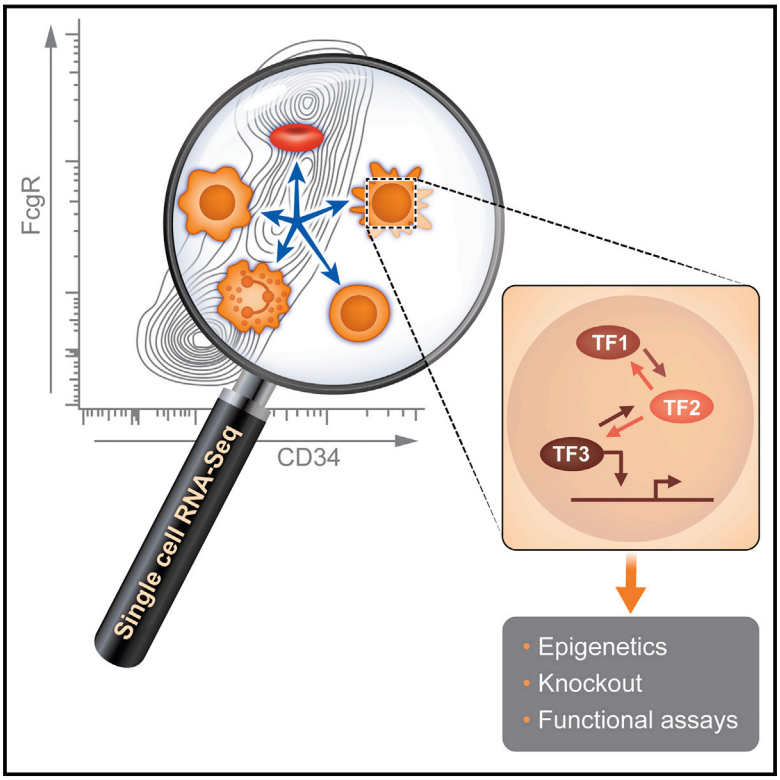
Article

Cell

Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors

Graphical Abstract



Authors

Franziska Paul, Ya'ara Arkin, Amir Giladi, ..., Bo Torben Porse, Amos Tanay, Ido Amit

Correspondence

amos.tanay@weizmann.ac.il (A.T.), ido.amit@weizmann.ac.il (I.A.)

In Brief

Single-cell transcriptomic analysis of bone marrow myeloid progenitor populations reveals early transcriptional priming toward seven different fates and absence of progenitors of mixed lineages, challenging the current models of hematopoiesis based on progressive loss of differentiation potential.

Highlights

- Transcriptionally primed single-cell subpopulations in early myeloid progenitors
- Transcription factors and epigenetic landscapes that regulate myeloid priming
- Mixed lineage states are not observed but appear when regulation is perturbed
- New reference model for studying hematopoiesis at single-cell resolution

Accession Numbers

GSE72857
GSE72858
GSE72859

Paul et al., 2015, Cell *163*, 1663–1677
December 17, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2015.11.013

Cell

Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors

Franziska Paul,[1,9] Ya'ara Arkin,[2,9] Amir Giladi,[1,9] Diego Adhemar Jaitin,[1] Ephraim Kenigsberg,[2] Hadas Keren-Shaul,[1] Deborah Winter,[1] David Lara-Astiaso,[1] Meital Gury,[1] Assaf Weiner,[1] Eyal David,[1] Nadav Cohen,[2] Felicia Kathrine Bratt Lauridsen,[3,4,5] Simon Haas,[6] Andreas Schlitzer,[7,8] Alexander Mildner,[1] Florent Ginhoux,[7] Steffen Jung,[1] Andreas Trumpp,[6] Bo Torben Porse,[3,4,5] Amos Tanay,[2,10,*] and Ido Amit[1,10,*]

[1]Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel

[2]Department of Computer Science and Applied Mathematics and Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel

[3]The Finsen Laboratory, Rigshospitalet, University of Copenhagen, Copenhagen 2200, Denmark

[4]Biotech Research and Innovation Centre (BRIC), Copenhagen 2200, Denmark

[5]Danish Stem Cell Centre (DanStem) Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

[6]Division of Stem Cells and Cancer, Deutsches Krebsforschungszentrum (DKFZ), 69120 Heidelberg, Germany

[7]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), BIOPOLIS 138648, Singapore

[8]Genomics and Immunoregulation, Life and Medical Sciences (LIMES) Institute, University of Bonn, 53115 Bonn, Germany

[9]Co-first author

[10]Co-senior author

*Correspondence: amos.tanay@weizmann.ac.il (A.T.), ido.amit@weizmann.ac.il (I.A.)

http://dx.doi.org/10.1016/j.cell.2015.11.013

SUMMARY

Within the bone marrow, stem cells differentiate and give rise to diverse blood cell types and functions. Currently, hematopoietic progenitors are defined using surface markers combined with functional assays that are not directly linked with in vivo differentiation potential or gene regulatory mechanisms. Here, we comprehensively map myeloid progenitor subpopulations by transcriptional sorting of single cells from the bone marrow. We describe multiple progenitor subgroups, showing unexpected transcriptional priming toward seven differentiation fates but no progenitors with a mixed state. Transcriptional differentiation is correlated with combinations of known and previously undefined transcription factors, suggesting that the process is tightly regulated. Histone maps and knockout assays are consistent with early transcriptional priming, while traditional transplantation experiments suggest that in vivo priming may still allow for plasticity given strong perturbations. These data establish a reference model and general framework for studying hematopoiesis at single-cell resolution.

INTRODUCTION

Hematopoiesis is the dynamic process by which self-renewing stem cells (HSCs) generate progeny that differentiate into progressively restricted cells of the erythroid, myeloid, or lymphoid lineages (Becker et al., 1963; Etzrodt et al., 2014; Orkin and Zon, 2008; Seita and Weissman, 2010; Wilson et al., 2009). This hierarchy is currently modeled using a canonical hematopoietic cell lineage tree (Figure S1A) that defines a set of progenitor and mature cell types (the tree nodes) and associates them with a discrete set of differentiation stages (the tree branches). In this model, each terminal cell type can be traced backward through the tree in a (unique) theoretical sequence of progenitors, ending up at the HSC origin. Despite the unequivocal evidence for the hierarchical and branching nature of hematopoietic differentiation, recent conflicting evidence relating to the function, differentiation fates, and unexplained heterogeneity within hematopoietic compartments calls for new approaches to complement our current models of blood and immune system development (Arinobu et al., 2007; Levine et al., 2015; Murre, 2007; Pronk et al., 2007; Schroeder, 2010; Wilson et al., 2008; Yamamoto et al., 2013).

The remarkable discoveries leading to the current models of hematopoiesis were founded on the technological breakthroughs enabling multicolored fluorescence-assisted cell sorting (FACS) using monoclonal antibodies. These technologies allow for the isolation and further analysis of cells that are characterized by the existence of specific combinations of surface markers (Seita and Weissman, 2010). Functional in vitro colony assays and in vivo transplantation assays led to characterization of the molecular and functional properties of such subpopulations (Akashi et al., 2000; Kondo et al., 1997, 2000) and formed the basis for our current models of hematopoietic hierarchy that are utilized extensively in both research and clinic. For example, the entire myeloid compartment (Figure S1A) can be isolated from bone marrow using a specific combination of molecular surface markers that separate the common myeloid progenitors (CMP) and their direct progenies: megakaryocyte/erythrocyte progenitors (MEP) and granulocyte/macrophage progenitors (GMP) (Akashi et al., 2000; Kondo et al., 2000). These cells have been shown to

lose their potential to differentiate into lymphocytes but are capable of giving rise to non-lymphoid white blood cells, red blood cells, and platelets—all of which feature a distinct morphology and carry out diverse functions. The success of marker-based approaches for dissecting hematopoiesis is dependent on the existence of cell-surface markers specific for distinct progenitors with highly organized and deterministic hierarchies. It also relies on the idea that captured cell populations are not perturbed during functional assays. However, recent studies suggest that these assays should be taken with some caution, as they involve severe perturbation of the host hematopoietic niche structure and major systemic stress in the recipient (Sun et al., 2014). Recently, as the functional and fate coherence of CMP and other progenitors is challenged by observations from multiple groups (Adolfsson et al., 2005; Arinobu et al., 2007; Bendall et al., 2014; Görgens et al., 2013; Levine et al., 2015; Murre, 2007; Pronk et al., 2007; Schroeder, 2010; Wilson et al., 2008; Yamamoto et al., 2013), it has become difficult to follow up functional evidence with continuous refinement of the tree model (Murre, 2007). Whether new and better markers can resolve this difficulty or whether it requires the incorporation of additional analytical and experimental tools is currently being debated (Etzrodt et al., 2014).

Differentiation of hematopoietic progenitors into distinct lineages is contingent upon activation of specific and tightly regulated gene expression programs, orchestrated by lineage-determining transcription factors (TFs) (Cantor and Orkin, 2002; Graf and Enver, 2009; Moignard et al., 2013; Orkin et al., 1998; Orkin and Zon, 2008; Pevny et al., 1991). Several factors have been shown to play critical roles in myeloid and erythroid development. *Pu.1* is an essential factor for reconstitution of the myeloid lineage (Graf and Enver, 2009). *Cebp-*α, *-*β, and *-*ε play major roles in commitment toward several myeloid cell types, primarily granulocytes, macrophages, and monocytes (Dahl et al., 2003; Laiosa et al., 2006; Tavor et al., 2002; Verbeek et al., 2001; Zhang et al., 1997). *Irf8* has also been implicated in myeloid lineage commitment, particularly for the monocyte and dendritic cell (DC) lineages (Becker et al., 2012; Kurotaki et al., 2014), while *Gata1*, *Klf1*, and *Gfi1b* are essential for erythrocyte and megakaryocyte lineage commitment (Iwasaki et al., 2003; Orkin et al., 1998; Pevny et al., 1991). The functions of many of these TFs are highly conserved from mouse to human, and their mutations cause severe lineage abnormalities and blood tumors (Cantor and Orkin, 2002; Rosenbauer and Tenen, 2007). Linking gene regulatory mechanisms to differentiation programs in myeloid populations thereby remains challenging, relying mostly on in vitro assays or perturbations followed by specific FACS readouts (Schroeder, 2010).

Single-cell RNA-seq enables unbiased characterization of cell types in complex tissues such as the spleen (Jaitin et al., 2014) or the brain (Zeisel et al., 2015). Applications to dynamic and partially differentiated cell populations such as those observed in hematopoiesis are currently restricted to the usage of conventional surface markers, single-cell qPCR, or mass cytometry (Bendall et al., 2014; Guo et al., 2013; Levine et al., 2015; Moignard et al., 2013). The results supported earlier data suggesting considerable heterogeneity within classically sorted hematopoietic populations, including CMP (Guo et al., 2013; Moignard et al., 2013; Pronk et al., 2007). These findings implicate deep genome-wide sampling by single-cell RNA-seq as an appropriate tool to systematically resolve myeloid progenitor heterogeneity and to assess whether early differentiation in this system occurs as classically suggested by hierarchical specification of CMP cells into the MEP and GMP lineages.

Here, we combine massively parallel single-cell RNA-seq (MARS-seq; Jaitin et al., 2014) with indexed FACS sorting, functional assays, chromatin profiling, genetic perturbation, and computational modeling to comprehensively characterize de novo the transcriptional and differentiation space of myeloid progenitor populations. Our integrated approach is defined by its sensitivity to uncover, without prior assumptions, in vivo subpopulations representing as little as 0.2% of the assayed compartment. This provides a definitive working model for the functional diversity of myeloid progenitors, defining 18 different subpopulations (and one outlier lymphocyte group) with variable degrees of transcriptional specificity toward erythrocytes, megakaryocytes, monocytes, dendritic cells, neutrophils, basophils, and eosinophils. Chromatin profiling suggests that transcriptional priming is coupled with in vivo developmental commitment. Functional assays are partially overlapping transcriptional sorting and chromatin profiling, which suggests that transplantation experiments may involve some degree of reprogramming that affects the in vivo priming of hematopoietic progenitors. Our model also outlines the potential activity of TF circuits within and between these subpopulations, and integration of genetic perturbation with single-cell RNA-seq enables further dissection of the TF drivers of the myeloid specification process. These data establish single-cell RNA-seq (Jaitin et al., 2014; Ramsköld et al., 2012) as a new tool for systematic characterization of developmentally plastic and highly dynamic cell populations in hematopoiesis and other tissues in both research and clinical applications.

RESULTS

Mapping the Transcriptional Landscape of Myeloid Progenitors

To characterize the distribution of transcriptional states within myeloid progenitor cells, we sorted c-Kit$^+$ Sca1$^-$ lineage (Lin)$^-$ bone marrow cells (Akashi et al., 2000) and sequenced their mRNA using MARS-seq (Experimental Procedures; Jaitin et al., 2014). Utilizing index sorting (Figure 1A; Experimental Procedures), we recorded the levels of the conventional surface markers that separate myeloid progenitors into CMP, MEP, and GMP (CD34 and FcgR) for each sorted single cell (Figures S1B–S1G; Experimental Procedures). Clustering analysis of 2,730 filtered single-cell profiles created a detailed map including 19 transcriptionally homogeneous subpopulations (Figures 1B and S2A–S2H). These clusters were based on differential expression patterns of >3,461 genes, allowing identification of subpopulations varying in frequency from 0.3% to 13%. Analysis of key marker genes (Figures 1C and S3) and global correlation analysis (Figure S2C) were used to order the transcriptional clusters and further examine their inter-cluster relationships. The sensitivity of

our clustering (i.e., n = 19) was determined using a two-stage process that first optimized a likelihood function and then ensured that no individual gene is maintaining a high intra-cluster variance (Experimental Procedures). We note that our model does not preclude the possibility of more sensitive or more coarse-grained clustering, since the population we analyze is not demarcated into simply defined cell types but involves variation at several levels. This resulted in specific and manually curated detection of all subpopulations with clearly distinct transcriptional profiles and in several clusters dissecting a continuum of broader groups of cells with strong co-variation among their genes. We found that, while some of the single-cell clusters (e.g. C1–C6, C14–C15, C16–C17) form groups that express marker genes corresponding to erythrocyte, monocyte, and neutrophil progenitors, other clusters (C7–C13) showed specific and distinct progenitor expression distributions that were not aligned with a simple hierarchical model of the population (Figures 1C and S3), as we discuss below. At the present sampling depth and MARS-seq coverage efficiency, our screen can detect subpopulations of 0.2% (5–6 cells) frequency or more. This provides us with a comprehensive and unbiased model for studying early myeloid and erythrocyte specification in the bone marrow.

Cell-Surface Markers Are Limited in Their Ability to Define Progenitor Populations

Analysis of the distribution of CD34 and FcgR indices for each of the above transcriptional clusters indicated that FACS markers are poor predictors for the single-cell structures identified by MARS-seq. We analyzed (Figures 1D–1E) the degree of correspondence between the standard FACS definitions for MEP, GMP, and CMP and the single-cell profiles, observing two types of FACS errors that are affecting current studies in the field. First, we observed misclassification based on CD34 and FcgR thresholds, leading, for example, to mispositioning of cells from clusters showing erythrocyte, neutrophil, monocyte, and basophil expression profiles out of the common GMP or MEP gate and into the CMP territory (e.g., C14, Figure 1E). Second, we observed lack of resolution in the selected marker classification, preventing separation of groups among themselves (Figure 1E). Lack of discriminative power was particularly noticeable in groups of progenitors that showed highly distinct transcriptional profiles (e.g., megakaryocyte, erythrocyte, DC, and basophil progenitors) but were clustered together as CMP based on FACS markers (C7–C12, Figure 1E). These effects explain some of the heterogeneity observed in CMP and the lack of clear understanding of the regulatory circuits and functional potential in this group. In the future, such data can improve surface marker selection to further enrich for subpopulation of interest (Figure S3). Nevertheless, our data suggest that any surface marker, and in particular CD34 and FcgR, is limited in its capacity to fully capture the internal state of progenitor cells and to faithfully reflect the mechanisms underlying its regulation and dynamic differentiation potential.

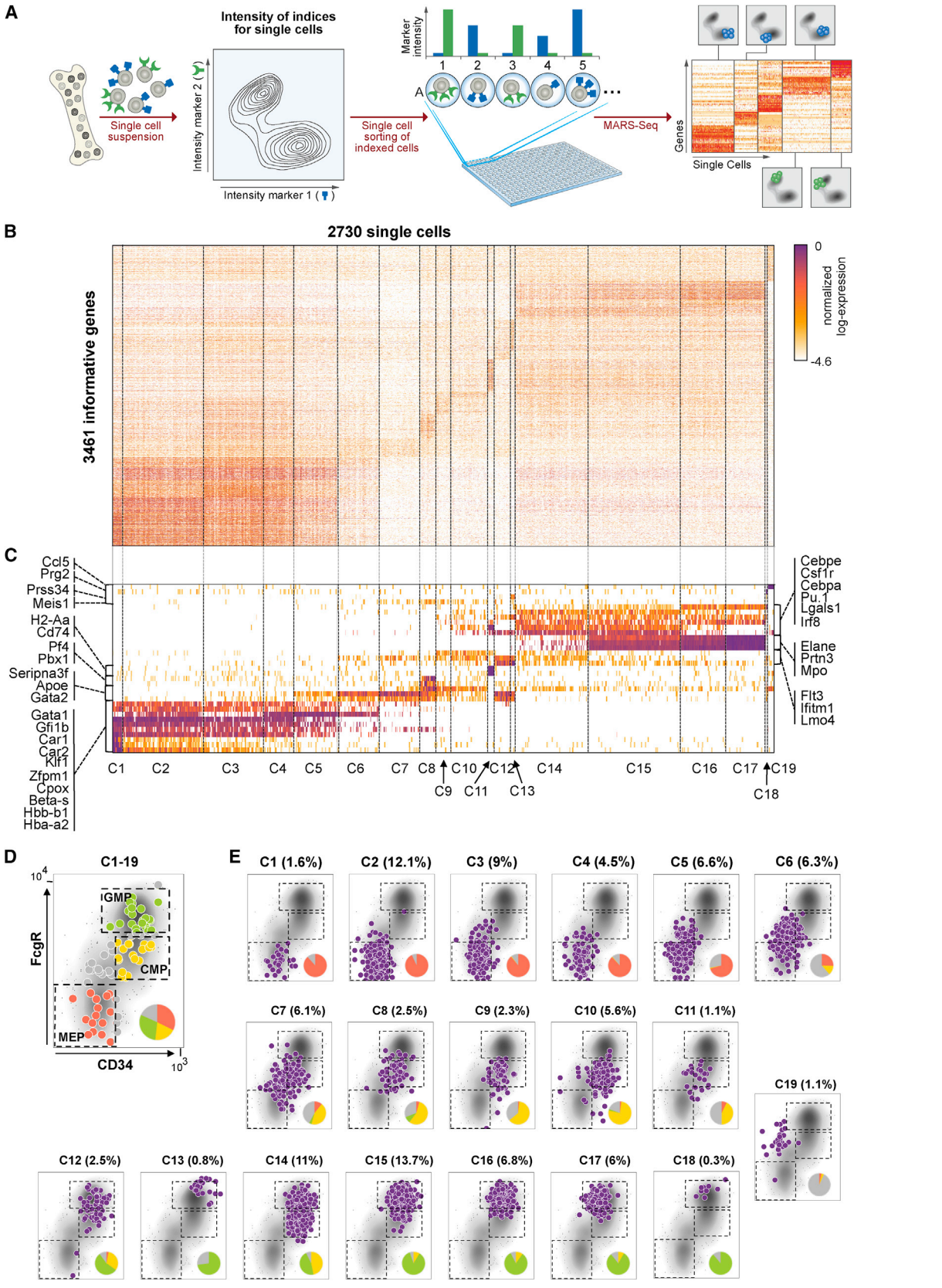
Transcriptional Regulators of the Myeloid Progenitor Populations

The sample depth and genome-wide nature of our single-cell transcriptional profiles enable us to quantitatively characterize cluster-specific gene expression. As shown in Figures 2A and 2B (and Figure S3B), we detected highly specific genes expressed almost exclusively in individual subpopulations. For example, expression of *Ccl5* associates a mis-sorted population of lymphoid progenitors (NK) with cluster 19, which we exclude from further follow up. Expression of *Prg2*, *Prss34*, *Cd74*, and *Pf4* specifically distinguishes clusters C18, C13, C11, and C8 and associates these with known signatures of eosinophil, basophil, DC, and megakaryocyte progenitors, respectively. Other marker genes distinguish broader groups of clusters. For example, *Car2* and *Mpo* discriminated between clusters C1–C6 (erythroid lineage progenitors) and clusters C12–C17 (myeloid lineage progenitors). A complete list of genes marking specific transcriptional subpopulations or defining broader classifications in the myeloid compartment is detailed in Table S3.

To characterize the gene-regulatory networks that underlie this marked diversification of transcriptional programs within the myeloid compartments, we quantified the expression level of all TFs that were correlated significantly with the identified cluster structure. We then linked TFs with clusters based on their expression signatures. This analysis resulted in a detailed model for putative gene-regulatory mechanisms in myeloid progenitors, including established as well as previously uncharacterized factors (Figures 2C and S3D–S3E).

Clusters C1–C6 showed strong association with known erythrocyte TFs, including *Klf1*, *Gata1*, *Gfi1b*, and factors that were not previously linked with erythropoiesis, including *Cited4* and *Phf10*. Interestingly, clusters C1–C7 represent a gradient of erythrocyte transcription, from early progenitor expression in cluster C7 toward cluster C1, which strongly expresses hemoglobin and other erythrocyte functional genes. This gradient is clearly reflected in the expression levels of *Klf1*, *Cited4*, and *Gfi1b* (Figure 2C). Other TFs showed different quantitative profiles along the gradient described by cluster C7-to-C1, suggesting that their regulation may participate in driving the differentiation process. Early progenitor TFs like *Gata2* and *Meis1* were still expressed in cluster C7 but were repressed in the clusters representing more differentiated progeny. Finally, several TFs, including cell-cycle regulators such as *Mbd4* and *E2f4*, were enriched in clusters C1–C7 while still showing expression in other clusters. Importantly, and unlike current models, none of these "MEP" clusters showed any expression of megakaryocyte markers or prominent megakaryocyte TFs and may therefore be termed erythrocyte progenitors (EP). Instead, both megakaryocyte markers (*Pf4* and *CD41*) and TFs (*Pbx1*, *Fli1*, *Mef2c*) were expressed in cluster C8, for which CD34 or FcgR marker intensities were incompatible with MEP standard gating and sorting. Given the single-cell data, we hypothesize that the erythrocyte clusters C7-to-C1 represent a developmental continuum of erythrocyte differentiation, which reflects progressive activation of the erythrocyte functional gene expression program and uncovers many previously unknown regulators of this process.

We observed a more combinatorial structure in clusters C12–C17. These clusters were characterized by high levels of both *Pu.1* and *Cebpa* (Figure 2C). Their subdivision into monocytes, granulocytes, and basophils was marked by specific expression of additional factors. *Irf8* (Kurotaki et al., 2014; Wang et al., 2014) was highly enriched in monocyte clusters (C14 and C15 in addition to specific enrichment in C11 as discussed below). *Cebpe*



(legend on next page)

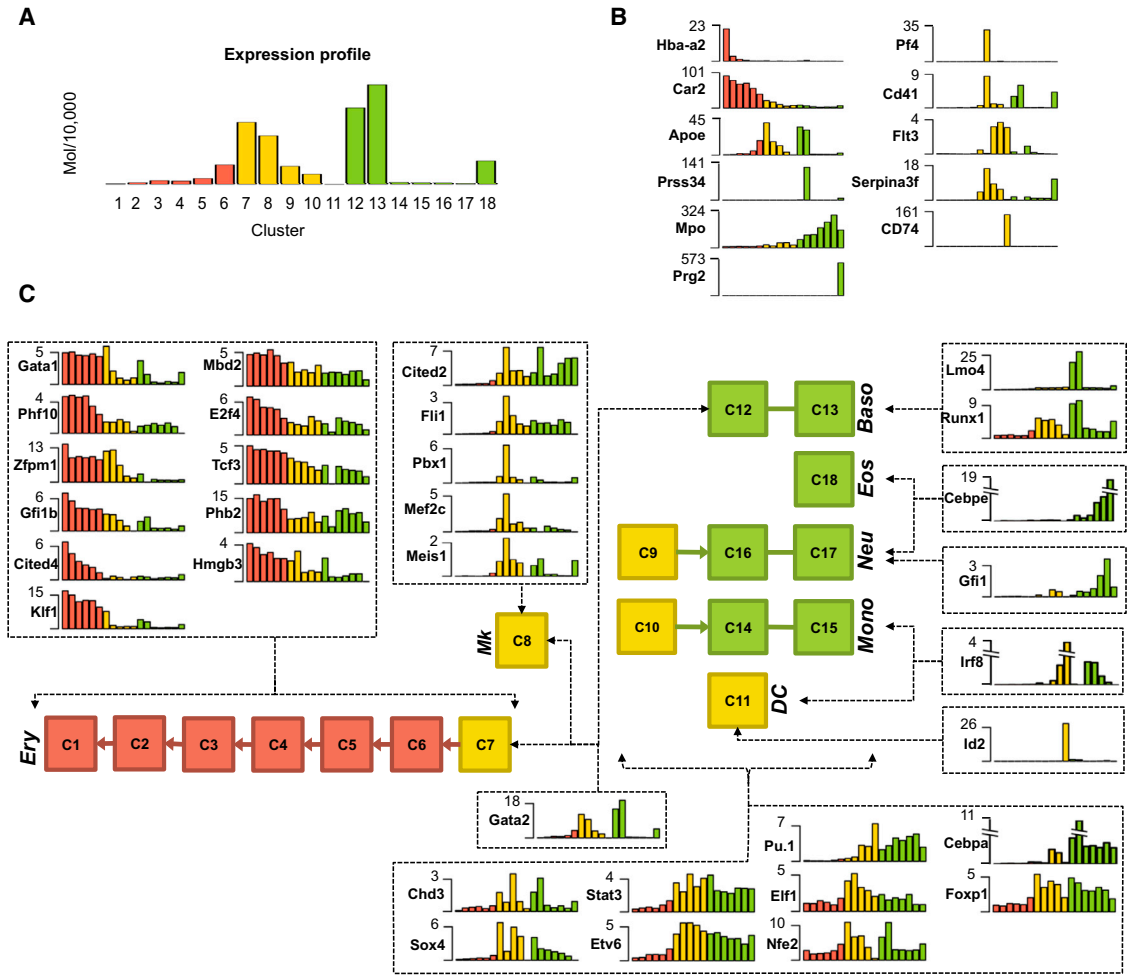


Figure 2. Transcriptional Networks Define Myeloid Progenitor Populations

(A) Schematic showing mean expression of a single gene across clusters C1–C18 colored by the index sorting territories.

(B) Mean expression per cluster (molecules/10,000 UMIs) of selected genes with expression constrained to one or several clusters.

(C) Bar graphs show mean expression per cluster (molecules/10,000 UMIs) for TFs that are associated specifically with the myeloid clusters C1–C18 (depicted as colored boxes). We grouped TFs and linked them with clusters (dashed lines) based on their expression. TF-cluster associations may therefore represent regulation of the cluster transcription program or of subsequent differentiation that is not captured within the myeloid progenitor populations.

and *Gfi1* were enriched in the neutrophil and eosinophil clusters (C16–C17 and C18), consistent with their known role in neutrophil differentiation and repression of the basophil lineage (Liu and Dong, 2012). *Lmo4* and *Runx1* were enriched in the basophil clusters (C12 and C13), suggesting that these TFs are major drivers of this lineage. We also detected TFs positively enriched for all of the above clusters (but not the erythrocyte clusters), including *Stat3*, *Etv6*, and *Foxp1*. Together, single-cell profiles provide a remarkably rich outline for the regulation of erythroid and myeloid lineages, distinguishing TFs with highly specific expression patterns from more global lineage regulators and allowing unbiased association of TFs to rare (e.g., basophil and

Figure 1. Unbiased Reconstruction of Myeloid Progenitors

(A) Schematic diagram of indexed single-cell sorting. Bone marrow suspension of heterogeneous cells (first panel) is single-cell sorted (without gating), while surface marker intensities are recorded (second panel) and linked to each sorted single cell by their well index (third panel). After MARS-seq (Jaitin et al., 2014), library preparation and sequencing, surface marker intensities are associated with single-cell transcriptome identities and plotted on an in silico FACS map.

(B) Expression of 3,461 differential genes across 2,730 myeloid progenitor (Lin^- c-Kit^+ $Sca1^-$) single cells. Cells are portioned into 19 distinct clusters (C1–C19) of gene expression profiles.

(C) Expression of key genes.

(D) Schematic showing index sorting intensities of 100 individual cells. Conventional CMP, GMP, and MEP gates are indicated by dashed boxes in the in silico FACS. Lower-right pie chart inlays show frequency of sorted cells falling into the CMP (yellow), GMP (green), or MEP (red) gates.

(E) FACS-measured FcgR and CD34 protein expression levels for single cells in each cluster (C1–C19) as in (D).

eosinophil) and more abundant (e.g., erythrocyte and neutrophil) groups. Linkage between expression of TF genes and target regulation cannot be automatically inferred from single-cell expression data alone due to possible time lags between expression of a TF and its downstream targets. Nevertheless, Figure S3A indicates that the physical targets (as inferred from chromatin immunoprecipitation sequencing [ChIP-seq] data) of the major putative TFs in our clusters show correlated expression to the TF gene dynamics outlined above.

Dissecting Heterogeneity between and within Small Progenitor Subpopulations

We defined clusters C7–C11 in our model (Figure 2C) as early progenitors since they lack strong expression of lineage-defining markers and TFs (e.g., *Car1,2/Klf1/Gata1* or *Mpo/Elane/Pu.1/Cebpa*) and showed expression of stem cell TF and marker genes (e.g., *ApoE*, *Meis1*, and *Serpina3f* [Riddell et al., 2014; Winkler et al., 2005]). Based on FACS, these subpopulations generally sort within the previously defined CMP territory, but their transcriptional distributions suggest more heterogeneity and subpopulation-specific expression than common stem cell characteristics. For example, cluster C8 was defined by high expression of *Pf4*, *Itga2b* (CD41), and the TFs *Pbx1*, *Mef2c*, and *Fli1*, which suggests a megakaryocyte fate. Cluster C11 was marked by distinctive expression of *Cd74* and other MHC-II related genes and by co-expression of the TFs *Irf8* and *Id2*, linking it with a DC fate. Cluster C7 showed early erythrocyte characteristics, with *Gata1* and *Klf1* already expressed at low levels. In both clusters C7 and C8, we observed high levels of *Gata2* and *Zfpm1*, linking the early megakaryocyte and erythrocyte progenitors by a possibly common regulatory mechanism. In contrast, clusters C9 and C10 were defined specifically by low but significant *Pu.1* and *Cepba* expression levels and also showed enrichment for several marker genes (e.g., *Flt3, Irf8* for C10; Figures 2B–2C and S3B–S3C), probably reflecting early monocyte and granulocyte progenitors.

To define the regulation and possible sub-structure of these early myeloid progenitors, we sought to increase their MARS-seq coverage by sub-sorting based on additional markers. We searched for currently available markers and transgenic mice that could enrich for three of our subpopulations: CD41$^+$ (*Itga2b*) to enrich for the C8 population expressing the megakaryocyte program (Mk), CD135 in combination with CD115 (Flt3$^+$/Csf1r$^+$) to enrich for the C10 (monocyte) subpopulation, and an Irf8 GFP-transgenic mouse along with MHC-II marker (Irf8$^+$/MHC-II$^+$) to enrich for population C11 (DC; Figure 3A). MARS-seq and projection of the derived RNA profiles onto our cluster model (Figures 3B and S4A) indicated that sub-sorting provided 68%, 53%, and 37% specificity for the targeted Mk, monocyte, and DC early progenitors, representing an enrichment by factors of 27, 10, and 35, respectively, when compared to standard gating protocols. This allows for sensitive analysis to identify structure within these small groups.

Analysis of the 230 cells that mapped to the Mk cluster C8 indicated that this group is homogeneously separated from the other sorted groups and is strongly marked by expression of Mk-related genes (Figures 3C and S4A). Importantly, clustering and additional analysis showed significant heterogeneity within the C8 population (Figure S4A), linking both Mk-specific genes and cell-cycle factors with several subgroups within C8. To define how the variation within this group is linked with the Mk fate, we formed five groups of cells by stratifying over normalized levels of the most strongly expressed Mk-specific gene in the data, *Pf4* (Figure 3D). We observed little correlation with the Pf4 groups for the regulator *Zfpm1*, suggesting that it is universally regulating this group. In contrast to this, we observed *Pf4*-contrasting patterns for the stem cell factors *ApoE* and *Gata2* and strong *Pf4* positive correlation with additional Mk genes, including *Vwf* and *Cd9*. Several additional genes were associated with negative and positive correlation to the Mk-specific signature represented by our stratification (Figure S4B). In summary, this analysis indicates that, within cluster C8, we can further detect a developmental continuum that is defined by a decrease in the expression of stem cell factors and a marked increase in Mk-specific gene expression program.

Interestingly, the 609 cells that were mapped to the monocyte cluster (C10) showed an analogous trend (Figures 3B, 3C, S4A, and S4C). We performed a stratified analysis using ten levels of *Mpo* gene expression and identified monocyte-specific genes (*Ly6e*, *Prtn3*, *Cstg*) as positively correlated with the *Mpo* signature and several other genes (including *Flt3* itself, *Ptpn18*, and *Lsp1*) as anti-correlated with the *Mpo* signature (Figure 3E). Similar to *Zfpm1* in cluster C8, cluster C10 subgroups showed generally stable levels for their main regulator, *Irf8*. Finally, the 417 cells that were mapped to the DC cluster C11 showed a more complex structure with multiple transcriptional profiles (Figures 3B, 3C, 3F, S4A, and S4D), including 5% (or 0.05% of the entire myeloid population) characterized by expression of *Cd7* and *Cfp* and 10% (0.1% of the entire population) characterized by expression of *Ccr7*, *Ccl5*, and *Ccl22*. Given the heterogeneity within the DC cluster, the origin of these cells and linkage to other clusters therefore remain unclear.

Although previous studies suggested some degree of heterogeneity within CMP—for example, identifying CD41 as a possible marker defining a subpopulation with Mk or naive potential (Miyawaki et al., 2015; Pronk et al., 2007)—our comprehensive MARS-seq screen of c-Kit$^+$ Sca1$^-$ Lin$^-$ bone marrow cells showed that, in fact, all myeloid progenitor subpopulations are transcriptionally primed toward one of at least seven different fates: erythrocytes, megakaryocytes, dendritic cells, monocytes, neutrophils, eosinophils, and basophils. Since such transcriptional priming may represent early developmental commitment or a more plastic regulatory state, we proceeded to examine the differentiation capacity of the myeloid clusters and the regulatory and epigenetic mechanisms that drive them.

Commitment of CMP Subpopulations in Transplantation Assays

One way to assess developmental potential of hematopoietic progenitors is through bone marrow transplantations (with the caveats of perturbing the in vivo context as discussed above). In order to evaluate the functional commitment of the CD41$^+$ and Flt3$^+$Csf1r$^+$ populations compared to total CMP, we sorted these populations from Ub-GFP transgenic mice. In these mice, GFP is ubiquitously expressed, including in platelets and erythrocytes (Yamamoto et al., 2013). We injected equal

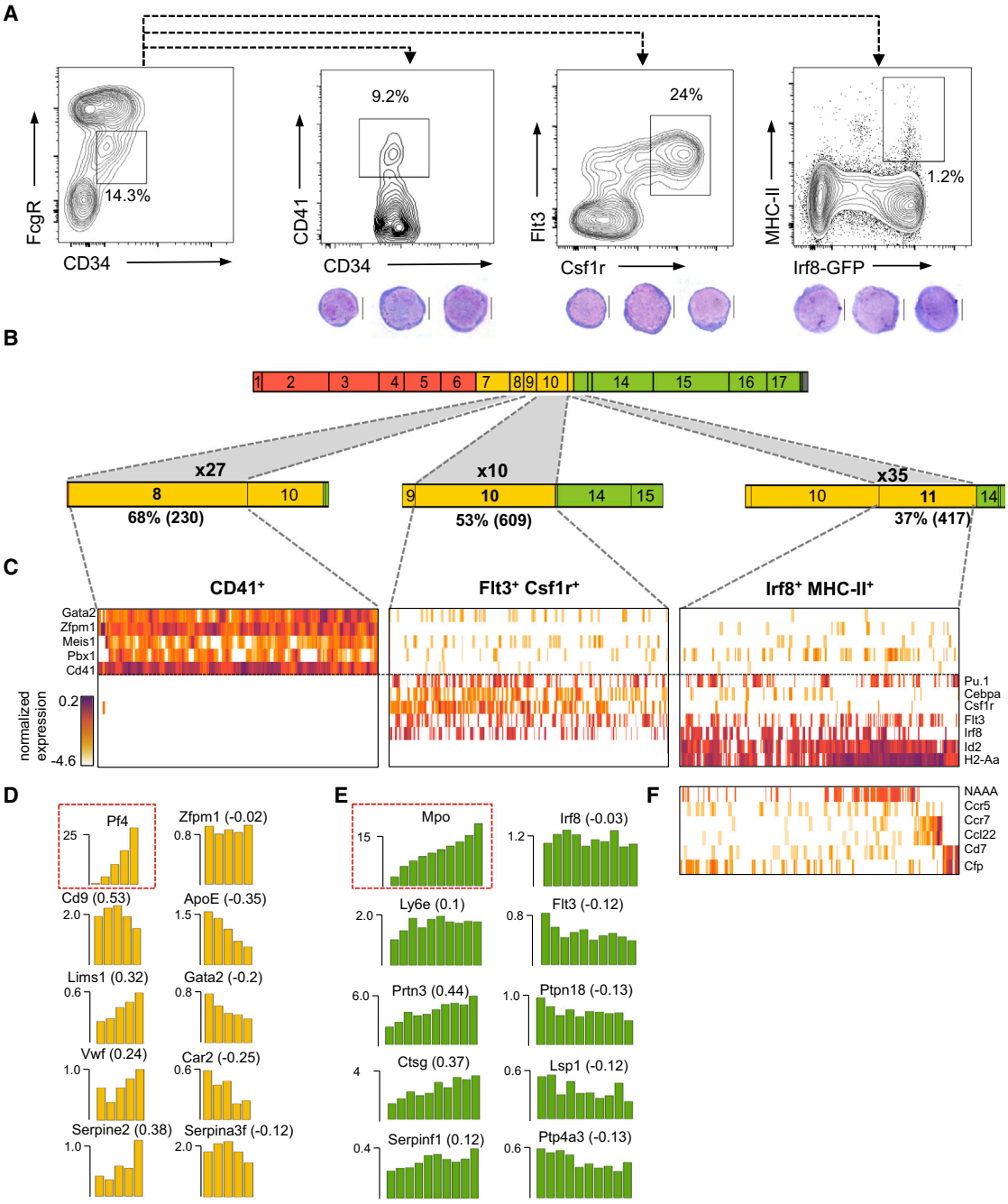


Figure 3. Surface-Marker-Based Enrichment of Myeloid Progenitor Subgroups

(A) Isolation of progenitor populations C8, C10, and C11 ($FcgR^{int}$ $CD34^{+}$ gates; Figure 1D) by the expression of cell-surface markers and transgenic markers: CD41, Flt3-Csf1r, and Irf8-MHC-II, respectively. These subtypes show hematopoietic progenitors' morphology as shown by Giemsa staining in lower panels (scale bar, 5 μm).

(B) Mapping of single-cell RNA-seq from sub-sorted populations to the myeloid model reveals enrichment of clusters C8 in $CD41^{+}$, C10 in $Flt3^{+}Csf1r^{+}$, and C11 in $Irf8^{+}MHC\text{-}II^{+}$ when compared to the original distributions (top). Colors are as indicated in Figure 2A.

(C) Color-coded expression of key genes that mark the enriched subpopulation. Data are shown for single cells that were mapped to the targeted expression cluster

(D and E) Shown are representative genes displaying intensified or repressed expression with *Pf4* expression (for C8, [D]) or *Mpo* expression (for C10, [E]). Expression correlations with *Pf4*/*Mpo* are shown in brackets.

(F) Expression profile of a subset of differential genes within the subgroups of $Irf8^{+}$ $MHC\text{-}II^{+}$ (C11) population.

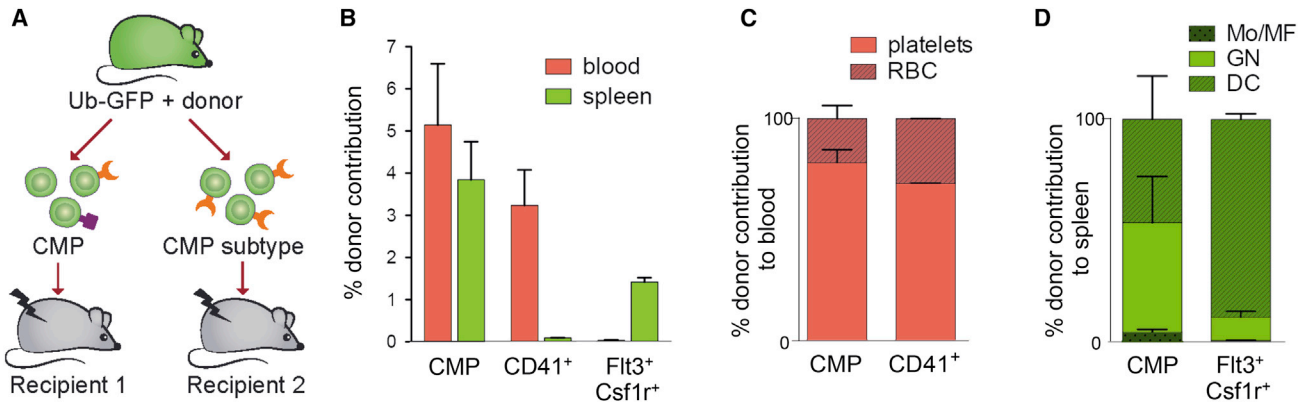


Figure 4. Myeloid Progenitors Are Partially Committed to Different Lineages In Vivo

(A) Myeloid progenitors (either CD41$^+$ [C8] or Flt3$^+$ Csf1r$^+$ [C10]) from Ub-GFP donor mice were injected into lethally irradiated recipients and the GFP$^+$ progeny analyzed compared to conventional CMP.

(B) Percent GFP$^+$ cells per 10,000 injected cells as measured in blood or spleen of recipient mice 10 days post transplantation.

(C) CD41$^+$ CMP (C8 enriched [Figure 3B]), which produced more GFP$^+$ progeny in the blood, contributed to erythrocyte and thrombocyte lineages; n = 2. RBC, red blood cells.

(D) Flt3$^+$Csf1r$^+$ CMP (C10 enriched [Figure 3B]), which differentially produced splenocytes, largely contributed to the DC lineage; n = 4. Mo/MF, monocytes/macrophages; GN, granulocytes. Mean + SEM. FACS plots are shown in Figure S5.

amounts of GFP$^+$ donor populations of interest mixed with unlabeled wild-type supportive bone marrow cells into lethally irradiated recipient mice (Figure 4A; Experimental Procedures). Analysis of blood repopulation (Figure S5A) showed that donor CD41$^+$ myeloid progenitors (C8-enriched) produced mostly platelets but also red blood cells (Figures 4B, 4C, and S5), whereas Flt3$^+$Csf1r$^+$ CMPs (C10-enriched) failed to generate either. In contrast, Flt3$^+$Csf1r$^+$ donor CMP were much more effective than CD41$^+$ myeloid progenitors in repopulating splenic niches (Figure 4B). Donor Flt3$^+$Csf1r$^+$ CMP preferentially produced dendritic cells, with a markedly lower production of granulocytes when compared to total CMP donor cells (Figures 4D, S5B, S5C). We failed to purify a sufficiently homogenous population of Irf8$^+$MHC-II$^+$ myeloid progenitors (C11), which prevented us from assessing the re-population potential of these cells. In summary, cells in cluster C8 and cluster C10 are restricted developmentally when compared to each other. Their heterogeneity cannot be regarded simply as transcriptional plasticity of a population with homogeneous differentiation fates. This restriction is, however, not complete, as observed for CD41$^+$ (C8) Mk progenitors that can also give rise efficiently to erythrocyte lineages, even though their transcriptional state is distinct from those of the erythrocyte clusters C1–C7 in our single-cell analysis.

Primed Enhancer Landscapes in Myeloid Progenitors

Development and lineage commitment involve changes in the chromatin landscape. In order to evaluate chromatin state changes in our subpopulations and to systematically identify putative enhancers that may regulate the myeloid differentiation process, we performed iChIP (Lara-Astiaso et al., 2014) on the three populations described above (enriched for C8, C10, and C11). Replicate experiments showed that iChIP was specific and sensitive for detecting thousands of genomic regions associated with H3K4me2 (enhancer and promoter) for each of the populations (Figures 5A–5I and S6A). Using these high-resolution maps of enhancers, we were able to annotate regulatory regions with differential H3K4me2 intensity, including genes specifically expressed in the Mk, monocyte, and DC progenitor populations. For example, we detected differential H3K4me2-marked regions for *Pf4* and *Vwf* in the megakaryocyte progenitor population, *Irf8* and *Csf1r* in the monocyte progenitors, and *Batf3* and *Cd40* in the DC progenitor population (Figures 5A–5I and S6A). Motif finding in subpopulation-specific enhancers further associated their epigenetic state with specific regulatory programs (Figure S6B), including *Gata2* for C8 and Pu.1 for C10–C11. To compare these three enhancer landscapes to previously characterized hematopoietic populations, we clustered enhancer loci into seven categories based on iChIP experiments for 12 cell types along the erythroid and myeloid branches (Lara-Astiaso et al., 2014; Figure 5J). We projected the CD41$^+$, Flt3$^+$/Csf1r$^+$, and Irf8$^+$/MHC-II$^+$ myeloid progenitor specific enhancer loci onto these clusters and calculated the overlap (Figures 5J and S6D–S6E). In addition we profiled the average H3K4me2 intensities for each of our populations across each category of enhancers (Figures 5K and S6C). According to these analyses, we found that CD41$^+$ myeloid progenitors were significantly more linked ($p < <10^{-10}$) to the enhancer landscapes of HSCs (progenitor I category) than the Flt3$^+$/Csf1r$^+$ and Irf8$^+$/MHC-II$^+$ group. On the other hand, the Irf8$^+$/MHC-II$^+$ population was significantly enriched with enhancers associated with DC ($p < <10^{-10}$). In summary, we show that the chromatin landscape of the three progenitor populations (C8, C10, C11) match their single-cell RNA-seq expression distribution, further supporting the idea that cells in these populations, under physiological conditions, are likely to commit to the fate defined by their transcriptional status.

Cebpa Knockout Blocks Granulocyte and Monocyte Initial Specification

Cebpa emerged in our detailed population maps as a major TF distinguishing myeloid (neutrophil, monocyte, basophils, but not DC) from erythroid and Mk populations, consistent with earlier studies of its function in myeloid development (Graf and Enver, 2009; Ye et al., 2013; Zhang et al., 1997). We used a conditional *Cebpa* knockout (KO) model (*Cebpa*$^{flox/flox}$*Mx1-Cre* [Lee et al., 1997]) and matching control (*Cebpa*$^{flox/flox}$; Figure S7A) and examined the structure of the resulting myeloid progenitor subpopulations at single-cell resolution. Three weeks prior to sorting, Mx1-Cre was activated by poly I:C injection (to both KO and control) and deletion efficiency was validated by FACS (Figure S7A). Sorting of *Cebpa* KO and control myeloid progenitors (Figures 6A and S7B) was followed by MARS-seq and mapping to the myeloid population cluster model (Figure 6B; Table S5; Experimental Procedures). Although these experiments were conducted in mice housed in another facility using a different experimental set-up (poly I:C injection), the myeloid population model in the control matched our C1–C19 cluster reference, with only slight differences in relative frequencies of the erythrocyte progenitor clusters (Figure 6B). Remarkably, the KO populations showed complete loss of all neutrophil, monocyte, and basophil clusters (C13–C18, Figure 6B).

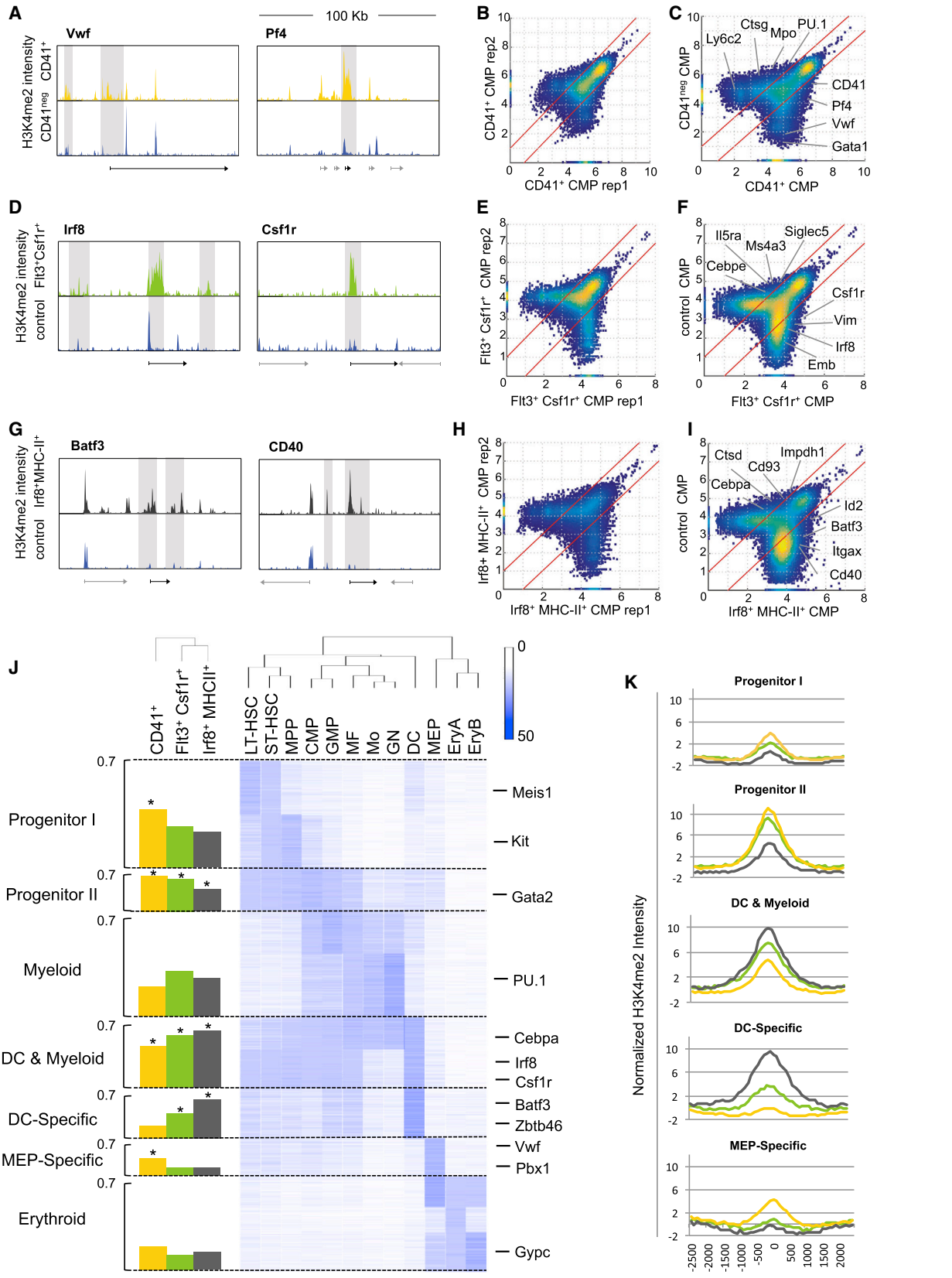
This loss was matched by dramatic increase in the frequency of the erythrocyte progenitor cluster C7 (4.5- to 6.3-fold; Experimental Procedures) and a smaller increase in the frequency of the monocyte cluster C10 (1- to 2.1-fold) but did not change the overall frequency of the erythrocyte and Mk progenitor spectrum (C1–C6 and C8) nor the expression of the TFs regulating them (Figures 6D and S7C–S7D). In-depth analysis of the expression of key regulating factors across myeloid progenitor clusters showed that the expression of *Pu.1* in expanded monocyte progenitor population (C10) was not affected by knockout of *Cebpa*, while *Irf8* expression was strongly diminished (Figure 6D). Clustering of 452 KO cells that were associated with the expanded C7 cluster showed that 407 out of the 452 cells were associated with normal erythrocyte progenitor expression profiles as defined by high levels of *Gata2*, *Zfpm1*, and *ApoE*. Importantly, we also detected a subpopulation within this cluster, which co-expresses unrelated genes that were not observed in the control mice, including the granulocyte gene *Mpo*, the erythrocyte β-globin gene (*Hbb-b1*, but not *Hba-a2*), and the lymphocyte genes *Gzmb* and *Tcb* (Figure 6C "outgroup"). This aberrant expression pattern supports a role for *Cepba* in repressing non-myeloid lineages and suggests that myeloid cells may express lymphocytic and erythrocytic genes in the absence of proper lineage-determining TFs. In summary, single-cell analysis identifies that *Cebpa* acts as a central and early regulator of the neutrophil, basophil, and monocyte fates. The accumulation of cells in cluster C7 following *Cebpa* loss suggests that the progenitor clusters C9 and C10 either compete with or are generated by differentiation of C7 cells. On the other hand, the remarkable stability of erythrocyte cluster C1–C6 abundance in the presence of a 5-fold expansion of cluster C7 progenitor population reflects tight control on the regulation of exit from this progenitor compartment toward erythropoiesis.

Cebpe Knockout Leads to Neutrophil Progenitor Arrest

Our analysis identified *Cebpe* as a factor specific to subpopulations expressing neutrophil or eosinophil genes (C16–C17 and C18). MARS-seq analysis of myeloid populations derived from a constitutive *Cebpe* KO model compared to matching control (Figures 7A and S7E–S7F; Yamanaka et al., 1997) indicated that the neutrophil clusters C16 and C17 were expanded by 4.5- and 4.8-fold, respectively (Figure 7B; see also Table S5). This suggests that *Cebpe* is not required for the initial differentiation toward a neutrophil progenitor state but for further differentiation and maturation out of the progenitor compartment. It also shows that, in a *Cebpe* KO background, cells that are committed toward a neutrophil fate but blocked from further maturation are incapable of switching their differentiation fate within the myeloid compartment (e.g., to expand monocyte clusters C14–C15 or others). Analysis of mean expression of neutrophil progenitor cells in control versus *Cebpe* KO mice identified relatively few genes that are directly regulated by *Cebpe* in this subpopulation (Figure S7G), also arguing toward a role for the factor later in the maturation processes. On the other hand, we identified strong induction of genes *S100a9*, *Lnc2*, *F10*, a possible weak induction of *Cebpa* in the knockout population (Figure 7C), and more global induction of *S100a8* in progenitors and myeloid clusters. We hypothesize that these changes are reflecting deregulated system-wide signaling into the bone marrow following *Cebpe* depletion and neutropenia. In conclusion, MARS-seq analysis of *Cebpa* and *Cebpe* KO myeloid progenitor populations presents a framework for systematic examination of perturbed/diseased in vivo myeloid differentiation dynamics coupled to identification of multiple regulatory pathways. Our analysis indicates that co-expression of TFs within subpopulations (e.g., as described in Figure 2C) may either reflect an early regulatory driver function (as shown here for *Cebpa*) or perform a function during later developmental stages (as shown here for *Cebpe*).

DISCUSSION

Recent observations from multiple studies challenged the highly organized and hierarchical model of the hematopoietic lineage tree (Paul and Amit, 2014). A refined and commonly accepted model of hematopoiesis was, however, difficult to develop thus far, since specific markers for progenitor subpopulations were not identified and commitment hierarchies based on bone marrow transplantation experiments involve disruption of the niche structure and natural development. By performing single-cell RNA-seq on thousands of mouse bone marrow single cells, we derive here a detailed map of the dynamic transcriptional states within the main myeloid progenitor populations. At single-cell resolution, we observed extensive transcriptional sorting of bone marrow progenitors into multiple clusters showing neutrophil, basophil, eosinophil, monocyte, DC, erythrocyte, and megakaryocyte properties. The data show that standard MEP definitions lead to exclusive isolation of the erythrocyte progenitor spectrum, while GMP and CMP definitions combine together multiple distinct groups with variable degrees of relatedness that we quantify and define here for the first time. Importantly, our results establish a framework for studying



(legend on next page)

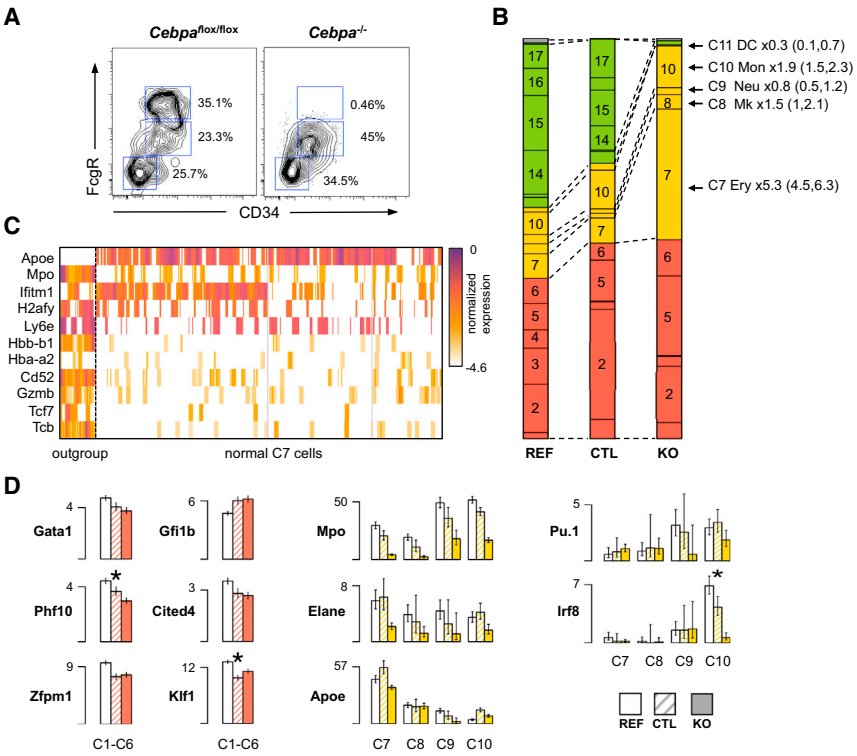


Figure 6. *Cebpa* Regulates Initial Differentiation into Myeloid Fates

(A) FACS sorting of myeloid progenitors from *Cebpa* control and KO mice. Relative frequencies (%) of sorted cells are shown.

(B) Changes in frequency of single cells mapping to clusters C1–C19 upon knockout of *Cebpa* (KO) as compared to *Cebpa*$^{flox/flox}$ control (CTL) and the wild-type reference (REF; Figure 1). Fold change of C1–C19 from reference is indicated with confidence intervals for selected clusters. Complete list of cluster fold changes and confidence intervals is shown in Table S5.

(C) Clustering of the expanded C7 erythroid progenitors in the KO identifies multi-phenotypic cell states ("outgroup") not present in REF or CTL mice.

(D) Mean expression of selected genes across clusters in *Cebpa* KO mice. x axis indicates molecule count per 10,000 counts. Asterisks mark transcription factors, which are differentially regulated compared to the control, as defined by Fisher's exact test. $q < 0.05$. Error bars represent 95% binomial confidence intervals, without correcting for multiple tests.

development and function in the hematopoietic compartment that go beyond the identification of distinct cell types. For example, we define a spectrum of gradual erythrocyte specification and complex combinatorial organization within early neutrophil, basophil, and monocyte progenitors. We further show that, by enriching for specific subpopulations prior to MARS-seq, we can enhance the resolution of our cluster maps and zoom in on small subpopulations of transcriptional clusters to identify additional developmental spectra within these subsets.

The observation of transcriptional sorting in the myeloid progenitor population is suggestive, but not proof, of the commitment of the observed transcriptional clusters to the developmental fates defined by their gene expression programs. Unlike previous reports (Miyamoto et al., 2002), in wild-type mice, we categorically do not observe cells that significantly express multiple lineage-specific genes or transcription factors regulating different fates. For example, cells expressing both erythrocyte factors (such as *Klf1* or *Gata1*) and myeloid (*Pu.1*, *Cebpa*), monocyte (*Irf8*), or neutrophil (*Cebpe*) TFs are never observed. This suggests that an intermediate state between these fates is rare in our gating strategy or only exists in a highly transient fashion. Additionally, knocking out the activity of the myeloid regulator *Cebpa* eliminates the buildup of transcriptionally sorted populations, suggesting that such sorting is not representing simple instability or stochastic fluctuations of a progenitor transcriptional program but is, rather, tightly regulated by distinct transcriptional circuits. Moreover, *Cebpa* KO bone marrow cells include a small subpopulation with highly aberrant transcriptional states marked by multi-phenotypic gene expression profiles that include erythrocyte, lymphocyte, and granulocyte genes. This result shows that such mixed states are possible in principle and are detectable by our technique but do not occur in wild-type transcriptional circuitries, at least at our current detection level. Taken together, our data argue in favor of regulated and consistent transcriptional priming within early myeloid cells, suggesting that, in vivo, primed cells will complete

Figure 5. Myeloid Subtypes Are Epigenetically Distinct

(A) Representative examples of regions marked by H3K4me2 in CD41$^+$ (C8 enriched [Figure 3B]), compared to the control CMP population (see also Figure S6). Displayed are normalized H3K4me2 intensities in 100 kb genomic regions. Putative differential enhancers are shadowed in gray. Lineage-specific genes (black) and other genes in the loci (light gray) are indicated with arrows below.

(B and C) Scatterplot of read density in H3K4me2 peaks from the CD41$^+$ population. Compared to biological replicate (B) or the remaining CMP populations (C)

(D–F) Same as (A)–(C) but for Flt3$^+$ Csf1r$^+$ (C10-enriched [Figure 3B])

(G–I) Same as (A)–(C) but for Irf8$^+$ MHC-II$^+$ (C11-enriched [Figure 3B])

(J) Heatmap showing K-means clustering of H3K4me2 read density in putative enhancers across the myeloid branches of the hematopoietic system as described in Lara-Astiaso et al. (2014). Hierarchical tree of myeloid populations is shown above. Left bar graphs (capped at 70%) show overlap of these enhancers with H3K4me2 signal in CD41$^+$ (yellow), Flt3$^+$ Csf1r$^+$ (green), and Irf8$^+$ MHC-II$^+$ (gray). Significant overlaps (Hypergeometric p value $< 10^{-10}$) are marked by asterisks. Hierarchical tree is shown above.

(K) Normalized H3K4me2 read density in 100 bp bins around putative enhancer groups. 0 marks the center of the H3K4me2 peaks.

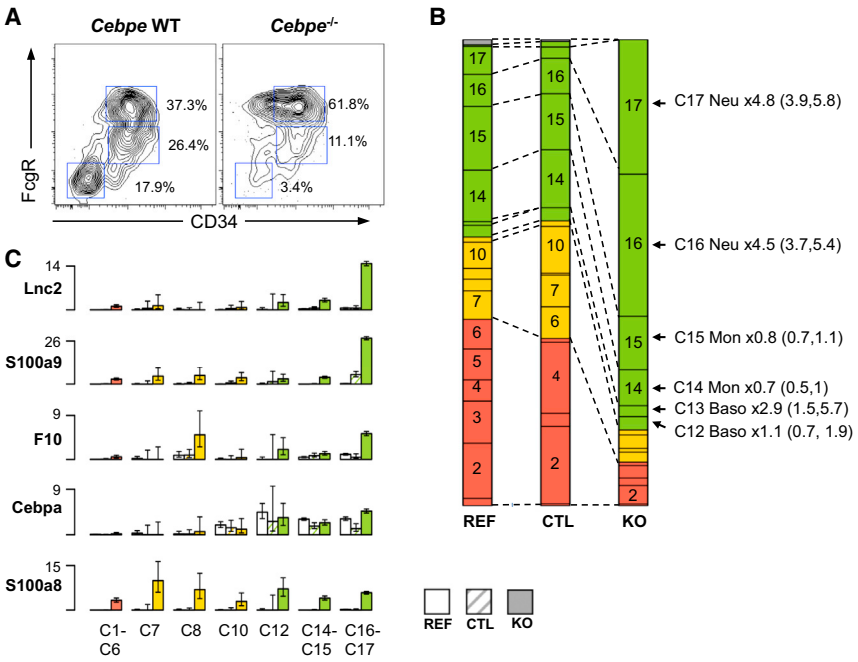


Figure 7. *Cebpe* Deficiency Is Manifested in Granulocyte Differentiation Arrest

(A) FACS sorting of myeloid progenitors from *Cebpe* knockout mice as compared to *Cebpe*$^{+/+}$ littermate controls.

(B) Changes in frequency of single cells mapping to clusters C1–C19 upon knockout of *Cebpe* (KO) as in Figure 6B.

(C) Mean expression of selected genes across all clusters in *Cebpe* KO mice as in Figure 6D. Error bars are calculated as in Figure 6D.

with high probability their differentiation as instructed by their initial gene expression patterns.

Our functional competitive assay showed that CD41^{+} Mk progenitors (C8 cells) and Flt3^{+}Csfr1^{+} monocyte/DC progenitors (C10 cells) have a restricted potential to reconstitute erythroid/megakaryocytic and myeloid lineages, respectively. This commitment was evident despite the fact that standard bone marrow transplantation experiments involve considerable perturbation of the native niche structure. Such perturbations may select for less-committed cells and/or favor reprogramming of partially committed progenitors. On the other hand, the fact that CD41^{+} progenitors give rise to a large proportion of red blood cells, despite their clear transcriptional association with the megakaryocyte and not the erythrocyte state, suggests that transcriptional diversification within the compartment does not completely restrict the potential of progenitors to switch fate once the environment is perturbed. Interestingly, chromatin analysis of CD41^{+} progenitors indeed suggests that their epigenetic landscape is still compatible, at least partially, with the profile of early hematopoietic stem cells. Furthermore, although we do not detect any trace of Mk-like cells within the classical MEP compartment, some of the major transcriptional regulators (*Gata2*, *Zfpm1*) and some potential markers (*ApoE)* are shared between early erythrocyte (C7) and Mk (C8) progenitors, but not with granulocyte and monocyte lineages. These observations may explain why Mk and erythrocyte progenitors retain some degree of plasticity in functional experiments. Given these considerations, we hypothesize that single-cell transcriptional analysis does provide a blueprint of the in vivo developmental trajectories of progenitors in the bone marrow and that transplantation-based functional assays outline a more plastic fate map due to the heterogeneity of sorted populations and the perturbed physiology of these experiments in the donor progenitors and host bone marrow niches. Additional analysis of chromatin and genetic perturbations, as shown here, and studies of DNA methylation (Jeong and Goodell, 2014), chromatin conformation (Lara-Astiaso et al., 2014), and niche structure (Boulais and Frenette, 2015) are all required to describe the interplay between gene expression and developmental commitment in early hematopoiesis in mechanistic terms.

In conclusion, we describe here the transcriptional sorting of myeloid progenitors at single-cell resolution with minimal biases in a screen with sensitivity that is restricted only by the depth of our single-cell RNA-seq sampling and its sub-sorting follow-up studies. The resulting model is based on multiple TFs that combinatorially drive differentiation within the myeloid compartments and toward multiple functional cell types. We accumulate significant evidences that support the linkage between transcriptional and functional commitment and demonstrate the integration of MARS-seq with additional strategies (functional assays, chromatin studies, and genetic perturbation) for characterizing these populations in greater detail. Future work eliminating any marker-associated biases (e.g., Sca1 and c-Kit markers) may outline the complete spectrum of blood development toward all lineages and the regulatory circuits driving these processes. Together, the new model and the underlying methodology are immediately applicable in normal and aberrant human hematopoiesis (Cantor and Orkin, 2002; Rosenbauer and Tenen, 2007) with substantial potential to define new markers, targets, and pathways shared across various hematological diseases.

EXPERIMENTAL PROCEDURES

Mouse Strains

Gene expression and chromatin analysis was performed on 6- to 8-week-old female C57BL/6 mice. For functional reconstitution assays, donor cells were isolated from C57BL/6 CD45.2 Ub-GFP mice and were transplanted into CD45.1 congenic mice. The conditional Cebpa$^{flox/flox}$Mx1-Cre was induced by three injections of 300 μg poly(I:C) into 11-week-old mice over the course of 4 days.

Isolation of Hematopoietic Progenitors from Bone Marrow

Bone marrow was isolated from mouse tibiae femora and ilia leg bones, filtered through a 70 μm cell strainer, and the cell suspension enriched for c-Kit expressing cells using magnetic cell separation (Miltenyi Biotec

Germany; #130-091-224) according to manufacturer's instructions. Cells were then stained and FACS sorted through a FACSAria Fusion cell sorter (BD Biosciences). Myeloid progenitors, including CMP, GMP, and MEP, were defined as Lin^- (lineage negative) c-Kit^+ $Sca1^-$ and gated by the levels of the FcgR and CD34 markers (Figure S1B). Lineage markers included: anti-mouse Ter-119, Gr-1, CD11b, B220, CD19, CD3, CD4, and CD8. Antibody clones used in this study are listed in Table S6.

Single-Cell Index Sorting

Isolated cells were single-cell sorted into 384-well cell capture plates containing 2 μl of lysis solution and barcoded poly(T) reverse-transcription (RT) primers for single-cell RNA-seq (Jaitin et al., 2014). Barcoded single-cell capture plates were prepared with a Bravo automated liquid handling platform (Agilent), as described previously (Jaitin et al., 2014). To record marker levels of each single cell, the FACS Diva 7 "index sorting" function was activated during single-cell sorting. While index sorting, single cells were sorted from the entire Lin^- c-Kit^+ Sca-1^+ space, and the intensities of the FcgR and CD34 FACS markers were recorded and linked to each cell's position within the 384-well plate (Figure 1A). Sorting efficiency was >95% of the wells analyzed.

Massively Parallel Single-Cell RNA-Seq Library Preparation

Single-cell libraries were prepared as previously described (Jaitin et al., 2014). In brief, mRNA from cells sorted into massively parallel single-cell RNA-seq (MARS) capture plates are barcoded and converted into cDNA and pooled using an automated pipeline (Jaitin et al., 2014). The pooled sample is then linearly amplified by T7 in vitro transcription, and the resulting RNA is fragmented and converted into a sequencing-ready library by tagging the samples with pool barcodes and Illumina sequences during ligation, RT, and PCR. Each pool of cells was tested for library quality and concentration is assessed as described earlier (Jaitin et al., 2014).

Sequencing and Low-Level Processing

All RNA-seq and ChIP-seq libraries (pooled at equimolar concentration) were sequenced using Illumina sequencers, either a HiSeq-1500 or a NextSeq 500, at an average sequencing depth of 78,682 reads per cell for RNA-seq and 6,446,930 reads per sample for ChIP-seq. We used statistics on empty-well spurious unique molecular identifier (UMI) detection to ensure that the batches we used for analysis show a low level of cross-single-cell contamination. Details on low-level processing and quality control are available in the Supplemental Experimental Procedures.

A Batch-Aware, Multinomial Mixture Model for Single-Cell RNA-Seq Data

We enhance our previously described mixture model for single-cell RNA-seq data (Jaitin et al., 2014) to accommodate larger datasets with multiple batches and possible associated biases. In brief, we implemented a multiplicative probabilistic model that represents the probability of sampling UMIs for each gene as a function of the cell's cluster and batch. We then inferred cluster and batch parameters using an expectation maximization (EM)-like procedure using non-linear optimization of the batch and cluster parameters' likelihoods in iterations. See Supplemental Experimental Procedures for complete description of the algorithm.

Clustering and Computational Post-processing

We refined our EM-based clustering approach by adding pre-filtering of gene modules associated with cell cycle or stress and post-processing of clustering results to identify remaining outlier genes. Complete details of our computational procedures can be found in the Supplemental Experimental Procedures.

In Vivo Differentiation Assay

Six hours prior to bone marrow transplantation, CD45.1 recipient mice were lethally irradiated with two subsequent X-ray doses of 550 cGy and 500 cGy that were administered 3 hr apart. After irradiation, drinking water was supplemented with 200 mg ciprofloxacin/ml. GFP^+ $CD41^+$ cells were isolated from CD45.2 Ub-GFP donor mice as described above. Between 10,000 and 25,000 donor cells were injected intravenously together with 50,000 CD45.1 flushed whole bone marrow carrier cells. After 10 days, blood and spleen was collected and analyzed by FACS.

ACCESSION NUMBERS

The accession numbers for the RNA-seq and ChIP-seq datasets reported in this paper are GEO: GSE72857, GSE72858, and GSE72859.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.11.013.

AUTHOR CONTRIBUTIONS

F.P., Y.A., A.G., D.A.J., A. Tanay, and I.A. conceived the project and designed the experiments; F.P., D.A.J., H.K.-S., D.L.-A., M.G., and A.M. performed the experiments; Y.A., A.G., E.K., D.W., A.W., E.D., and N.C. analyzed the data. S.H., A.S., F.G., and A. Trumpp contributed to the functional assay; F.K.B.L. and B.T.P. implemented the knockout experiments; F.P., Y.A., A.G., D.A.J., E.K., D.W., A. Tanay, and I.A. wrote the paper; A. Tanay and I.A. supervised the project.

ACKNOWLEDGMENTS

We thank members of the Tanay and Amit labs for critical discussions and Genia Brodsky for artwork. Research in I.A. and A. Tanay laboratories is supported by the European Research Council, the I-CORE for chromatin and RNA regulation, and personal grants from the Israel Science foundation (782/11, 1050/12) and the BLUEPRINT FP7 consortium. I.A. is supported by the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, a Minerva Stiftung research grant, and the National Human Genome Research Institute Center for Excellence in Genome Science (1P50HG006193). A. Tanay is a Kimmel investigator. F.P. is a fellow of the German-Israeli Helmholtz Research School in Cancer Biology. Work in the B.P. laboratory was supported through a center grant from the NovoNordisk Foundation (The Novo Nordisk Foundation section for Stem Cell Biology in Human Disease).

Received: July 20, 2015
Revised: September 30, 2015
Accepted: November 5, 2015
Published: November 25, 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.

Becker, A.J., McCULLOCH, E.A., and Till, J.E. (1963). Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature *197*, 452–454.

Becker, A.M., Michael, D.G., Satpathy, A.T., Sciammas, R., Singh, H., and Bhattacharya, D. (2012). IRF-8 extinguishes neutrophil production and promotes dendritic cell lineage commitment in both myeloid and lymphoid mouse progenitors. Blood *119*, 2003–2012.

Bendall, S.C., Davis, K.L., Amir, A.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell *157*, 714–725.

Boulais, P.E., and Frenette, P.S. (2015). Making sense of hematopoietic stem cell niches. Blood *125*, 2621–2629.

Cantor, A.B., and Orkin, S.H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. Oncogene *21*, 3368–3376.

Dahl, R., Walsh, J.C., Lancki, D., Laslo, P., Iyer, S.R., Singh, H., and Simon, M.C. (2003). Regulation of macrophage and neutrophil cell fates by the PU.1:C/EBPalpha ratio and granulocyte colony-stimulating factor. Nat. Immunol. *4*, 1029–1036.

Etzrodt, M., Endele, M., and Schroeder, T. (2014). Quantitative single-cell approaches to stem cell research. Cell Stem Cell *15*, 546–558.

Görgens, A., Radtke, S., Möllmann, M., Cross, M., Dürig, J., Horn, P.A., and Giebel, B. (2013). Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. Cell Rep. *3*, 1539–1552.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Guo, G., Luc, S., Marco, E., Lin, T.W., Peng, C., Kerenyi, M.A., Beyaz, S., Kim, W., Xu, J., Das, P.P., et al. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. Cell Stem Cell *13*, 492–505.

Iwasaki, H., Mizuno, S., Wells, R.A., Cantor, A.B., Watanabe, S., and Akashi, K. (2003). GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. Immunity *19*, 451–462.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Jeong, M., and Goodell, M.A. (2014). New answers to old questions from genome-wide maps of DNA methylation in hematopoietic cells. Exp. Hematol. *42*, 609–617.

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

Kondo, M., Scherer, D.C., Miyamoto, T., King, A.G., Akashi, K., Sugamura, K., and Weissman, I.L. (2000). Cell-fate conversion of lymphoid-committed progenitors by instructive actions of cytokines. Nature *407*, 383–386.

Kurotaki, D., Yamamoto, M., Nishiyama, A., Uno, K., Ban, T., Ichino, M., Sasaki, H., Matsunaga, S., Yoshinari, M., Ryo, A., et al. (2014). IRF8 inhibits C/EBPα activity to restrain mononuclear phagocyte progenitors from differentiating into neutrophils. Nat. Commun. *5*, 4978.

Laiosa, C.V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L., and Graf, T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity *25*, 731–744.

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. Science *345*, 943–949.

Lee, Y.H., Sauer, B., Johnson, P.F., and Gonzalez, F.J. (1997). Disruption of the c/ebp alpha gene in adult mouse liver. Mol. Cell. Biol. *17*, 6014–6022.

Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184–197.

Liu, Q., and Dong, F. (2012). Gfi-1 inhibits the expression of eosinophil major basic protein (MBP) during G-CSF-induced neutrophilic differentiation. Int. J. Hematol. *95*, 640–647.

Miyamoto, T., Iwasaki, H., Reizis, B., Ye, M., Graf, T., Weissman, I.L., and Akashi, K. (2002). Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. Dev. Cell *3*, 137–147.

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.

Moignard, V., Macaulay, I.C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F.J., Kinston, S., Joshi, A., Hannah, R., Theis, F.J., et al. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. Nat. Cell Biol. *15*, 363–372.

Murre, C. (2007). Defining the pathways of early adult hematopoiesis. Cell Stem Cell *1*, 357–358.

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

Orkin, S.H., Shivdasani, R.A., Fujiwara, Y., and McDevitt, M.A. (1998). Transcription factor GATA-1 in megakaryocyte development. Stem Cells *16* (*Suppl 2*), 79–83.

Paul, F., and Amit, I. (2014). Plasticity in the transcriptional and epigenetic circuits regulating dendritic cell lineage specification and function. Curr. Opin. Immunol. *30*, 1–8.

Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature *349*, 257–260.

Pronk, C.J., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell *1*, 428–442.

Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Riddell, J., Gazit, R., Garrison, B.S., Guo, G., Saadatpour, A., Mandal, P.K., Ebina, W., Volchkov, P., Yuan, G.-C.C., Orkin, S.H., and Rossi, D.J. (2014). Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. Cell *157*, 549–564.

Rosenbauer, F., and Tenen, D.G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. Nat. Rev. Immunol. *7*, 105–117.

Schroeder, T. (2010). Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior. Cell Stem Cell *6*, 203–207.

Seita, J., and Weissman, I.L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip. Rev. Syst. Biol. Med. *2*, 640–653.

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.

Tavor, S., Vuong, P.T., Park, D.J., Gombart, A.F., and Cohen, A.H. (2002). Macrophage functional maturation and cytokine production are impaired in C/EBPε-deficient mice. Blood *99*, 1794–1801.

Verbeek, W., Wächter, M., Lekstrom-Himes, J., and Koeffler, H.P. (2001). C/EBPepsilon−/−mice: increased rate of myeloid proliferation and apoptosis. Leukemia *15*, 103–111.

Wang, H., Yan, M., Sun, J., Jain, S., Yoshimi, R., Abolfath, S.M., Ozato, K., Coleman, W.G., Jr., Ng, A.P., Metcalf, D., et al. (2014). A reporter mouse reveals lineage-specific and heterogeneous expression of IRF8 during lymphoid and myeloid cell differentiation. J. Immunol. *193*, 1766–1777.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E., et al. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell *135*, 1118–1129.

Wilson, A., Laurenti, E., and Trumpp, A. (2009). Balancing dormant and self-renewing hematopoietic stem cells. Curr. Opin. Genet. Dev. *19*, 461–468.

Winkler, I.G., Hendy, J., Coughlin, P., Horvath, A., and Lévesque, J.-P.P. (2005). Serine protease inhibitors serpina1 and serpina3 are down-regulated in bone marrow during hematopoietic progenitor mobilization. J. Exp. Med. *201*, 1077–1088.

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.

Yamanaka, R., Barlow, C., Lekstrom-Himes, J., Castilla, L.H., Liu, P.P., Eckhaus, M., Decker, T., Wynshaw-Boris, A., and Xanthopoulos, K.G. (1997). Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. Proc. Natl. Acad. Sci. USA *94*, 13187–13192.

Ye, M., Zhang, H., Amabile, G., Yang, H., Staber, P.B., Zhang, P., Levantini, E., Alberich-Jordà, M., Zhang, J., Kawasaki, A., and Tenen, D.G. (2013). C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. Nat. Cell Biol. *15*, 385–394.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc. Natl. Acad. Sci. USA *94*, 569–574.