Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing

Roy Drissen1,2, Natalija Buza-Vidas2, Petter Woll1,3, Supat Thongjuea1, Adriana Gambardella1,2, Alice Giustacchini1,3, Elena Mancini4, Alya Zriwil5, Michael Lutteropp1,3, Amit Grover1,2,4, Adam Mead1,3, Ewa Sitnicka5, Sten Eirik W Jacobsen1,3,6 & Claus Nerlov1,2,4,6

According to current models of hematopoiesis, lymphoid-primed multi-potent progenitors (LMPPs) (Lin−Sca-1+c-Kit+CD34+Flt3(hi)) and common myeloid progenitors (CMPs) (Lin−Sca-1+c-Kit+CD34+CD41(hi)) establish an early branch point for separate lineage-commitment pathways from hematopoietic stem cells, with the notable exception that both pathways are proposed to generate all myeloid innate immune cell types through the same myeloid-restricted pre–granulocyte-macrophage progenitor (pre-GM) (Lin−Sca-1−c-Kit+CD41−FcγRII/III−CD150−CD105+). By single-cell transcriptome profiling of pre-GMs, we identified distinct myeloid differentiation pathways: a pathway expressing the gene encoding the transcription factor GATA-1 generated mast cells, eosinophils, megakaryocytes and erythroid cells, and a pathway lacking expression of that gene generated monocytes, neutrophils and lymphocytes. These results identify an early hematopoietic-lineage bifurcation that separates the myeloid lineages before their segregation from other hematopoietic-lineage potential.

Each cell type of the myeloid innate immune system (neutrophils, macrophages, eosinophils, mast cells and basophils) makes a unique contribution to immunity to bacteria and parasites1. All of these cell types are generated from multi-potent hematopoietic stem cells (HSCs) in a process controlled by a complex network of transcription factors2 and extrinsic signals3. Knowledge of the cellular pathways by which hematopoietic cell types are generated and how their production is regulated by infectious organisms and cytokines is critical to understanding how the immune system functions and how its dysfunction results in deficiency in or overproduction of immune cells under pathological conditions4,5. Notably, current models for hematopoiesis commonly indicate that all myeloid cell types develop from the same granulocyte-macrophage progenitor populations (pre–granulocyte-macrophage progenitor (pre-GM) and granulocyte-macrophage progenitor (GMP); defined as Lin−Sca-1−c-Kit+CD41−FcγRII/III+))2,6–9, which according to the CMP-LMPP hematopoietic model are produced by both the CMP branch (which is able to generate monocyte cells, megakaryocytes and erythroid cells but lacks lymphoid-lineage potential10,11) and LMPP branch (which can generate myeloid and lymphoid cell types but lacks megakaryocyte-erythroid (Mk-E)–lineage potential12). Published studies have shown that GMPs have combined monocyte-macrophage and neutrophil potential at the clonal level6,10. However, granulocytes also include functionally distinct mast cells, basophils and eosinophils, each specified through the action of critical transcription factors13, and the presence of such potential in GMPs has so far not been analyzed systematically. Therefore, whether the LMPP and CMP pathways generate the same complement of myeloid cell types has remained an open question.

Here we used global gene profiling of single pre-GMs to identify molecularly distinct pre-GM subsets defined by differential expression of Gata1 (which encodes the transcription factor GATA-1) and Flt3 (which encodes the receptor tyrosine kinase Flt3). Gata1 was of particular interest, as in addition to its critical role in megakaryocyte and erythrocyte development14,15, GATA-1 is expressed in eosinophils, basophils and mast cells (but not monocytes-macrophages or neutrophils) and is important for their differentiation16–18. By generating reporter mice expressing enhanced green fluorescent protein (EGFP) from Gata1 (Gata1-EGFP), we were able to fractionate myeloid progenitors on the basis of their Gata1 expression and identify sub-populations with distinct myeloid lineage potential. Gata1-EGFP+ (GE+) pre-GMs and GE+ GMPs gave rise to mast cells and eosinophils but lacked monocyte-macrophage potential and had little or no neutrophil potential. Conversely, Gata1-EGFP− (GE−) pre-GMs and GE− GMPs (as well as LMPPs) generated monocytes and neutrophils but no mast cells and few eosinophils. Notably, monocyte-neutrophil potential and mast cell-eosinophil potential segregated before separation from other blood lineages, as GE− pre-GMs generated megakaryocytes and erythroid cells, whereas LMPPs contained robust B lineage and T lineage potential. Our findings therefore identify an early blood-lineage-fate ‘decision’ that segregates lymphoid-macrophage-neutrophil potential from mast cell–eosinophil–megakaryocytic–erythroid potential and thereby establishes separate
myeloid-erythroid and myelo-lymphoid differentiation pathway and generates two distinct myeloid-restricted progenitor subsets, instead of a common GMP harboring all monocyte-macrophage potential and granulocyte potential.

RESULTS
Identification of pre-GM molecular heterogeneity
To address whether the same myeloid differentiation pathways emerge from CMPs and LMPPs, we performed global single-cell gene profiling of bone marrow pre-GMs\(^6\), the earliest progenitors predicted to be shared by the two pathways\(^6\)–\(^8\). RNA sequencing of 63 single pre-GMs was followed by unsupervised clustering with the 100 genes with greatest variance across the entire cell population. This led to the identification of two distinct pre-GM clusters (Supplementary Fig. 1a) and 55 genes showing differential expression by the clusters (P < 0.05; false-discovery rate < 0.05; Supplementary Fig. 1b), as potential sub-population classifiers. Clustering with this gene set identified three distinct pre-GM subpopulations defined by expression of Gata1 but not Flt3, expression of neither Gata1 nor Flt3, and expression of Flt3 but not Gata1 (Fig. 1a), a pattern that was validated by targeted single-cell gene-expression analysis (Fig. 1b). This analysis also identified Gata1 as an optimal classifier that was homogeneously and selectively expressed in a distinct subpopulation of pre-GMs that did not express Flt3 but was not detected in pre-GMs that did express Flt3 (Fig. 1b). Notably, Gata1 expression and Flt3 expression have been used to define CMPs\(^11\) and LMPPs\(^12\), respectively, within the CD34\(^+\) Lin\(^−\) Sca-1\(^+\) c-Kit\(^+\) (LSK) population, which would suggest that expression of Gata1 and Flt3 might have the potential to identify pre-GM subsets derived from these distinct upstream progenitors.

Gata1 expression defines distinct myeloid progenitors
The Gata1-EGFP transgene used to define CMPs\(^11\) lacks key regulatory elements\(^19\)–\(^21\) and is not expressed in GMPs\(^22\). We therefore generated a Gata1-EGFP reporter mouse line using a bacterial artificial chromosome transgenic construct containing all known Gata1 regulatory sequences, in which an expression cassette encoding EGFP replaced the coding part of the second exon of Gata1 (Supplementary Fig. 1c). In these Gata1-EGFP reporter mice, distinct GE\(^+\) and GE\(^−\) subsets were present among pre-GMs and GMPs (Fig. 1c and Supplementary Fig. 1d). Within the bone marrow multi-potent LSK compartment, HSCs are CD150\(^+\) LSK cells\(^23\), whereas LMPPs are Flt3\(^hi\) LSK cells\(^12\). The CD150\(^+\) LSK cell population contained a substantial fraction of GE\(^+\) cells (Fig. 1d and Supplementary Fig. 1e), but only GE\(^−\) CD150\(^−\) LSK cells had HSC function in vivo (Supplementary Fig. 2a–d). We therefore defined HSCs as CD150\(^+\) GE\(^−\) LSK cells here. Similarly, a small fraction (2–3\%) of LMPPs had low Gata1-EGFP expression (Fig. 1d). Published studies have shown that approximately 2\% of LMPPs possess detectable megakaryocyte potential\(^12\), and we found that this megakaryocyte potential was confined to the small GE\(^−\) sub-fraction of Flt3\(^hi\) LSK cells (Supplementary Fig. 2e), which supported the proposal of the existence of an LMPP population devoid of platelet-forming capacity\(^12\)\(^,\)\(^24\)\(^,\)\(^25\) and led us to stringently define and sort LMPPs as Flt3\(^hi\) GE\(^−\) LSK cells in this study. Notably, Gata1-EGFP expression mirrored endogenous Gata1 mRNA expression both at the population level (Fig. 1e) and in analysis of single pre-GMs (Supplementary Fig. 2f). The Gata1-EGFP reporter, therefore, identified transcriptional heterogeneity within the phenotypic HSC, LMPP, pre-GM and GMP populations.

Early separation of macrophage and mast cell potential
Quantitative PCR analysis of lineage-specific gene expression showed that genes affiliated with Mk-E cells (Hga2b, KIf1, Epor and Gata2) were expressed ‘preferentially’ in GE\(^+\) pre-GMs and GMPs, the Mk-E progenitor (pre-Meg-E) population (Lin\(^−\)Sca-1\(^−\)c-Kit\(^−\)CD41\(^+\)–FcyRII/III\(^−\)/CD150\(^−\)/CD105\(^−\)), the megakaryocyte progenitor (MkP) population (Lin\(^−\)Sca-1\(^−\)c-Kit\(^−\)CD150\(^−\)CD41\(^+\)) and erythroid colony-forming unit progenitor (pre-CFU-E) population (Lin\(^−\)Sca-1\(^−\)c-Kit\(^−\)CD41\(^−\)–FcyRII/III\(^−\)/CD150\(^−\)/CD105\(^−\)) (Fig. 1f), as were mast cell–related genes (Cma1 and Mcpt1; Fig. 1g), whereas macrophage-related genes (Cd68 and Emr1; Fig. 1h), neutrophil-related genes (Csf3r and Ctsg; Fig. 1i) and lymphoid cell–related genes (Rag1 and Cmah; Fig. 1j) were expressed selectively in LMPPs, GE\(^−\) pre-GMs and GE\(^−\) GMPs. The divergent expression of macrophage-neutrophil–related genes versus mast cell–related genes in GE\(^−\) pre-GMs and GMPs versus GE\(^−\) pre-GMs and GMPs suggested that these progenitors might generate different myeloid cell types. We therefore cultured GE\(^+\) pre-GMs and GE\(^−\) pre-GMs for 8 d under ‘pan-myeloid’ conditions (with the cytokines interleukin 3 (IL-3), IL-5, IL-9, stem-cell factor (SCF) and granulocyte-macrophage colony-stimulating factor (GM-CSF)) and assessed the resulting cultures by cytospin morphology and gene-expression analysis. Morphological analysis showed that LMPPs and GE\(^−\) pre-GMs generated monocytes but no mast cells, whereas GE\(^+\) pre-GMs generated mast cells but no monocytes (Fig. 1k). Polymorphonuclear (PMN) cells were present in both cultures, with neutrophil morphology (small cell size and a highly condensed and segmented nucleus) evident in those derived from GE\(^−\) pre-GMs and LMPPs, whereas those derived from GE\(^+\) pre-GMs were generally larger with less-condensed nuclei (Fig. 1k). Consistent with the observed separation of mast cell potential and monocyte-macrophage potential, mast cell–specific genes were expressed exclusively in the cultures derived from GE\(^−\) pre-GMs (Fig. 2a), and monocyte-macrophage–specific genes were expressed selectively in those derived from LMPPs and GE\(^−\) pre-GMs (Fig. 2b). Finally, while the genes encoding the myeloid transcription factors PU.1 and C/EBP\(\alpha\) were expressed in all cultures, expression of Gata1 and Gata2 was confined to cultures derived from GE\(^−\) pre-GMs (Fig. 2c,d).

To determine the frequency and distribution of granulocyte and monocyte-macrophage potential within the different progenitor populations at the single-cell level, we individually cultured sorted progenitor cells and analyzed their myeloid lineage output at several time points. While single LMPPs, GE\(^−\) pre-GMs and GE\(^−\) GMPs produced large numbers of monocytes, mast cell potential was never detectable in cultures of single LMPPs or GE\(^−\) GMPs and was very rare (<2\% of cultures) in GE\(^−\) pre-GM single cell–derived cultures, at all time points investigated (Fig. 2e). In contrast, monocytes were generated only very rarely (<2\% of single cell cultures) from GE\(^−\) pre-GMs or GE\(^+\) GMPs, whereas mast cell potential was very abundant in GE\(^−\) pre-GMs or GE\(^+\) GMPs (Fig. 2e). Combined monocyte and mast cell morphology was exceptionally rare, found in only 2 of over 1,000 single cell–derived clones of all progenitors analyzed (Fig. 2e). Notably, this was not due to any inability of the culture system to support the development of these two cell types simultaneously, as culture of multi-potent HSCs or co-culture of GE\(^−\) pre-GMs and GE\(^−\) pre-GMs generated combined mast cells and monocytes with high frequency (Supplementary Fig. 2g,h). Both monocytes and mast cells were observed in conjunction with other granulocytes with high frequency in single cell–derived clones from GE\(^−\) pre-GMs (22–33\%), GE\(^−\) pre-GMs (~50\%) and LMPPs (65\%) and also in single cell–derived...
Figure 1 Gata1 expression identifies distinct myeloid progenitor subsets. (a) Hierarchical clustering of single pre-GM cells (with the gene set (right margin) from Supplementary Fig. 1b) in the main molecularly distinct cell clusters distinguished according to the presence (+) or absence (−) of Gata1 expression and Flt3 expression (below plot). FPKM, fragments per kilobase of transcript per million mapped reads. (b) Expression of genes (right margin) selectively expressed in single pre-GMs (grouped according to descending expression of Gata1 and Flt3 expression (below plot)) in the three clusters in a; results were normalized to those of the control gene Hprt and are presented as deviation from mean expression of each gene. (c) EGFP expression in pre-GM, MkP, GMP, pre-CFU-E, pre-Meg-E and CFU-E populations in bone marrow from adult mice heterozygous for the Gata1-EGFP transgene, with gates set on the basis of wild-type cells (gating strategy, Supplementary Fig. 1d). Numbers above bracketed lines indicate percent EGFP+ cells. (d) EGFP expression in LSK CD150+Flt3hi and LSK Flt3hi cells (gating strategy, Supplementary Fig. 1e). Numbers above bracketed lines as in c. (e–j) Quantitative PCR analysis of the expression of mRNA from Gata1 (e), genes associated with megakaryocytes-erythrocytes (f), mast cells (g), monocytes-macrophages (h), neutrophils (i) and lymphoid cells (j) in purified stem and progenitor populations (gated as in key); results are presented relative to those of Hprt. (k) Microscopy of cells in cultures of LMFPs, GE− pre-GMs and GE+ pre-GMs, showing the morphology of monocytes (Mo), PMN granulocytes (PMN) and mast cells (Ma) at day 8 of culture. Scale bars, 25 μm. Data are representative of one (a), two (b), twenty (c), five (d) or ten (k) independent experiments or are from one experiment (e–j; mean and s.e.m. of n = 2 (GE− pre-GM) or n = 3 (all others) biological replicates).
clones from GE+ GMPs and GE− GMPs at a lower frequency (Fig. 2c). The PMN cells associated with monocytes showed neutrophil morphology, whereas those associated with mast cells seemed larger, with less-condensed nuclei (Fig. 2f). These results demonstrated the distinct lineage potential of GE+ myeloid progenitors versus that of GE− myeloid progenitors.

**Eosinophil potential of GE+ pre-GMs and GE+ GMPs**

While the data presented above showed that mast cell potential and monocyte-macrophage potential were separated before the formation of pre-GMs and GMPs, the nature of the additional granulocyte-lineage potential associated with these progenitors remained unclear. To address this issue, we performed microarray-based global gene profiling of pre-GMs, GMPs, HSCs, LMPPs, CLPs (Lin−B220−Sca-1loc-KitloFlt3+IL-7R+), and pre-Meg-E cells. Consistent with their similar granulocyte-monocyte lineage ‘readouts’, LMPPs, GE− pre-GMs and GE+ GMPs clustered together by principal-component analysis, in close association with CLPs, whereas GE− pre-GMs and GE+ GMPs formed a separate cluster, associated with pre-Meg-E cells (Fig. 3a).

Direct comparison of the transcriptomes of GE+ GMPs and GE− GMPs showed that GE+ GMPs had higher expression of mast cell– and eosinophil-specific genes, whereas GE− GMPs had higher expression of numerous monocyte-macrophage– and neutrophil-associated genes (Supplementary Table 1). We therefore investigated the possibility that the PMN cells derived from GE− pre-GMs and GE+ GMPs were eosinophils. First, we cultured GE+ GMPs and GE− GMPs in the presence of SCF, GM-CSF and IL-5 (Supplementary Fig. 3a) and assessed the morphology and gene expression of the resulting cultures. Under these conditions, GE+ GMPs did not generate mast cells but generated only large PMN cells with uncondensed nuclear morphology, while GE− GMPs generated cells with monocyte and neutrophil morphology (Fig. 3b). Gene-expression analysis indicated selective expression of eosinophil-specific genes (Il13ra, Ccr3, Siglec5 and Prg2) in GE+ GMP cultures (Fig. 3c) and of monocyte-macrophage–specific genes (Cd68, Lpl, Lrp1 and Mpeg1) in GE− GMP cultures (Fig. 3d). In both cases, the cell-type specificity of gene expression was confirmed by analysis of macrophages and eosinophils sorted from the peritoneal cavity as controls. Cytospins indicated that the morphology of the eosinophil-like cells derived in vitro from GE+ GMPs was similar to that of purified peritoneal eosinophils (Supplementary Fig. 3b). These results indicated the selective development of eosinophils from GE+ GMPs.

We also performed flow cytometry of the cells differentiated from GE+ pre-GMs and GE− pre-GMs, as well as from GE+ GMPs and GE− GMPs, following culture under ‘pan-myeloid cytokine’ conditions. Cells with a FcεRIα+ c-Kit+GE+ mast cell surface phenotype were abundant in GE+ pre-GM and GE+ GMP cultures, whereas Mac-1+Ly6G−GE− monocytes and Mac-1+Ly6G−GE− neutrophils were the principal cell types generated from GE− pre-GMs and GE− GMPs (Fig. 4a,b); this indicated that neutrophil potential was selectively associated with GE− pre-GMs and GE− GMPs. Eosinophil, neutrophil and mast cell identities were further confirmed by analysis of the expression of lineage-specific genes (Supplementary Fig. 3c–f). FcεRIα+ c-Kit−SiglecF+GE+ eosinophils were inefficiently generated under the ‘pan-myeloid’ conditions (Fig. 4a,b), probably because of overgrowth by mast cells. To circumvent this, we cultured...
GMPs in the presence of SCF, GM-CSF and IL-5 and observed selective development of eosinophil-like cells in GE+ GMP cultures, while GE- GMPs gave rise to monocytes and neutrophils but few or no eosinophils, even under conditions favoring eosinophil development (Fig. 4c). Thus, on the basis of analyses of morphology, surface markers and gene expression, we concluded that GE+ GMPs gave rise to eosinophils and mast cells but did not produce detectable neutrophils or monocytes. Finally, to accurately ‘co-localize’ lineage potential in single progenitor cells, we combined morphological analysis of cytospins of single cell–derived cultures with immunocytochemistry directed against EGFP to distinguish GE+ eosinophils from GE- neutrophils (Fig. 4d). GE+ pre-GM and GE+ GMP populations included a large proportion of bi-potent mast cell–eosinophil progenitors, whereas GE- LMPP, GE- pre-GM and GE- GMP populations included an abundance of bi-potent neutrophil–monocyte progenitors (Fig. 4e). In addition, under ‘pan-myeloid’ conditions, we observed the emergence of FcγRII/III+ GE+ cells from GE- pre-GMs and FcγRII/III+ GE- cells from GE- pre-GMs and LMPPs that in both cases displayed the clonogenic lineage potential of the corresponding GE+ or GE- GMP population (Supplementary Fig. 4a,b), in further support of the proposal of the existence of independent and functionally distinct GE+ and GE- progenitor hierarchies. Comparison of the gene-expression profiles of GE+ GMPs and GE- GMPs identified genes encoding cell-surface markers that were ‘preferentially’ expressed by the GE+ subset (such as Csd5 and Il1rl1) or the GE- subset (such as Ly6c) (Supplementary Table 1). While cell-surface expression of the cytokine receptor ST2 (encoded by Il1rl1) was not detectable on pre-GMs or GMPs with the antibodies available, the convertase inhibitor CD55 (encoded by Csd5) and surface marker Ly6C (encoded by Ly6c) were expressed selectively on GE+ pre-GMs and GMPs and on GE- pre-GMs and GMPs, respectively (Supplementary Fig. 4c). In agreement with such selective expression, the expression of these markers allowed the separation of progenitors with a restricted monocyte–neutrophil potential (Ly6C+ pre-GMs and GMPs) or restricted mast cell–eosinophil potential (CD55+ pre-GMs and GMPs), without the use of the Gata1-EGFP transgene (Supplementary Fig. 4d). Overall, these results indicated that mast cell–eosinophil potential was present in GE+ pre-GMs and GE+ GMPs, whereas monocyte–neutrophil potential was present in LMPPs, GE- pre-GMs and GE- GMPs.

**Definition of two progenitor domains by Gata1 expression**

We further investigated the extent to which the progenitors identified had other lineage potential. Comparison of ’lineage programming’ in GE- pre-GMs and GE+ pre-GMs by gene-set–enrichment analysis showed considerable enrichment for the expression of Mk-E–associated genes in GE+ pre-GMs and of lymphoid cell–related genes in GE- pre-GMs (Fig. 5a). All the pre-GM and GMP populations identified generated colonies when cultured under myeloid conditions (Fig. 5b). In contrast, megakaryocytic or erythroid colony–forming ability was associated mainly with GE+ pre-GM populations, which showed an Mk-E potential similar to that seen for pre-Meg-E cells (Fig. 5c,d) and included multi-potent cells with combined
granulocyte and Mk-E potential but no monocyte-macrophage potential (Fig. 5e,f). Conversely, B lymphoid and T lymphoid potential was high in LMPPs but was also substantial in GE− pre-GMs but not GE+ pre-GMs (Fig. 5g and Supplementary Fig. 4e).

Antagonistic transcription factor pairs, including PU.1—GATA-1 and C/EBP—FOG-1, have been proposed to be co-expressed in multi- or bi-potent progenitors, and resolution of the antagonism has been proposed to underlie lineage bifurcations. We therefore

---

**Figure 4** Derivation of eosinophils from GE+ GMPs and of neutrophils from GE− GMPs. (a) Flow cytometry of cells produced from GE− or GE+ GMPs (far left) cultured for 8 d under conditions sustaining all myeloid cell types. Numbers in outlined areas indicate percent Mac-1+Ly6Ghi monocytes (Mo) and Mac-1+Ly6Glo neutrophils (Ne) (left middle), FcεRIα+c-Kit+ mast cells (Ma) (right middle), and SiglecF+ eosinophils (Eo) (far right), among live single cells. (b) Quantification of cell types identified as in a after culture of GMPs and pre-GMs for 8 d as in a, presented as frequency of live cells. (c) Quantification of cell populations identified as in a after culture of GMPs under eosinophil-favoring conditions. (d) Immunocytochemistry detecting EGFP expression (brown) on cytospins of GE− GMP, GE+ GMP and pre-Meg-E, respectively) or mean and s.d. of en enrichment score (NES, normalized enrichment score. (e) Gene-set–enrichment analysis of gene expression by GE− and GE+ pre-GMs, with CLP gene sets (top) and pre-Meg-E gene sets (bottom). NES, normalized enrichment score.

**Figure 5** Segregation of granulocyte-monocyte potential before dissociation from the lymphoid and Mk-E lineages. (a) Gene-set–enrichment analysis of gene expression by GE− and GE+ pre-GMs, with CLP gene sets (top) and pre-Meg-E gene sets (bottom). NES, normalized enrichment score.

---

**ARTICLES**

---

© 2016 Nature America, Inc. All rights reserved.
measured the expression of those and other key transcription factors in GE+ pre-GMs and GMPs, as well as in GE− pre-GMs and GMPs, and in HSCs, LMPPs, pre-CFU-E cells, prec-Meg-E cells and CLPs, to determine if their expression was consistent with the associated lineage potential of each population (Fig. 6). Most notably, Cebpα (which encodes C/EBPα), Cebpβ (which encodes C/EBPβ) and Zfp101 (which encodes FOG-1) were all expressed in GE+ pre-GMs, with Zfp101 being downregulated and Cebpα and Cebpβ being upregulated in GE+ GMPs. The converse change, upregulation of Zfp101 and downregulation of Cebpα and Cebpβ, was seen in pre-Meg-E cells. The FOG-1 target gene Trib2 (ref. 29), which encodes a product (TRB2) that mediates the degradation of C/EBPα and C/EBPβ30,31, was upregulated in pre-Meg-E cells, consistent with the proposed role of C/EBP–FOG-1 antagonism in resolving the multi-potency of GE+ pre-GMs32 and with the observed ability of FOG-1 to suppress differentiation of the eosinophil and mast cell lineages32,33. Sfpi1 (which encodes PU.1) and Gata1 were expressed in GE+ pre-GMs, with Sfpi1 being downregulated in Mk-E–committed pre-Meg-E cells. Gata1 and Sfpi1 were co-expressed in GE+ GMPs, which indicated that downregulation of Gata1 was not decisive for the commitment of GE+ pre-GMs to granulocyte lineages and consistent with PU.1’s being required for mast cell differentiation34. Finally, genes encoding key transcriptional regulators of the differentiation of monocytes–macrophages (Irfβ and Klf4)35 and neutrophils (Gjl1)36 had higher expression in LMPP and GE+ pre-GM and GMP populations than in GE+ pre-GMs and GMPs. These data identified LMPPs and GE+ pre-GMs as distinct multi-potent progenitors with non-overlapping lineage potential, with the former producing lymphocytes, neutrophils and monocytes–macrophages and the latter producing erythrocytes, platelets, eosinophils and mast cells.

Functional heterogeneity of GE+ pre-GMs

Lymphoid–lineage potential has been associated with Flt3+ pre-GMs37. By flow cytometry we observed that Flt3+ pre-GMs were exclusively GE− (Fig. 7a). The selective presence of lymphoid potential in Flt3+ GE− pre-GMs but not Flt3− GE+ pre-GMs raised the possibility that these two populations also differed in other lineage potential. Analysis of bulk and single-cell cultures showed that both Flt3+ GE− pre-GMs and Flt3− GE+ pre-GMs generated monocytes and neutrophils (Fig. 7b,c), with the Flt3+ GE− pre-GMs being perhaps more biased toward neutrophil production. While mast cells did not develop from either progenitor population, a small number of eosinophils developed in the bulk Flt3+ GE− pre-GM cultures (Fig. 7b), and 7% of clones derived from single Flt3+ GE− pre-GMs showed combined neutrophil and eosinophil differentiation. Quantitative RT-PCR profiling of single pre-GMs confirmed the expression of Gata1 in GE+ pre-GMs and of Flt3 in Flt3+ GE− pre-GMs, with Sfpi1 mRNA
expression being similar across the three pre-GM subsets (Flt3+ GE+, Flt3− GE− and GE−) (Fig. 7d–f), consistent with their myeloid-progenitor identity. Mk-E lineage–specific genes and mast cell lineage–specific genes were co-expressed in >90% of single GE+ pre-GMs, and genes encoding antagonistic transcription factor pairs, such as Sfp1-Gata1 and Cebpa-Zfp41, were co-expressed with high frequency (32 of 48 cells and 17 of 48 cells, respectively) in single GE+ pre-GMs (Fig. 7g), suggestive of the molecular properties of multi-potent progenitors. Co-expression of Sfp1-Gata1 and Cebpa-Zfp41 was not observed in Flt3+ or Flt3− GE− pre-GMs, which instead had high expression of monocyte-macrophage- and neutrophil-related genes such as Ifi30 and Csf3r (Fig. 7g), consistent with the presence of such lineage potential in both GE+ sub-populations. In addition, the proportion of Flt3− pre-GMs was lower in mice deficient in the Flt3 ligand than in those with hypomorphic mutation of the gene encoding the tyrosine kinase receptor c-Kit (Wt1/Wt1), which failed to express the recombinant Gata1 allele by Cre recombination expressed under control of the interferon-inducible gene Mx1, after induction by the synthetic RNA duplex poly(I:C) (Supplementary Fig. 5a, b). Consistent with published findings33, Gata1-null GE+ progenitors differentiated into mast cells in vitro (Supplementary Fig. 5e, f), possibly due to redundancy with Gata2 in these progenitors. The gene expression of the myeloid progenitors identified was therefore consistent with the lineage potential observed.

**In vivo** myeloid potential of GE+ and GE− progenitors

Finally, we investigated lineage potential in vivo by transplanting progenitor populations expressing Gata1 together with progenitor populations that lacked Gata1 expression (CD45.2+ and CD45.1+CD45.2−), via an intrafemoral route, into non-irradiated recipient mice (CD45.1+) with hypomorphic mutation of the gene encoding the tyrosine kinase receptor c-Kit (Wt1/Wt1), which gave the transplanted Kit+/− progenitors a competitive reconstitution advantage (Supplementary Fig. 6a). Flow cytometry after 7 d and 10–11 d indicated that Lin−CD11b−Ly6Ghi neutrophils and Lin−CD11b−Ly6G−c-Kit−FceR1α−SiglecF− monocytes were generated by GMPs that did not express Gata1 and Flt3+ GE+ pre-GMs, and Lin−CD11b−Ly6G−c-Kit−FceR1α−SiglecF− eosinophils were generated by GE+ GMPs and GE+ pre-GMs, in both the spleen and bone marrow of the reconstituted mice (Fig. 8a and Supplementary Fig. 6b, c). However, overall engraftment levels, in particular of eosinophils, were low, and no Lin−Ly6G−c-Kit−FceR1α− mast cells were detected (Fig. 8a and Supplementary Fig. 6b, c).

To overcome that obstacle, we assessed the ability of GMPs and pre-GMs to engraft after intraperitoneal injection and found that total engraftment of the peritoneum by both pre-GMs and GMPs was one to two orders of magnitude greater than that in bone marrow and spleen after intrafemoral transplantation (Supplementary Fig. 6d), and mast cell development was also readily detected (Fig. 8b). Following intrafemoral transplantation, GE+ GMPs generated exclusively peritoneal mast cells and eosinophils, whereas only monocytes and neutrophils were derived from GE− GMPs (Fig. 8c). Similarly, GE− pre-GMs produced predominantly mast cells and eosinophils, whereas Flt3+ GE− pre-GMs and LMPPs generated almost exclusively neutrophils and monocytes (Fig. 8d, e). The cellular phenotypes defined by flow cytometry were further confirmed by morphology (Supplementary Fig. 6e–h) and selective GE expression in eosinophils and mast cells (Supplementary Fig. 6i). GE− pre-GMs generated low numbers of neutrophils (but no monocytes), and LMPPs and Flt3+ GE− pre-GMs generated limited numbers of eosinophils (but no mast cells) (Fig. 8d–e), which indicated that the more upstream myeloid progenitors (LMPPs and
pre-GMs) retained some lineage plasticity restricted to the eosinophil and neutrophil lineages *in vivo*, which was lost in both GE+ GMPs and GE− GMPs (Fig. 8c–e). These results confirmed the same lineage potential *in vivo* as that noted *in vitro* for both GE+ myeloid progenitors (GE+ GMPs and GE+ GMPs) and GE− progenitors (LMPPs, GE− GMPs and GE− GMPs). We therefore propose to designate GE+ GMPs as candidate erythroid-megakaryocyte-primed multi-potent progenitors (EMkMPPs) and GE− GMPs as eosinophil–mast cell progenitors (EoMPs). In contrast, since GE− GMPs and GE− GMPs generated predominantly neutrophils and monocytes, we propose to designate them as pre–neutrophil-monocyte progenitors (pre-NMs) and neutrophil–monocyte progenitors (NMPs), respectively. In various current hematopoietic models, all monocyte-macrophage lineage potential and granulocyte lineage potential is proposed to reside in one common GMP, our results suggested that the separation of lineages that expressed *Gata1* (megakaryocytes, erythrocytes, eosinophils, and mast cells) and lineages that did not express *Gata1* (lymphocytes, neutrophils, and monocytes) represented an early lineage bifurcation of the multi-potent HSCs. These findings substantially change the view of how myeloid cell types are specified by suggesting that mast cells and eosinophils are generated from an EMkMPP shared with platelets and erythrocytes, rather than from a common GMP. This type of lineage relationship has been suggested before on the basis of observations of myeloid malignancies. Platelets have important innate immunological functions, acting as sentinels and providing chemotaxis of myeloid effector cells and the platelet-ancestral thrombocyte is capable of bacterial phagocytosis. These are all functions shared with mast cells, consistent with the evolution of mast cells and platelets through specialization from a common ancestral cell with both hemostatic function and phagocytic function. Our results suggest that eosinophils might have evolved from the same GATA-1-expressing hematopoietic branch, possibly as an adaptation to parasites. In contrast, we found that monocytes–macrophages were developmentally closely related to lymphoid cells, as has already been suggested by the identification of progenitors restricted to macrophage and lymphoid differentiation in both fetal hematopoiesis and adult hematopoiesis.

Published lineage analysis of myeloid colonies derived from heterogeneous human CD34+ hematopoietic cells is compatible with a similar separation of human GATA-1-expressing granulocyte lineages along with megakaryocytic lineage potential and erythroid lineage potential, with the important distinct that considerable macrophage potential was also observed together with the GATA-1-expressing lineages. This might be related to the fact that the human cell populations investigated typically also had extensive lymphoid potential of unknown relationship to the granulocyte lineages investigated.

**DISCUSSION**

Here we have prospectively isolated and characterized distinct, myeloid progenitor populations restricted to either the neutrophil-monocyte fate (pre-NM and NMP) or eosinophil–mast cell fate (EoMP). We have also prospectively isolated and characterized candidate multi-potent progenitor cells with combined Mk-E potential and eosinophil–mast cell potential (EMkMPPs). In addition, we found that LMPPs produced neutrophils and monocytes but no mast cells and few or no eosinophils.

While existing models indicate that the segregation of monocyte-macrophage potential and granulocyte potential occurs downstream of a common GMP, our results suggested that the separation of lineages that expressed *Gata1* (megakaryocytes, erythrocytes, eosinophils, and mast cells) and lineages that did not express *Gata1* (lymphocytes, neutrophils, and monocytes) represented an early lineage bifurcation of the multi-potent HSCs. These findings substantially change the view of how myeloid cell types are specified by suggesting that mast cells and eosinophils are generated from a common GMP. This type of lineage relationship has been suggested before on the basis of observations of myeloid malignancies. Platelets have important innate immunological functions, acting as sentinels and providing chemotaxis of myeloid effector cells, and the platelet-ancestral thrombocyte is capable of bacterial phagocytosis. These are all functions shared with mast cells, consistent with the evolution of mast cells and platelets through specialization from a common ancestral cell with both hemostatic function and phagocytic function. Our results suggest that eosinophils might have evolved from the same GATA-1-expressing hematopoietic branch, possibly as an adaptation to parasites. In contrast, we found that monocytes–macrophages were developmentally closely related to lymphoid cells, as has already been suggested by the identification of progenitors restricted to macrophage and lymphoid differentiation in both fetal hematopoiesis and adult hematopoiesis.

**Figure 8** *In vivo* myeloid reconstitution of GE− and GE+ myeloid progenitor cells. (a) Gating strategy for the identification of peritoneal neutrophils (Ne), mast cells (Ma), monocytes (Mo) and eosinophils (Eo). (b) Gating strategy for the identification of donor-derived mast cells, eosinophils, neutrophils, and monocytes at 7 d (GMP) or 11 d (pre-GM) after intraperitoneal transplantation of GMPs or pre-GMs. Numbers adjacent to outlined areas indicate percent CD45.1*CD45.2* donor cells (top right) or CD45.2* donor cells (top left). (c–e) Quantification of the *in vivo* contribution of GE+ or GE− donor-derived GMPs (c), Flt3+ GE+ or Flt3− GE− donor-derived pre-GMs (d) and donor-derived LMPPs (e) to peritoneal mast cell, eosinophil, neutrophil and monocyte populations (horizontal axes) at various times (keys) after intraperitoneal transplantation of 1 × 103 to 4 × 103 cells per donor (presented per 106 mononuclear cells analyzed per 105 injected cells). *, reconstitution below detection limit (Supplementary Fig. 6). Data are representative of ten experiments (a), two experiments (b), or two experiments (days 7 and 11) or one experiment (days 4, 15 and 20) (c–e; mean and s.e.m.).
CMPs have been defined as having myeloid-lineage potential and Mk-E–lineage potential but lacking lymphoid-lineage potential and would potentially generate both pre-NM–NMP populations and MKEMPP-EoMP populations. However, it has not been established if the CMP population as currently defined (Lin−Sca-1−c-Kit+CD34+CD41b+) includes cells with monocyte-macrophage potential as well as all granulocyte potential at the clonal level, and further studies will therefore be needed to firmly establish its relationship to the myeloid progenitor populations identified here. In addition, our studies do not preclude the possibility that an upstream progenitor with the same lineage potential as the EMkMPP exists, for example, within the proposed CMP compartment.

In summary, our findings have provided an improved cellular and molecular template for elucidation of the evolutionary and developmental relationships of hematopoietic cell types, as well as the identification of cells-of-origin of hematopoietic malignancies. Our results emphasize the power of single-cell transcriptome analysis to simultaneously resolve cellular heterogeneity and identify the molecular determinants selectively and uniformly expressed in novel cell subsets to allow their purification and functional characterization.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: microarray data, GSE49241; single-cell RNA sequencing data, GSE77029.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank A. Cuman (Pasteur Institute, Paris) for OP9 and OP9-DL1 stroma cells; the EMBL Monterotondo Gene Expression Service and Transgenic Core Facility for generating the Gata1-EGFP bacterial artificial chromosome and the corresponding transgenic mouse line, and flow cytometry facilities at the ISCR (supported by Wellcome Trust Equipment Grant WT08371MA (C.N.) and WIMM (supported by Medical Research Council Grants MC_UU_12009, MC_UU_12025 and G0902418) for flow cytometry and cell sorting. Supported by the Association for International Cancer Research (11-0724 to C.N.), the Medical Research Council (UK Program Grant G0801073 and Unit Grant MC_UU_12009/5 to S.E.W.J.; Strategic Grant G0701761, Program Grant G0900892 and Unit Grant MC_UU_12009/7 to C.N.) and the European Commission FP7 (CardioCell project (C.N.) and EuroSyStem project (S.E.W.J.)).

AUTHOR CONTRIBUTIONS

R.D., N.B.-V., P.W., A.M., S.E.W.J. and C.N. designed the experiments; R.D., N.B.-V., P.W., A.Ga., A.Gi., A.Z., M.L. and A.Gr. performed the experiments; N.B.-V., P.W., R.D., A.Ga., S.T., A.M., E.S., S.E.W.J. and C.N. designed the experiments; R.D., N.B.-V., P.W., A.M., E.S., S.E.W.J. and C.N. designed the experiments; R.D., N.B.-V., P.W., A.M., E.S., S.E.W.J. and C.N. designed the experiments; R.D., N.B.-V., P.W., A.M., E.S., S.E.W.J. and C.N. designed the experiments; R.D.,
40. Luc, S., Buza-Vidas, N. & Jacobsen, S.E. Biological and molecular evidence for existence of lymphoid-primed multipotent progenitors. *Ann. NY Acad. Sci.* **1106**, 89–94 (2007).
41. Arock, M., Schneider, E., Boissan, M., Tricottet, V. & Dy, M. Differentiation of human basophils: an overview of recent advances and pending questions. *J. Leukoc. Biol.* **71**, 557–564 (2002).
42. Vieira-de-Abreu, A., Campbell, R.A., Weyrich, A.S. & Zimmerman, G.A. Platelets: versatile effector cells in hemostasis, inflammation, and the immune continuum. *Semin. Immunopathol.* **34**, 5–30 (2012).
43. Nagasawa, T. et al. Phagocytosis by thrombocytes is a conserved innate immune mechanism in lower vertebrates. *Front. Immunol.* **5**, 445 (2014).
44. Abraham, S.N. & St John, A.L. Mast cell-orchestrated immunity to pathogens. *Nat. Rev. Immunol.* **10**, 440–452 (2010).
45. Cumano, A., Paige, C.J., Iscove, N.N. & Brady, G. Bipotential precursors of B cells and macrophages in murine fetal liver. *Nature* **356**, 612–615 (1992).
46. Lacaud, G., Carlsson, L. & Keller, G. Identification of a fetal hematopoietic precursor with B cell, T cell, and macrophage potential. *Immunity* **9**, 827–838 (1998).
47. Kawamoto, H., Ohmura, K. & Katsura, Y. Direct evidence for the commitment of hematopoietic stem cells to T, B and myeloid lineages in murine fetal liver. *Int. Immunol.* **9**, 1011–1019 (1997).
48. Montecino-Rodriguez, E., Leathers, H. & Dorshkind, K. Bipotential B-macrophage progenitors are present in adult bone marrow. *Nat. Immunol.* **2**, 83–88 (2001).
49. Görgens, A. et al. Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. *Cell Rep.* **3**, 1539–1552 (2013).
50. Miyawaki, K. et al. CD41 marks the initial myelo-erythroid lineage specification in adult mouse hematopoiesis: redefinition of murine common myeloid progenitor. *Stem Cells* **33**, 976–987 (2015).
### ONLINE METHODS

**Mouse lines.** The Gata1-EGFP transgene was generated by bacterial artificial chromosome recombinant engineering in bacteria at the EMBL Genome Engineering Core Facility (http://www.embl.de/services/genome-engineering-service/index.html). In this transgene an EGFP-polyA cassette replaced the coding part of the second exon of the mouse Gata1 gene. Gata1-EGFP mice were generated by intracytoplasmic sperm injection at the EMBL Monterotondo Transgenic Facility (http://www.embl.de/services/transgenic_facility/index.html). Mice were backcrossed for more than six generations onto a C57BL/6J background. Mice were bred and maintained at the animal facilities of Edinburgh University and Oxford University, UK. Experimental protocols were approved by the Edinburgh University School of Biological Sciences Ethical Review Committee, and the Oxford University Clinical Medicine Ethical Review Committee. Flt3l knockout, Kit W/W, Gata1 conditional knockout mice and Mx1-Cre transgenic mice were previously generated. All experiments were performed under projects licensed from the UK Home Office. Mice were sacrificed for harvesting hematopoietic cell populations at an age of 6–17 weeks. Both male and female mice were used.

**Flow cytometry.** BD LSRRFortessa, LSRII, a FACSARia II, FACSARia III and FACSARia Fusion (BD Biosciences) were used for flow cytometry and cell sorting. Cells were prepared for cell sorting by either CD117 enrichment by autoMACS Pro Separator (Miltenyi Biotec) or lineage depletion using immunomagnetic beads. In staining where anti-FcRII/III antibody was not included, cells were incubated with Fc-block. Flowjo analysis software (TreeStar) was used for subsequent data analysis. Antibodies used are listed below. Populations were defined as follows: HSC, Lin−c-Kit+Flt3−CD150+Gata1-EGFP−; LMPP, Lin−Lin− Sca-1−c-Kit+Flt3+Gata1-EGFP−; CLP, Lin−B220−Sca-1−c-Kit+Flt3+IL-7Rα−; MkP, Lin−Lin−c-Kit+CD150−CD41−; pre-GM, Lin−Lin−c-Kit+CD41+FCR/ III+CD105−; GMPC, Lin−Lin−c-Kit+CD41+FCR/II+CD105+; pre-Meg-E, Lin−Lin−c-Kit+CD41+FCR/III+CD105−; pre-CFU-E, Lin−Lin−c-Kit+CD41+FCR/III+CD105+; and CFU-E, Lin−Lin− c-Kit+CD41+FCR/III+CD105+. Eosinophils from the peritoneal cavity: Lin−CD45.1+; CLP, Lin−B220−Sca-1−c-Kit+Flt3+IL-7Rα−; eosinophils from peripheral blood: Lin−CD4−CD5−CD8−Ter119−B220−; eosinophils, Lin−CD11b+Ly6G+SiglecF+FcεRic F4/80+Gata1-EGFP−. Macrophages from the peritoneal cavity: Lin−CD45.1+CD45.1+; CLP, Lin−B220−Sca-1−c-Kit+Flt3+IL-7Rα−; Lin−CD11b+Ly6G+SiglecF+FcεRic F4/80+Gata1-EGFP−. Neutrophils from peripheral blood: Lin−CD45.1+CD45.1+; CLP, Lin−B220−Sca-1−c-Kit+Flt3+IL-7Rα−; neutrophils, Lin−CD11b+Ly6G+SiglecF+FcεRic F4/80+Gata1-EGFP−.

**Transplantation studies.** In order to directly compare the in vivo potential of pre-GM and GMP progenitors subfractionated based on Gata1 expression, equal numbers of GE+ and GE− cells were co-transplanted competitively into W/W or W/W recipient mice (5–10 weeks of age) on a C57BL/6J and CD45.1+ background. By using Gata1-EGFP/p+ mice on a CD45.2+ or a mixed CD45.1+CD45.2+ background, the contribution of the two competitor populations to the myeloid lineages could be separated and distinguished from each other endogenous cells based on CD45 allotype, and directly compared in a single recipient. Briefly, GE+ GMPs or GE−Flt3− pre-GMs were sorted from a CD45.2+ donor and mixed with equal cell numbers of GE− GMPs or Flt3+ GE+ pre-GMs, or vice versa. LMPP cells were isolated from either CD45.1+CD45.2+ or CD45.2+Gata1-EGFP mice. Purified cells were either injected intraperitoneally (1,000–4,000 cells from each donor) or intravenously (1–2,000–11,000 cells from each donor) into W/W or W/W CD45.1+ mice. For intraperitoneal transplantation, recipients were injected intraperitoneally 12–24 h before transplantation with 1.5 ml of distilled water to deplete endogenous mast cells. Contribution to the myeloid lineages was assessed by peritoneal lavage 4–20 d post-injection for mice injected intraperitoneally and by analysis of injected femur and spleen 7 and 10 days post-injection for mice injected intravenously. The contribution of CD45.2+ and CD45.1+CD45.2+ donor cells to mast cells, eosinophils, monocytes and neutrophils was assessed by flow cytometry using the following markers: mast cells, Lin−Ly6G−c-Kit+Fcer1T−; eosinophils, Lin−CD11b+Ly6G−c-Kit+Fcer1T−SiglecF+; neutrophils, Lin−CD11b+Ly6G−c-Kit+Fcer1T−SiglecF+; monocytes, Lin−CD11b+Ly6G−c-Kit+Fcer1T−SiglecF+; and Lin−CD45.1+CD45.2+ to each lineage was adjusted to the number of cells injected and number of total mononuclear cells analyzed, and is presented as cells per million mononuclear cells per 1,000 transplanted cells analyzed. Average values below two events per 1 million cells analyzed per 1,000 cells injected were not above the non-specific background and therefore not used to evaluate lineage-contribution of donor-derived cells.

**In vitro cultures.** To test myeloid potential, cells were cultured in IMDM with L-glutamine (Invitrogen), 20% heat-inactivated FCS (Gibco), penicillin/streptomycin (Invitrogen) and 0.1 µM β-mercaptoethanol (Sigma), supplemented with 20 ng/ml mSCF (Peprotech), 20 ng/ml mIL-3 (Peprotech), 50 ng/ml mIL-7 (R&D Systems) and 50 ng/ml mIL-9 (R&D Systems) at 37 °C, 5% CO2. Bulk cultures of LMPP and pre-GM were started with 100–200,000 cells per culture. Myeloid potential of single cells was tested by sorting single cells directly into Terasaki plates, unless stated differently in the text, each well containing 20 µl medium. GMP cells were derived in vitro by culturing progenitors (pre-GM or LMPP) for 3 days with 100 ng/ml mSCF and 10 ng/ml mIL-3 in round bottom 96 well plates and FCRII/III+ cells were in this experiment considered as GMPS. In vitro derived GMPs were manually plated at an average of 5 cells/well into 60 well Terasaki plates. For eosinophil favoring culture conditions, GMP cells (200 cells/culture) were grown with 20 ng/ml mSCF, 10 ng/ml mGM-CSF and 50 ng/ml IL-5 in round bottom 96 well plates. After 5 days, the cells were washed and put back in culture for another 2 days without mSCF and mGM-CSF. All eosinophils were compared with a Shandon cytoSpin at 1,000 RPM with low acceleration, followed by May-Grunwald-Giemsa stain (VWR).

**Immunocytochemistry.** Air dried cytoSpins were fixed with 4% PFA in PBS for 5 min. Slides were first blocked with PBS/5%BSA containing Fc-block for 30 min at room temperature and then with 3% H2O2 in water for 5 min. Slides were incubated with chicken polyclonal anti-GFP (Abcam, ab13970) in PBS/5% BSA for 4 h at room temperature and next with peroxidase donkey anti-chicken antibody (Stratech, 703-035-155-JIR) for 1 h at room temperature. Peroxidase was detected with DAB substrate kit (Vector Laboratories, SK-4100) according to manufacturer's instructions. Cells were counterstained with Harris Modified Hematoxylin Solution (Sigma-Aldrich, HHS32) for 10 min.

** Colony forming assays.** For evaluation of myeloid potential from progenitor populations, cells were seeded in Methocult M3436 and myeloid potential was evaluated with Methocult M3534. Colonies were scored after 8 days of culture. CUFX-Mix colonies were scored in Methocult M3434 based on colonies containing red cells after 8 days of culture. Megakaryocytic potential was evaluated using the Megacult collagen-based assay. All these media were from StemCell Technologies, Vancouver, Canada. Megakaryocyte cultures were supplemented with 10ng/ml mIL-3, 20 ng/ml hIL-6 (Peprotech), 50 ng/ml hIL-11 (Peprotech), 50 ng/ml hThpo (Peprotech), and megakaryocytes detected using acetylthiocholinioide staining (Sigma) according to manufacturer's instructions after 7–8 days of culture. For evaluation of Mk potential from single GE+ and GE−Lin−ScA1−c-Kit+Flt3+ cells, these cells were sorted into X-vivo15 medium containing glutamax and gentamycin (BioWhittaker) supplemented with 10% FCS (HyClone), 0.1 µM β-mercaptoethanol, 50 ng/ml mSCF, 50 ng/ml hLT3 ligand (hLF; Immunex), 50 ng/ml hThpo, 20 ng/ml mIL-3 and 5 U/ml hePO (Roche). For each experiment, 240 cells
were manually plated at 1 cell per well into 60 well Terasaki plates. MK potential was evaluated 8 or 12 days after culture using an inverted microscope.

**Measurement of lymphoid lineage potential.** B lymphocyte and T lymphocyte potential of purified progenitor populations were evaluated following co-culture on OP9 and OP9-DL1 stroma cells, respectively (provided by A. Cumano, Pasteur Institute, Paris, France). LMPPs, GE+ and GE− pre-GMs, and GE− GMPs were seeded onto stromal monolayers. OP9 cultures were supplemented with 25 ng/ml mSCF, 25 ng/ml hFL (Peprotech) and 10 ng/ml hIL-7 (Peprotech), while OP9-DL1 cultures were supplemented with 25 ng/ml mSCF and 25 ng/ml hFlt3L. SCF was removed after 7 days of culture. Readout was by flow cytometry at 14–21 days of culture. B cells were defined as CD19+ and T cells were defined as CD4+CD8− and/or Thy1.2−CD25− (Supplementary Fig. 4.e.f).

**Quantitative PCR.** For gene expression analysis of bulk cultures and cultured cell populations, RNA was purified from 2,000–6,000 cells per replicate with Trizol (Invitrogen, Carlsbad, CA, USA) and reverse-transcribed with SuperScript VILO cDNA synthesis kit according to the manufacturer’s instructions (Invitrogen, Carlsbad, CA, USA). For single cells and progenitor populations (200 cells per replicate), CellsDirect One-Step qRT-PCR kit (Life technologies, 11753–100) was used according to manufacturer’s protocol for preparation and amplification of cDNA. The BioMark 96.96, 48.48 or 192.24 Dynamic Array platform (Fluidigm) and either Taqman assays (for progenitor populations and single cells) or the Universal probe library system (Roche) (for bulk cultures and cultured cell populations) were used according to the manufacturers’ instructions. UPL primers and probes were chosen by the Assay Design Center web service for mouse genes. Default parameters were used and highest-ranking primers-probe combination was chosen for each gene.

**Quantitative PCR analysis of single cells.** ΔCT values relative to Hprt (ΔCT(Hprt) – ΔCT(Gene)) were zero centered for each gene by subtraction of the average expression values 1. We selected top 100 genes with the highest absolute correlation coefficient (PCA component loadings) in one of the first three components for the hierarchical clustering analysis. Genes differentially expressed between the two main cellular were identified using analysis of variance (P < 0.001; false-discovery rate <0.05) and used for the final clustering of single cells.

**Statistical analysis.** Student’s t-test was used to determine statistical significance for normally distributed data. For single-cell expression levels, the Mann-Whitney U-test was used, and Fisher’s exact test was used to compare expression frequencies at the single cell level between populations. No statistical method was used to predetermine sample size, and experiments were not randomized. No data were excluded from analysis. The Investigators were not blinded to allocation during experiments or outcome assessment.

**Antibodies.** Information on antibodies (including clone, concentration and vendor) is provided in Supplementary Table 2.

**Taqman assays.** Information on these assays is provided in Supplementary Table 3.

---

51. Moreira, P.N., Pozuelo, J., Girald, P., Gutierrez-Aadán, A. & Montoliu, L. Generation of yeast artificial chromosome transgenic mice by intracytoplasmic sperm injection. Methods Mol. Biol. 349, 151–161 (2006).
52. McKenna, H.J. et al. Mice lacking R3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells. Blood 95, 3489–3497 (2000).
53. Lindeboom, F. et al. A tissue-specific knockout reveals that Gat1 is not essential for Sertoli cell function in the mouse. Nucleic Acids Res. 31, 5405–5412 (2003).
54. Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. Science 269, 1427–1429 (1995).
55. Kanakura, Y. et al. Changes in numbers and types of mast cell colony-forming cells in the peripheral cavity of mice after injection of distilled water: evidence that mast cells suppress differentiation of bone marrow-derived precursors. Blood 71, 573–580 (1988).
56. Inizan, R.A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Bioinformatics 4, 249–264 (2003).
57. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
58. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27, 2325–2329 (2011).