Redefining hematopoietic progenitor cells and reforming the hierarchy of hematopoiesis

Lipeng Chen[1,4], Qing Sun[2,4], Guoqiang Li[3], Qijun Huang[1], Sujin Chen[1], Yingyun Fu[1], Yongjian Yue[1,*]

[1]Shenzhen Institute of Respiratory Disease, the First Affiliated Hospital of Southern University of Science and Technology, Shenzhen People's Hospital, Shenzhen, Guangdong, China

[2]Shenzhen Key Laboratory of Reproductive Immunology for Peri-implantation, Fertility Center, Shenzhen Zhongshan Obstetrics & Gynecology Hospital (formerly Shenzhen Zhongshan Urology Hospital) , Shenzhen, China

[3]Department of hematology, The Second Clinical Medical College of Jinan University, Shenzhen People''s Hospital, Shenzhen, Guangdong, China

4 Lipeng Chen and Qing Sun contributed equally to this work.

*** Correspondence:** Yongjian Yue

Address: Shenzhen People's Hospital, The First Affiliated Hospital of Southern University of Science and Technology, No. 1017 Dongmen North Road, Luohu District, Shenzhen 518020, China.

Email: yueyongj@163.com

Tel: +86 0755-2553 3018

Summary

Deciphering the mechanisms underlying progenitor cell definition and fate decisions is critical to answering fundamental questions regarding hematopoietic lineage commitment. Here, we redefine the entire spectrum of original hematopoietic progenitor cells (HPCs) using a comprehensive transcriptional atlas, effectively subverting the conventional definition of HPCs. We have defined the transcription factors associated with the key fate decisions at each level of the hematopoietic hierarchy, providing additional insights into the underlying molecular mechanisms. Initial hematopoietic progenitors are simultaneously primed into erythroid, lymphoid, and neutrophilic progenitors during the first differentiation stage of hematopoiesis. Moreover, plasma progenitor cells were identified and defined. A hematopoietic hierarchy roadmap was reformed, demonstrating that the megakaryocytic–erythroid lineage commitment process is continuous. However, the lymphoid lineage commitment process was identified as a discrete stepwise process with distinct lineage-committed populations during hematopoiesis. Our study has revealed numerous possibilities for precisely controlling progenitor cell differentiation, facilitating research development in regenerative medicine and disease treatment.

Keywords: hematopoietic progenitor cells, redefinition, lineage commitment, fate decision, hierarchy

Highlights

- Hematopoietic progenitors are redefined using comprehensive transcriptional atlas
- Fate decision-related transcription factors are revealed in the hematopoietic hierarchy
- Progenitor lineage commitment includes continuous and stepwise processes
- The hierarchy of hematopoiesis was reformed

Introduction

Hematopoietic stem cells (HSCs) comprise a rare cell population in the peripheral blood that can self-renew and differentiate into full blood cell lineages with tree-like differentiation patterns in classical hematopoietic hierarchical models [1]. Whereas the diversity of human peripheral blood cells is known, their conflicting definitions, lineage heterogeneity, and dynamic differentiation present challenges in determining their precise characteristics. Single-cell sequencing technology has been successfully implemented to explore the molecular heterogeneity and hierarchy of HSCs [2-4]. Lineage commitment is affected by transcription factors (TFs) within a continuum of the fate-decision process [5, 6]. Although *in vitro* and *in vivo* studies have demonstrated the critical role of TFs in lineage commitment, the function of, and the regulatory mechanism underlying HSC differentiation remains unclear [7]. In addition, a transcriptional atlas, and the fate decisions of entire spectrum progenitor cell subtypes in adult peripheral blood have not yet been elucidated.

Characterizing the molecular mechanisms underlying lineage commitment and hematopoietic hierarchy is challenging for several reasons. First, the definition of hematopoietic progenitor cells (HPCs) has not been standardized. Based on classical and traditional definitions, HPCs are characterized by surface markers, including CD34, CD38, FLT3, and KIT[5]. Combined surface markers, including CD10 (MME), FLT3, and CD38, have also been proposed to identify common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs). Nevertheless, owing to their rare and transient nature, CLP and CMP subtypes of HSCs have not been strictly characterized nor accurately captured. Furthermore, surface markers are limited in capturing the internal state of progenitor cells and reflecting the dynamic differentiation mechanisms underlying hematopoiesis [8]. Second, HPCs have been characterized in functional studies using selected CD34-positive cells; however, this strategy is affected by the culture methods and surface markers used for identification. Therefore, combining surface markers for HPC characterization could lead to bias and arbitrariness. Third, bone marrow progenitors, widely used in functional studies and subgroup identification, are quiescent, lacking the entire spectrum of progenitor

stages[9]. Finally, capturing the entire spectrum of adult peripheral blood HPCs is challenging in scope.

Recently, many studies have identified progenitor cell lineages using single-cell sequencing. Although several hematopoietic hierarchy have been proposed, they contradict each other [1, 10, 11]. Moreover, there are several limitations associated with defining the transcriptional atlas. In t-distributed stochastic neighbor embedding (T-SNE) plots, the single-cell RNA sequencing (scRNA) expression data of $CD34^+$ cells exhibits a scattered and mixed distribution of two well-defined progenitor subtypes, CLPs and CMPs, implying that the transcriptional atlases of these cell types are not distinguishable using unsupervised dimensionality reduction and clustering. Most established lineage fate decision models are based on *in vitro* and *in vivo* functional studies or cell fate induction experiments [12, 13]. However, cellular conditions can frequently lead to inaccurate interpretations of HPC fate decisions as the cell lineages are induced by diverse experimental conditions that may not apply to all possible branches of lineage commitment [13]. Similarly, single-cell studies tend to modify aspects of a classical hematopoietic hierarchy as hematopoietic progenitors change continuously during the progenitor differentiation process [5, 14]. Moreover, progenitor cells have been defined and selected using classical surface markers [5, 6]. Notably, each type of progenitor is rare in the blood; thus, the number of captured cells is frequently very low (< 100 cells) [5], implying the omission of key progenitor subgroups. Hence, the scRNA sequencing of bone marrow quiescent progenitors does not reflect the full spectrum of hematopoietic progenitor lineage commitment, affecting the reliability of results. Furthermore, an inefficient and inaccurate classical definition of hematopoietic progenitors can result in disorganized subgrouping without a reliable boundary [2]. Additionally, the unsupervised dimensionality reduction and clustering of lineage commitment will result in a farraginous model [4]. Therefore, determining lineage definitions and fate decisions and establishing reliable lineage commitment during hematopoiesis remain challenging [10, 15].

In this study, we aim to redefine adult hematopoietic progenitors, identify key TFs involved in lineage fate decisions, and reform the hematopoietic hierarchy. Our

HPC atlas redefines the cellular properties of progenitors at different stages throughout the hematopoietic lineage commitment process. As such, the results provide new perspectives for HPC characterization, definition, and classification.

Results

Determining characteristics of progenitor cells in adult peripheral blood using scRNA

To enrich the entire HPC spectrum and intermediate-state progenitors, mature lymphocytes and granulocytes were removed using a surface marker panel (Lineage−), while progenitor cells were enriched via negative deletion (Figure 1A). We performed 10× genomic scRNA sequencing, unsupervised dimensionality reduction, and clustering analysis using Seurat and uniform manifold approximation and projection (UMAP), identifying 32 clusters (Figure 1B). Overall, 159,603 cells were captured and sequenced, and our single-cell expression atlas showed the heterogeneity of several $CD34^+$ progenitor clusters. We further analyzed significant genes identified among progenitor and other cell clusters. *SXO4*, *CDK6*, *CYTL1*, and *EGFL7* were highly expressed across the progenitor clusters (Figures 1C and D), consistent with the expression patterns in bone marrow progenitors (Figure S1). *EGFL7* and *CYTL1* were highly and specifically expressed in the progenitor clusters, indicating that they are likely the key functional genes for HPC lineage differentiation and maturation. The top 30 highly expressed genes in the progenitor cells are presented in Figure 1E. The gene set enrichment analysis of differentially expressed genes showed that the hematopoietic cell lineage, MAPK, and PI3K–Akt pathways were enriched in progenitor cells, but not significantly (Figure 1F). Gene Ontology (GO) analysis also revealed that biological processes related to DNA replication, cell differentiation, and the regulation of hemopoiesis were enriched (Figure 1G). Therefore, based on the surface markers and the aforementioned key functional genes associated with HSCs, we confirmed that the identified $CD34^+$ clusters were HPCs.

Conventional surface markers are inadequate and ineffective for accurately defining HPCs

A re-clustering analysis was performed on 27,566 $CD34^+$ cells to evaluate diversity, resulting in 17 clusters, including 21,244 mobilized and 6,322 quiescent progenitors (Figure 2A). Most clusters comprised many progenitor cells (Table S1), and the identified progenitor cell subgroup revealed category and functional heterogeneity. Based on our findings, the cell subgroup aggregated in three main differentiation directions. Conventionally, membrane markers of CMPs and granulocyte–monocyte progenitors (GMPs) were defined based on the CD38+MME-FLT3+ phenotype, whereas megakaryocytic–erythroid progenitors were defined using the CD38+MME-FLT3− phenotype. However, progenitor clusters based on these markers were not distinguishable among the subgroups. *KIT* (*CD117*) and *FLT3* were scattered expressed in different clusters (Figure 2B), indicating that the $CD34^+KIT^+$ and $CD34^+FLT3^+$ progenitors comprise two different differentiation directions. *CD38* and *FLT3* showed contradictory expression patterns in clusters (Figure 2B), implying that surface markers, such as CD38, FLT3, KIT, and MME, can be used to distinguish myeloid and lymphoid progenitor differentiation directions but cannot be used for the classification of subtypes and intermediate stages.

Canonical markers, including *CD49f* (*ITGA6*) and *THY1*, were present only in a small proportion of HPCs (Figure S1A). Additionally, our data showed that using *CD49f*, *THY1*, *FLT3*, and *KIT* to classify or define progenitor cell groups was artificial and subjective. We further analyzed the data of four bone marrow single-cell projects to determine whether the conventional classification markers *THY1* and *CD49f* were expressed by progenitor cells (Figure S1B). Overall, data for 8,296 $CD34^+$ cells were extracted from the four merged project datasets, exhibiting a scattered expression distribution, with 84 *THY1*-positive and 824 *CD49f*-positive cells. This result is similar to that for the other identified markers, including FLT3, CD38, and KIT (Figure S1C). Unsupervised dimensionality reduction and clustering in bone marrow $CD34^+$ cells could not distinguish the progenitor cells (CMPs or CLPs) according to the traditional classification criteria. The progenitors of $FLT3^+$ cells were

identified as a mixed and diverse population, which could explain why previous functional studies of selected FLT3^{+} progenitors presented varying differentiation potential[16]. Classical and conventional surface markers were also inadequate, specious, and ineffective for accurately distinguishing CMPs, GMPs, and CLPs and characterizing HPCs.

Redefinition of adult peripheral blood progenitors

Unsupervised dimensionality reduction analysis was performed to identify new efficient marker genes for distinguishing progenitor clusters. The progenitors were clustered in several directions (Figure 2A). *AVP* was expressed in HSCs and, along with *CSF3R*, was primarily expressed in clusters 0 and 3. Hence, these two clusters were likely the original HSC progenitors (Figures 2B and C). The myeloid progenitors were found in clusters 1, 2, 5, and 6, with the gradual upregulation of *CD38*, *GATA1*, and *KIT* expression (Figure 2B). The one-directional lineage clusters 1, 2, and 6 could be distinguished based on the expression of TFs, such as *KLF1*, *GATA1*, and *GATA2* (Figure 2B), indicating that the clusters were megakaryocyte or erythrocyte progenitors. Meanwhile, cluster 5 presented myeloperoxidase (*MPO*) and *CSF3R* marker genes (Figure 2C). MPO is the most abundant and uniquely expressed protein in human neutrophils, with a vital role in neutrophil antimicrobial responses [17, 18]. Hence, the progenitor branches were likely granulocyte–monocyte progenitors, oriented entirely differently from megakaryocyte or erythrocyte progenitors. *IGHM*, *MME*, and *HOXA9* [19] were expressed in cluster 4, confirming that this cluster comprised lymphoid progenitor cells (Figure 2C). Expression of the TF HOXA9 increased sequentially from clusters 0 to 4. Accordingly, TFs could fully distinguish the intermediate state and reflect the mechanisms underlying hematopoietic lineage commitment. The use of the expression and activation of TFs with unique marker genes was more efficient for defining HPC lineages.

To examine differences between mobilized and quiescent (unperturbed) progenitor cells, cell distribution analysis was performed. The total number of progenitor cells was remarkably increased after G-CSF mobilization. A comparison of

the progenitor cells before and after mobilization revealed that all progenitor cell subtypes existed in the quiescent state (Figure 2D). *CSF3R* expression was consistent with G-CSF mobilization in the peripheral blood during HSC transplantation [20]. This signifies that the entire spectrum of progenitor cells exists in a quiescent state in the peripheral blood, consistent with the cluster distributions in the mobilized state. This might also indicate that unperturbed adult hematopoiesis depends on replenishment from multiple potential progenitor cells. Whereas the progenitor cells gradually differentiated into mature cells in the peripheral blood, we expected that fold-changes for the more primitive progenitors would be higher than those for mature progenitors. The highest cell numbers and fold-changes were found in clusters 0 and 3 (Figure 2E), the most original multi-potent clusters known as multi-potent progenitor cells (MPCs). MPCs can differentiate into multiple types of lineages.

To identify the common myeloid progenitor, all early priming stage clusters were re-clustered. The common intermediate progenitor stage was not observed among myeloid lineages, indicating that CMPs were absent during lineage commitment (Figure S2). The original MPCs were instantaneously primed to generate lymphoid, neutrophil, and megakaryocytic–erythroid progenitor cell types, and they subsequently matured into different hematopoietic lineages (Figure S2A). The self-renewal capacity of MPCs is generally reduced when they undergo differentiation into the subsequent stages. Based on the successful capture of the entire spectrum of adult peripheral blood progenitor cells, including those in mobilized and quiescent progenitor states, we generated the transcriptional atlas of adult peripheral blood progenitor cells.

To redefine the HPC subgroups, we analyzed the characteristics of various progenitor cell subtypes, revealing the selected unique and specific marker genes and key TFs of progenitor cells in each lineage (Table 1 and Figure 2F). Additionally, negatively expressed genes were listed to distinguish each similar or identical redefined subgroup (Table 1). The transcriptional atlases of clusters 3 and 0 (MPCs) were remarkably comparable, with cluster 3 (C3) presenting an early tendency toward lymphoid progenitor cells. We redefined the three MPC-primed intermediate

directions of the CLPs, *GATA* gene-controlled progenitors (GAPs), and neutrophilic and monocyte progenitors (NMPs), which could not be characterized using CD38, FLT3, and KIT (Figure 3A and Table 1). Most of the top five specific marker genes for each progenitor lineage were canonical marker genes of classically defined hemopoietic cells, indicating a reliable subgroup characterization (Figure 2F).

Reforming the hematopoietic hierarchy

Based on the reliable redefinition and precise characteristics of the entire spectrum of progenitors, a new hematopoietic hierarchy roadmap was reformed (Figure 3). During the differentiation of the megakaryocytic–erythroid lineages, unique genes, such as *GATA2*, *CPA3*, *CNRIP1*, and *SLC40A1*, gradually accumulated. The GAP lineages differentiated independently of the other two branches. The shape of each branch was controlled by the expression of key regulated genes, which functioned similarly to a gradual activation switch. During the myeloid progenitor differentiation process, GATA2 and CSF3R regulated the formation of the first branches formulation step (Figure 3), followed by the downregulation of *CSF3R* expression and upregulation of *GATA2* expression, implying that the progenitors were primed to generate oligo-potent progenitor GAPs. The expression of GATA1 and KLF1 was presented and gradually upregulated at later mature stages, defined as megakaryocytic–erythroid progenitors (MEPs) (Figure 2B). CSF3R expression was upregulated in the original MPCs and nearly absent in the later stages of MEPs. Cytokine receptors were co-expressed or co-silenced with relevant key lineage regulators, indicating that multiple genes regulated the progenitor lineage-priming process. The enhanced differentiation of MEPs resulted in the expression of the second branch of mature MEPs. Toward the final maturation stages of MEPs, KIT, CDK4, GATA1, KLF1, or LMO4 controlled the differentiation of progenitors to megakaryocyte-erythrocyte (Pro-ME), or mast cell and basophil progenitors (MBPs). KIT, MS4A2, and TPSB2 expression levels were upregulated or uniquely expressed by MBPs, indicating that this cluster comprised MBPs (Figure S1A). GATA2 and KLF1 expression was restrained when the progenitors were primed to generate the

other myeloid progenitor branch of NMPs (Figures 3A and B). Further, the expression of *GATA2* and *CSF3R* genes marked the boundary for GAPs and NMPs.

Our study revealed that the original progenitors of basophils, eosinophils, and neutrophils differed. The continuous differentiation procedure of neutrophils is the shortest in the hematopoietic hierarchy. The neutrophilic progenitor is directly priming from MPC and has only two stages of the maturing process (Figure 3C). However, the identity of the original progenitor of eosinophils was not confirmed. The original progenitor cells were primed and differentiated into various progenitors through gradual activation/deactivation followed by TF accumulation, synergism with key receptor genes, and signal activation. MEP-representative genes gradually accumulated during differentiation, revealing their distinct continuous differentiation characteristics (Figure 3C). These findings suggest that the hierarchical organization of megakaryocytic–erythroid lineages in hematopoiesis is a continuous process, controlled by the activation and suppression of fate decision-associated genes.

During priming of the lymphoid progenitor branch, expression of the critical TFs, including HOXA9 and HOPX, was up-regulated (Figure 2C). Similarly, marker genes of lymphocytic lineages, such as *MS4A1*, *GNLY*, and *GATA3*, were specifically expressed, their expression was uniquely presented in specific stages and upregulated in a stepwise manner. Certain genes were prominently and specifically co-expressed in lineage-restricted progenitors and CLPs: CLPs expressed *IGHM*, *FCMR*, *BLNK*, and *LTB*; B progenitors expressed *CD79A* and *JCHAIN*; monocyte progenitors expressed *SPI1*, *FOS*, *PECAM1*, and *LST1*; NK progenitors expressed *NKG7* and *GZMA*. The expression characterization of these genes also indicated the underlying mechanisms of lymphoid lineage commitment.

The lymphoid lineage commitment process is likely a discrete stepwise process, with distinct lineage-committed populations during hematopoiesis. This was also observed for monocyte lineage differentiation into macrophages and dendritic cell (DC) progenitors. We also identified the co-expression of ETV6, LYL1, and NFE2 in NMPs and monocyte lineage-restricted progenitors, which was like that of SPI1, FOS, PECAM1, and LST1 co-expression in monocytes and CLPs (Figures 2C and 3B). The

differentiation of each progenitor stage was controlled by the expression of key regulatory genes, which together functioned as a stepwise activation switch. By switching TF expression in a stepwise manner, which generally showed a synergistic relationship with key receptor gene expression and signal activation, the original progenitor cells were primed and assigned to the next stages of lymphoid lineage progenitors (Figure 3).

Discovery of fate decision-associated TFs during lineage commitment

Several fate decision-associated factors were identified, including NFE2, LYL1, MYB, CDK4, ETV6, TESPA1, ZBTB16, and GATA2 of myeloid progenitors and RUNX2, HOXA9, JUN/FOS, TCF4, DDIT4, HOPX, KLF10, HOXA3, and TSC22D1 of lymphoid progenitors (Figure 3B). Our data revealed that RUNX1 monitors the differentiation of most progenitors (Figure S3), whereas RUNX2 mainly controls the differentiation of lymphoid progenitors. Importantly, these key factors were determined to have a contrasting functional role during HPC differentiation. HPCs differentiated into diverse myeloid progenitors by upregulating the expression of NFE2, LYL1, MYB, CDK4, ETV6, TESPA1, and GATA2 and through restraining the expression of RUNX2, HOXA9, TCF4, DDIT4, HOPX, KLF10, HOXA3, and TSC22D1. Together with the activation of other receptors and unique TFs, such as CSF3R for neutrophils and GATA2 for GAPs, in each progenitor lineage, the entire hematopoietic lineage commitment process was precisely controlled and gradually occurred (Table 1 and Figure 3C).

To confirm the essential genes involved in maintaining HPC states and their differentiation, we compared expression levels based on three single-cell datasets, including bone marrow cells, control peripheral blood mononuclear cells (PBMCs), and $CD34^+$ cell subsets. We found that ITGA4 (CD49d) was highly expressed in most progenitors, and this is important for the molecular regulation of HSC homing [21]. HSC functional marker genes were broadly expressed in most progenitor cells but were not specific, and these included *PTPRC*, *HMGB1*, *CD74*, *CD37*, *PFN1*, *TXN*, *ZFP36L2*, *SERPINB1*, *TMSB4X*, *SMIM24*, *MYADM*, and *HSP90AA1* (Table 1 and

Figure 1E). As most of these genes were confirmed via functional validation, our data demonstrated that the highly expressed genes in $CD34^+$ progenitor subsets were essential for the progenitor state and differentiation. These genes mediate the self-renewal, differentiation, mobilization, and maintenance of HPCs. Additionally, two new genes, *PPIA* and *EIF1*, were identified, similar to previously reported genes that exhibit pan-expression in progenitors and are highly expressed in mature subgroups of peripheral blood cells.

CDK6, *SOX4*, *CD82*, *EGFL7*, *SPINT2*, *IMPDH2*, *TSC22D1*, *CYTL1*, *FOXP1*, *SPI1* [6], *LY6E (SCA-2)*, and *ETV6* were also widely expressed in most progenitors and progenitor-differentiated cells, however, were relatively absent in all other cell types (Table 1). These genes were more highly or specifically expressed in the bone marrow and our identified progenitor cells than in mature PBMCs (Figure S4), highlighting their significance in regulating genes in the HPC state for maintenance and differentiation. Moreover, *ERG*, *BCL11A*, *LMO2*, and *RUNX1* were widely expressed in approximately 50% of the cells in each progenitor subgroup (Figure S3).

Dynamics of progenitor cell differentiation

To confirm the differentiation trajectory of the progenitor cells, pseudo-time analysis was performed using Monocle3, which revealed three distinct differentiation directions of the progenitor cells. The results clearly showed that along the trajectory of progenitor differentiation, MPCs were gradually primed to generate three separate progenitor cell branches and subsequently transitioned into diverse lineage stages (Figure 4A). However, CMPs were not identified during the lineage-priming process. Additionally, the pseudo-time analysis revealed that lineage signature genes, such as *KLF1*, *GATA1*, and *CSF1R*, exhibited unique expression patterns along the differentiation trajectory, and were separated from those of other lineages. The expression trajectory of signature genes was highly consistent with the differentiation trajectory of the lineages (Figure 4B).

Similarly, the inference of the single-cell regulatory network and clustering assessment based on the activity of TFs and regulons showed that the activation and

expression of key TFs differed among different progenitor subgroups (Figure 4C). The original progenitor of MPCs exhibited activated RFX2, whereas the TFs MAZ, HOXA9, KLF3, STAT1, JUND, POLR2A, and KLF13 were identified as key regulators of CLP priming. Several TFs, including SF1, RUNX1, ZMIM1, and BCLAF1, were identified in eosinophilic progenitors. Moreover, GATA1, MYC, SOX4, and HMGB1 were found to play a critical role in the progression of megakaryocytic–erythroid lineages (Figure 4C). XBP1 and IRF7 were activated by plasma progenitor cells, and IRF8, CEBPD, JUNB/FOSB, SPI1, KLF2, and CEBPB were the TFs that determined fate decisions during monocytic lineage differentiation. The characteristics of TF activation were similar among the same original lineages, including megakaryocytic–erythroid and monocyte lineages (Figure 4C). Unsupervised dimensionality reduction analysis showed that the two nearby progenitor CLPs and eosinophilic clusters had similar TF activation characteristics, indicating that they might have been primed by similar original cells (C3). Therefore, combining multi-omics and epigenetic analyses using the scRNA assay for transposase-accessible chromatin will provide strong evidence for the HPC-fate-decision process. Our results showed highly consistent TF expression and activation patterns throughout the lineage-priming process, reflecting the synergistic effects of the suppression and activation of key regulators. These findings suggest that the original progenitor cells can be primed and differentiate into various progenitors during hematopoietic lineage commitment.

Pathways associated with each subgroup

Diverse pathways affect and regulate HPC differentiation. Our analysis showed that most progenitors were enriched in and highly expressed *JUN/FOS (AP-1)*, *JUNB*, and *JUND* (Figures 2C and 3B). This indicates that the JUN-related signaling pathway is critical for the differentiation, mobilization, and maintenance of HPCs. The *DEPTOR* gene was highly expressed by MEP clusters (Figure S3), and it acts as a negative regulator of the mTORC1 and mTORC2 signaling pathways [22], indicating that the mTOR signaling pathway should be suppressed during MEP maturation and

differentiation. NFKBIA was widely expressed by >80% of the progenitor cells, whereas NFKBIZ was highly expressed in progenitors of monocytes and macrophages (Figure S3). NFKBIZ is a member of the NF-κB inhibitor family [23], suggesting that the NF-κB signaling pathway is suppressed in progenitors, particularly during monocyte and macrophage maturation and differentiation.

Gene set enrichment analysis of clusters using gene set variation analysis revealed the pathways and signals that affect the differentiation of each branch and progenitor subgroup (Figure 4D). The transforming growth factor-beta (TGF-β) signaling pathway was enriched in the original progenitor cells (C0, MPCs), demonstrating its importance in maintaining the quiescent state of HPCs [24]. The enriched gene sets differed among myeloid progenitors (C2, C5, C6, and C13) and lymphoid progenitors. Moreover, the Notch, KRAS, and TNF-α signaling pathways were enriched, indicating their likely crucial role in maintaining monocyte progenitors (Figure 4D). Other studies have reported the novel role of Notch in regulating monocyte survival during hematopoiesis [25]. Heme metabolism and hedgehog pathway gene sets were enriched in erythroid progenitors. The activation of TGF-β and the hedgehog pathway was gradually decreased, but oxidative phosphorylation was enriched during the maturation stage of erythroid progenitors. Hallmark cell proliferation-related gene sets (G2M checkpoint, E2F targets, MYC targets V1 and V2, and p53 pathway), defined by the molecular signature database, were highly enriched among erythroid and neutrophilic progenitors. (Figure 4D). That meanings erythroid (C6) and neutrophil (C5) progenitors are already in S stage of cell cycle, indicating those cells have finish differentiation and tend to proliferation. However, the TNF signaling gene set was suppressed (Figure 4D). The Wnt/β-catenin signaling pathway was enriched in NK and T cell progenitors, and was reported to regulate aspects of peripheral T cell activation and migration [26]. Moreover, interferon response gene sets were enriched in B cells and monocyte progenitors.

Fate decision-associated genes of each redefined subgroup

Advanced analysis was performed to identify fate decision-associated genes,

particularly those encoding TFs (Figure 5A). Six genes (*SPINK2*, *HOPX*, *C1QTNF4*, *CSF3R*, *NPDC1*, and *SELL*) were highly expressed by MPCs, and their expression was gradually restrained during differentiation from GAPs to MEPs (Figure 5B). CSF3R, S100A10, and C1QTNF4 expression levels were gradually downregulated during the differentiation of early lymphoid progenitors (C4 cluster, Figure 5B). We also identified genes with the opposite expression pattern that were specific among megakaryocytic–erythroid lineages (C2, C6, and C13) and other progenitors, such as *SPINK2*, *HOPX*, *C1QTNF4*, *FCER1A*, *GATA1*, and *SLC40A1* (Figures 5A and B). This indicated that the fate decisions among megakaryocytic–erythroid lineages (C2, C6, and C13) and other progenitors might be controlled by those genes during lineage commitment. The hemoglobin subunit delta was determined to be the early protein involved in erythrocytic development. Mini-chromosome maintenance protein genes, including *MCM2*, *MCM6*, *MCM5*, *MCM4*, *MCM3*, and *MCM7*, which encode components of the MCM2–7 complex, were highly expressed by Pro-ME and NMPs (Table 1). The MCM complex was found to be the replicative helicase that play an essential role in eukaryotic DNA replication licensing system. That was consistent with the genes sets analysis result (Figure 4D), which showed the status of the two progenitors may switch from differentiation to proliferation. *LMO4* was the most specific TF-encoding gene in MBPs (Figure 3B).

Three key TFs (GATA3, TCF7, and DDIT4) were identified in T progenitor cells (Figure 5C), particularly two novel genes, *GATA3* and *TCF7*. These three TFs are critical for fate decisions during T-cell lineage differentiation (Table 1 and Figure 3B). Previous studies have considered TCF7 and GATA3 critical TFs during T-cell development and differentiation [27, 28], which is highly consistent with our redefined characterization of T progenitor cells. We observed that LTB was widely expressed by CLPs, Pro-T, and Pro-B progenitor cells and that its expression was upregulated in the later stages of lymphoid progenitor differentiation. Two distinct types of B progenitor cells were identified and redefined as expressing different fate decision-associated TFs and marker genes in C9 and C12 clusters. Two factors, PRDM1 and IRF4, were specifically expressed in cluster C9, whereas PAX5 was

specifically expressed in cluster 12 (Table 1). *XBP1* was widely expressed in both clusters, and upregulated in cluster 9. This indicates that the development and differentiation of B cells occur in the early progenitor stages. Cluster 9 comprised the progenitors of plasma cells, whereas cluster 12 represented classical B progenitor cells (Table 1). Plasma cells are believed to represent the final stage of B cell proliferation and are incapable of further proliferation [29]. However, our single-cell data confirmed the existence of plasma progenitor cells in the whole blood samples. Immunoglobulin variable (IGV) genes were mainly expressed by the progenitors of cluster 9 cells, and the rearrangements in the *IGHV*, *IGKV*, and *IGLV* genes had already occurred during the progenitor state of plasma cells. These four TFs play a critical role in the fate decision of B lineage cells. HOPX was very highly expressed in the progenitor and mature NK cells but negatively expressed in the MEP lineages (C2, C6, and C13), and was determined to be one of the most important TFs for NK cells (Figures 3B and 5A; Table 1).

Two types of monocyte progenitor cells in the C7 and C14 clusters were identified and redefined. One of the most highly expressed genes in monocyte lineages was *IFI30*, which is related to the interferon-gamma signaling pathway. SPI1 and LST1 were widely expressed in most progenitors but were highly expressed in the monocyte subgroups (C7), and also highly expressed in the mature states of original monocyte cells, such as macrophages and DCs (data not shown). This suggests that *SPI1* and *LST1* are critical genes for monocyte progenitors. The monocyte progenitors of C7 and C14 were controlled by various fate decision-associated TFs, including KLF4, EGR1, MAFB, IRF8, and DDIT4 (Table 1 and Figure 5B). The two clusters, C7 and C14, included $CD86^+$ and $CD68^+$ cells. The C7 cells expressed high levels of EGR1, CEBPB, and MAFB but were negative for IRF8 and DDIT4 expression. Furthermore, C7 included $FCGR3A^+$ cells and showed high levels of CSF1R and TNF expression, indicating that they are more likely to be macrophage progenitors. The monocyte progenitor C14 positively expressed CLEC9A, ANPEP, and THBD (Table 1), indicating that this cluster is more likely to differentiate into DCs. Overall, the expression characteristics of fate decision-associated TFs in lymphoid, monocyte, and

MEP lineages were highly consistent with the redefinition of progenitors. Fate decision-associated TFs in each branch of the hematopoietic hierarchy were clearly recovered.

Functional validation and primary culture confirm the reliability of the redefinition of HPCs

The functional validation of HSC fate decision-associated genes has been performed in previous studies. We hypothesized that the phenotypic results of functional studies would be consistent with or explained by our HPC definitions. Furthermore, the reformed model of the hierarchical organization of hematopoiesis can be used to interpret conflicting or confusing results. KLF1 and GATA1 are key fate decision-associated TFs involved in erythroid and megakaryocytic differentiation [30, 31]. Functional studies have already confirmed that GATA1 progenitors generate mast cells, basophils, megakaryocytes, and erythroid cells [32], and their absence results in the production of monocytes, neutrophils, and lymphocytes [30]. Our transcriptional atlas revealed that GATA1 was uniquely expressed in the MEP lineage but negatively expressed in the monocyte, neutrophil, and lymphocyte progenitors (Figure 2B). These expression characteristics explain the findings of the functional study in that the absence of GATA1 will halt MEP lineage differentiation. Conversely, the absence of GATA1 causes progenitor cells to differentiate into NMPs and monocytes. In primary cultures, HPCs differentiated into erythroid cells and mast cells under the same *in vitro* conditions, utilizing IL3 and (SCF); this indicates that the two types of cells originate from the same lineage. These results confirmed that the reformed hematopoietic hierarchy of the MEP lineage was reliable.

Macrophage colony-stimulating factor (M-CSF) is critical for macrophage survival during *in vitro* primary cultures [33]. This is consistent with our transcriptional atlas where the M-CSF receptor, CSF1R, was only expressed in monocyte progenitors and matured macrophages (Figure 5B). Additionally, stem cell growth factor (SCGF), also known as CLEC11A, is a secreted glycoprotein that exhibits erythroid burst-promoting activity in the absence of the SCF/c-kit ligand [34,

35]. Our results showed that SCGF was highly expressed in NMPs and MEPs, whereas CSF3R, a G-CSF receptor, was highly expressed in NMPs. SCGF cooperates with GM-CSF or G-CSF and enhances the growth of granulocyte-macrophage colonies; however, if only SCGF is present in culture, erythroid-burst promoting activity occurs [34, 35]. The expression characteristics of CSF3R and G-CSF strongly suggest that SCGF with G-CSF activate and induce the differentiation of neutrophil and macrophage progenitors, suppressing MEP lineage generation. The IL15 receptor IL2RB expression in NK progenitors was unique (Figure 5D). The common beta chain receptor CSF2RB of IL-3, IL-5, and GM-CSF, was unique in erythroid lineage progenitors. The expression pattern of IL3RA and CSF2RB in our atlas revealed that the IL3 can non-specifically bind to the two receptors, which induces the progenitor differentiation into other types of lineages. That validated the expression pattern of the megakaryocytic–erythroid lineages unique markers. In addition, IL7R, an essential driver of T-cell development and homeostasis, was solely expressed in T-cell progenitors, whereas *IL2RG* and *IL1B* were expressed across most progenitor subgroups (Figure 5D). IL18 was identified as one of the most essential cytokines in early progenitors (data not shown), and our study indicated that this cytokine had a critical role in hematopoiesis. The transcription of *IL1B* was regulated by the TFs *SPI1* and *CEBPB* [36], which were highly expressed in monocyte progenitors (Table 1). Moreover, *ALDH1A1* was highly expressed in erythroid-lineage progenitors (Figure S5), and its product was previously shown to participate in retinoic acid synthesis. Retinoic acid is required for progenitor cell maintenance[37]. The expression characterization of these cytokine receptors and genes explains why the cytokine ligands are essential for cell survival and proliferation in *in vitro* primary cultures. In summary, the expression characteristics of fate decision-associated TFs and cytokine receptors provide evidence that the redefinition of progenitor lineages is verifiable, reliable, and reasonable (Figure 3B).

Are the definition and hierarchy reliable?

In vivo or *in vitro* functional studies can be affected by various conditions, which

can’t accurately reflect the true physiological state of original progenitors. And colony forming assay also exhibits obviously defects due to variations in cytokines, samples, and selection methods (see introduction). Most scRNA studies that enrich bone marrow progenitors using flow cytometry enrichment demonstrate inefficient reproducibility, false positives, and subjectivity definitions [4, 5]. Some CD34-positive cells are transcriptionally negative, and subsets are either missing or inconsistently distribution among samples. Fate mapping is nearly impossible for adult original progenitors. Here, we present three aspects of evidence supporting the reliability of our definition and hierarchy.

Robustly Validating the Definition by a Reliable Atlas

Here, we utilized original progenitors from adult peripheral blood, which faithfully reflect the physiological state of human progenitors. We comprehensively captured the entire differentiation stages of these physiological progenitors. The substantial numbers of cells in each progenitor subset provided a reliable transcriptional atlas (Table S1). By utilizing unsupervised data analysis, we avoided bias and subjective effects on results. The subset distribution reproducibility across sample is exemplary (Figure 2D). Furthermore, the expression patterns of some unique genes in each redefined subset are highly consistent with conventional markers (Table 1). Based on truly physiological original progenitors and a reliable atlas, the scRNA data robustly validates the definition, and is more reliable than definition based on *in vivo* or *in vitro* experiments.

Straightforward Results Confirm the Reliability of the Hierarchy

Our T-SNE plot clearly and perfectly displayed the three differentiation directions of progenitors after the initial priming step of MPC (Figure 3A). The different stages of the megakaryocytic–erythroid lineages are distinctly distributed in the same directions, expressing key well-known fate decision genes (GATA1,KLF1,TAL1) (Figure 2B).The unique expression characteristics of these well-known genes (GATA1,KLF1,TAL1) confirmed the definition of MEP lineages, and we are the first

to identify and define the intermediate stage of GAP(Figure 3A). The subsets of the megakaryocytic–erythroid lineages are adjacent and lack clear boundaries, indicating continuous differentiation trajectories. The gradual and continuous expression characteristics of other well-known genes (GATA2,SLC40A1) also provide robust evidence of the continuous lineage commitment. Furthermore, the unique expression patterns of well-known receptor genes (EPO, MPL, CSF2RB) confirm the definition of MEP lineages, which is consistent with *in vitro* culture. The termination of MEP lineages differentiation yields two type of cells (mast and erythliod), which already validated by numerous studies that mast and erythroid cell are generated under the same culture conditions, and further validation is needn't.

Physiological Evidence

The megakaryocytic–erythroid lineage commitment process is continuous, aligning with the constant production of erythroid cells every second in the blood. While the commitment process in the lymphoid lineage is discrete and stepwise, reflecting the obviously increase in numbers and heightened activity of lymphoid cells specifically in response to infections or pathological conditions (adaptive immunity). The continuous differentiation process of neutrophils is the shortest in the hematopoietic hierarchy (only two stages) (Figure 3A, C), which also matching with physiological phenomenon. Neutrophils are the primary and rapid responders during infections of innate immunity and can quickly generate an enormous number of mature neutrophils for immune defense [38]. The shortest differentiation trajectories of neutrophils enable a rapid and robust response against infection. In summary, the reformed hierarchy of hematopoiesis perfectly explains the physiological phenomenon of immunity. Often, the truth is simple, yet we have been confused so long time.

Discussion

Overall, we redefined the characteristics of original progenitor lineages and

identified specific fate decision-associated TFs and marker genes critical for hematopoiesis. Our study further reformed and established a comprehensive hierarchy of hematopoiesis that reconciled previous models and explained conflicts among functional studies. This provides detailed insights into the lineage of HPCs, revealing the molecular mechanism underlying their self-renewal and multi-potent differentiation. Our single-cell transcriptional atlas provides an integrated landscape of the heterogeneity, development, and hierarchy of adult peripheral blood progenitor cells. Collectively, this study makes significant breakthroughs in the fundamental discipline of hematopoiesis, disciplinary challenges that have plagued the field for more than half a century.

Currently, no standard consensus for defining all HPC lineages exists, owing to the inconsistent classification, with conflicting functional outcomes, in the current literature [11, 15]. In our study, nearly the entire spectrum of HPCs was successfully enriched in adult peripheral blood. This will be particularly attractive for large single-cell studies to establish a comprehensive hierarchy of hematopoiesis with a clear HPC definition and reliable lineage commitment. We identified progenitor subgroups using single-cell sequencing, and these were inconsistent with the classical definition by specific surface or characteristic marker genes, such as *GATA1*, *MPO*, and *KLF1*. Similarly, we demonstrated that classical surface markers, including CD38, KIT, THY1, CD49f, and FLT3, were non-specific and inefficient when identifying HPCs. Our data showed that none of the MME-positive cells co-expressed CD38, suggesting that classifying lymphocytes using $CD38^+$ and MME^+ phenotypes is insufficient[4, 5]. This indicates that the conventional lineage definition of HSCs using the aforementioned markers is imperfect and should be improved. All hematopoietic progenitors were redefined using unique markers and key distinguishable TF-encoding genes. Some unique marker genes of progenitor subgroups were highly consistent with previous characterizations, such as *KLF1*, *HPGDS*, and *MPO* of MEPs, mast cells, and neutrophils, respectively [17, 39, 40]. Combining TFs with unique marker genes would be highly efficient for HPC lineage definitions. For the first time, we clearly defined most HPC lineages using

unsupervised and inartificial clustering and provided an unambiguous transcription atlas of each HPC lineage type.

Our transcriptional atlas revealed the major fate decision-associated factors at each level of the hematopoietic hierarchy. Progenitor cells are primed via the activation of key TFs, which drives progenitor differentiation into downstream-specific mature lineages [41]. The expression of TFs in the original and mature stages of progenitors revealed a basic split between myeloid–lymphocyte and mast–erythroid development. In the later stage, MEPs differentiate in two directions owing to changes in the expression of three key TFs (KLF1, LMO4, and GATA1). Our findings highlight the importance of TFs and the characterization of genes involved in hematopoietic lineage fate decisions, providing additional insights into the molecular mechanisms underlying hematopoietic lineage commitment. The expression characteristics of fate decision-associated factors provide an accurate interpretation of the mechanism underlying hematopoietic hierarchy, which can be used to resolve the conflicts among *in vitro* and *in vivo* progenitor cell studies [30]. The expression characteristics of known fate decision-associated TFs, such as NFE2, TAL1, and LYL1 [42-44], provide evidence that the redefinition of progenitor lineages is verifiable, reliable, and reasonable. A previous study showed that transcriptional programs are not strictly lineage-affiliated, and expression changes between HSCs and early progenitors occur independently of the lineage selection [45]. This conclusion might be attributed to the unreliable definitions of the subgroups and different stages of progenitors during lineage commitment. Based on a reasonable classification and the redefinition of HPCs, our study found that the expression of unique TFs is strictly lineage-affiliated. Induced pluripotent stem cells were previously established by activating four transcription factors, and the reprogramming of pluripotent stem cells from various somatic cells has been widely investigated [46]. For example, GATA1 and TAL1 have been validated in reprogramming megakaryocytes and erythroid cells [47, 48]. These fate decision-associated TFs of HPCs, identified using our atlas, are crucial for reprogramming progenitor cells. The artificial and precise commitment to progenitor cell differentiation can be achieved by

controlling TF and receptor genes. Moreover, this is the first study to clearly identify the cell fate decision-associated TFs in each branch during the formation of hematopoietic hierarchy in adult peripheral blood. Precise control of the priming and differentiation of HPCs can subsequently be interfered based on the key TFs and receptor signals, which we refer to as the Artificial Operation of HPCs.

CSF3R, the G-CSF receptor [49], was mainly expressed in the original clusters (MPCs). However, the molecular mechanisms underlying the G-CSF induced-enhancement of hematopoiesis remain unclear. Our results provide a new hierarchy of hematopoiesis based on the scRNA analysis of unperturbed and G-CSF-mobilized progenitor lineages in peripheral blood. Here, the abundance of all progenitors, particularly MPCs, remarkably increased. G-CSF directly stimulated HPC downstream activation signals by binding to CSF3R, which resulted in the conversion of quiescent-state HPCs into activated progenitors. After the mobilization procedures, HPCs were primed and differentiated into all types of progenitors, which further differentiated into mature myeloid and lymphocyte cells. The key TFs and cytokine regulators synergistically controlled the hematopoietic lineage commitment, validated using primary culture-induced lineage formation and reprogramming studies. The mechanisms underlying lineage commitment raise certain questions. How are the key TFs activated during the lineage commitment process? Are the TFs activated independently (internal) or via the cytokine signaling cascade (external)? After the first step of HPC priming with G-CSF, which factors regulate the activation of lineage fate decision-associated factors (TFs and cytokine signaling cascade) and which determine the follow-up differentiation direction? Generally, the environment and epigenetics also play a key role.

The following three aspects strongly support the definition and classification of HPCs as distinctly verifiable, reliable, and reasonable. First, the unique markers were highly consistent with the characteristics of most progenitor subgroups. Second, the definition addresses the conflicting results associated with functional studies. Third, the expression of distinguished cytokine receptors perfectly explains the phenomenon observed in the primary culture of progenitor subgroups, such as CSF3R and CSF1R

for neutrophil and monocyte cultures, respectively [50].

Recent studies have reported that HSC lineage commitment is a continuous process [5, 14], which challenge the classical stepwise hierarchy of hematopoiesis[51]. However, based on HSC heterogeneity and the conflicting definitions of HSC subgroups, the mechanism underlying this hierarchy is enigmatic, and the established hematopoietic hierarchy are inconsistent and incompatible. This might be due to the limitations associated with the number of captured cells and the inefficient definition based on surface markers. Based on our clear definition and atlas of HPCs, a gradient of hematopoietic progenitors exists during lineage commitment, which can be distinguished by the expression of unique TFs. Bias during HPC differentiation gradually accumulated through the activation of multiple lineage-specifying TFs, such as NFE2, LYL1, MYB, and GATA2 of the myeloid cells. Throughout this process, individual progenitors gradually acquired lineage biases in multiple directions during homeostasis. Our atlas results showed that the key transcriptional factors acted synergistically with cytokine receptors to prime HPCs for commitment to distinct downstream lineages, resulting in full lineage commitment. Our findings will facilitate mapping the entire transcriptional signaling and pathways involved in lineage commitment and provide strong evidence of a heterogeneity gradient among lymphoid and monocyte lineage progenitors. The expression of key TFs could be considered a relatively reliable boundary between stages during progenitor maturation. The clear definition and characteristic of HPCs showed stepwise and continuous process exists during lineage commitment, which can be distinguished by the expression of unique TFs and marker genes. The megakaryocytic–erythroid lineage commitment process is continuous, aligning with the constant production of erythroid cells every second in the blood. While the commitment process in the lymphoid lineage is discrete and stepwise, reflecting the remarkable increase in numbers and heightened adaptive immunity of lymphoid cells only in response to infections or pathological conditions. Based on the reliable redefinition and hierarchy, the clonal dynamics of clonal hematopoiesis will be clearly understood.

Our elaborate and comprehensive transcriptional atlas of HPCs provides accurate

definitions of the hierarchy of and biological insights into the hematopoiesis of progenitors. Our study further explains how diversified progenitor cells are primed, induced, and differentiated into all lineages of immune system cells; this enables the HPC transcriptional atlas to provide a novel perspective and adequate explanation of how the immune system is reconstituted in adults. Cytokine storms have been reported during severe infections, including coronavirus disease [52, 53], and our atlas demonstrates the expression of cytokine receptors in specific progenitors. Controlling the initial stage and generation of cytokines, particularly suppressing initial cytokine-mediated activation of the cascade process, could inhibit or reduce the effect of subsequent cascade activation. Technically, this would help to control the excessive activation of cytokines and prevent cytokine storms as a novel therapeutic strategy. Our reformed taxonomy will thus facilitate accurate preventative and treatment methods and improve immune monitoring in healthy and diseased states.

This study has certain limitations. First, our study did not define the original progenitors of HSCs, as the canonical markers CD49f and THY1 were rarely or negatively expressed. Granulocyte, neutrophilic, and monocyte progenitors were identified as originating from the same myeloid progenitors [15]. However, our results showed that the fate decision-associated TFs and characteristic genes between the two types of progenitors were remarkably different, implying that the original association between these two types of lineages should be reconsidered. Our study further showed that the expression patterns in CLPs were similar to those in monocyte lineage-restricted progenitors. We further speculated that monocyte progenitors could also originate from CLPs or have an independent origin. Although we identified the fate decision-associated factors of the two types of B progenitor cells and monocyte progenitors, the association between the two cell types remains unknown during the HPC differentiation process. Therefore, further investigations are required to determine whether the two types of progenitor cells were derived from the same or independent original cells. Second, our enrichment method might have caused the loss of some intermediate-state progenitors. Given that our entire atlas was established based on unsupervised dimensionality reduction and clustering from a statistical

approach, its accuracy and efficiency are imperfect. Artificial intelligence and deep learning could provide an accurate, reliable, and distinguishable definition of each state and all progenitor subgroups during hematopoiesis. Moreover, DNA barcoding and the clustered regularly interspaced short palindromic repeats (CRISPR)/ CRISPR-associated protein 9 (Cas9)-mediated genome editing system will prove helpful in evaluating the progenitor cell clonal dynamics during hematopoiesis.

In conclusion, our study reformed the comprehensive hierarchy of hematopoiesis and provided the key regulators involved in controlling the hematopoietic process in human adult peripheral blood. Our new model provides an accurate definition of a complete hierarchy of biological insights into the hematopoiesis and lineage commitment of HPCs, enabling new research in cell engineering, regenerative medicine, and disease treatment. The reformed lineage commitment results provide critical insights into and implications regarding the etiology and treatment of hematologic disorders. Moreover, this hematopoietic hierarchy could lead to various possibilities for artificial, precise, and controlled progenitor cell differentiation and the targeted treatment of immune diseases, providing hope for presently untreatable diseases.

Acknowledgments

We acknowledge Professor Chen Qiu, Rongchang Chen, and Yingyun Fu for supporting this research. Moreover, we are grateful to the subjects for their participation in this study. We would like to express our appreciation of the laboratory support provided by the Key Laboratory of Shenzhen Respiratory Diseases (ZDSYS201504301616234). We also thank Chi-Wai Man of the Chinese University of Hong Kong for manuscript revision. The study was supported by the National Natural Science Foundation of China (21807072).

Authors' contributions

Y.Y conceived the original idea, designed the study, and performed HPC-enrichment experiments. Y.Y and L.C conducted the bioinformatics and statistical analyses of the

sequencing data. Y.Y and Q.S wrote and revised the manuscript. All authors approved the final draft of the manuscript. Q.H, Y.F, and G.L conducted the sample collection, and S.C generated the schematic diagram. Y.Y and S.C conducted the primary culture and lab work. Thanks to the partly content publication of the study in the special supplement of Blood journal (2023.142. Supplement 1: 4057.) by the 65th ASH Annual Meeting & Exposition (ASH 2023).

Declaration of interests

The authors declare no competing interests. One PCT patents (PCT/CN2024/070855) related to this work by Dr. YJ YUE.

Declaration of Generative AI and AI-assisted technologies in the writing process

Not applicable

Figure Legends

Figure 1. Progenitor subtypes identified using unsupervised dimensionality reduction and clustering analysis in adult peripheral blood.

A, Schematic showing progenitor cell enrichment and sequencing workflow. B, Thirty-two integrated clusters identified via UMAP analysis. C, Expression feature plots of the progenitor markers CD34 and CYTL1. D, Violin plot of the newly identified and known progenitor cell marker genes *CDK6*, *EGFL7*, *CYTL1*, and *SOX4*. E, Dot plot illustrating the expression of progenitor cell-related marker genes in all clusters. F, Enrichment results of three signaling pathways of progenitor cells. G, Dot plot displaying the enriched categories in the GO analysis. GO, Gene Ontology; UMAP, uniform manifold approximation and projection

Figure 2. Characterization of progenitor subtypes in quiescent and activated states. A, UMAP illustrating 17 clusters of integrated $CD34^+$ progenitor cells. B, Violin plots of the expression of conventional surface markers and unique marker genes, showing that *GATA2*, *CSF3R*, and *MME* were more efficient and accurate

markers than the conventional surface markers, *KIT* and *FLT3*, for myeloid, granulocyte, and lymphoid progenitor classification. C, Dot plots showing the expression of representative marker and lineage-specific genes. D, UMAP showing the distribution of cells in the quiescent (Control) and activated (G-CSF) states, displaying the persistence of all types of progenitors in the quiescent state of adult peripheral blood. E, Column plot displaying the percent changes in all progenitor clusters. The primitive clusters 0 and 3 exhibited the most pronounced changes in cell numbers and fold changes. F, Volcano plot showing the top five genes with the highest fold-changes in progenitor clusters. UMAP, uniform manifold approximation and projection

Figure 3. Newly reformed hematopoietic hierarchy of hematopoietic progenitor cell differentiation. A, UMAP plots demonstrating the redefined progenitor clusters. B, Dot plots illustrating the expression patterns of representative fate decision-associated transcription factor genes in each subgroup of progenitors. C, Reformation of the lineage commitment model based on a scRNA transcriptional atlas. In the new hematopoietic hierarchy, the MPCs are simultaneously primed into three progenitor cell types, which subsequently mature into diverse lineages. The first progenitor cell type is the CLPs, which differentiate into T, B, and NK progenitor cells stepwise. The second progenitor cell type, comprising GAPs, continuously differentiates into the intermediate-stage MEPs first, controlled by GATA2 and GATA1. In the later stages, MEPs differentiate into two directions, controlled by KIT, KLF1, and LMO4. The high-expression branch differentiates into MBPs. The other branch differentiates into Pro-MEs. The third type of progenitor cells include NMPs, which differentiate into neutrophils and monocytes. Dashed lines represent partial lineage connections that lack strong direct evidence of differentiation relationships. The shaded area indicates the GATA2 and GATA1 expression domains. UMAP, uniform manifold approximation and projection; MPC, multi-potent progenitor cells; scRNA, single-cell RNA; GAPs, gene-controlled progenitors; CLPs, common lymphoid progenitors; MEPs, megakaryocytic–erythroid progenitors; MBPs, mast cell,

basophil, or eosinophil progenitors; Pro-MEs, megakaryocytic and erythroid pro-progenitors; NMPs, neutrophilic and monocytic progenitors

Figure 4. Dynamics of lineage signature genes and activities of lineage-priming transcription factor regulons. A, UMAP visualizing the differentiation trajectory of progenitor cells using pseudo-time analysis and Monocle 3. B, UMAP feature plots of transcriptional activities of lineage signature genes related to the differentiation trajectory. C, Heatmap of the differential activation of transcription factor regulons of all progenitor subgroups. D, Heatmap plot of gene set enrichment analysis results of GSVA clusters. UMAP, uniform manifold approximation and projection; GSVA, gene set variation analysis

Figure 5. Expression of marker genes and genes associated with fate decisions in redefined hematopoietic progenitor cells. A, Heatmap showing the expression of marker genes and genes associated with the fate decisions of redefined progenitor clusters. B, Violin plot illustrating the expression of feature marker genes, SPINK2, HOPX, C1QTNF4, CSF3R, NPDC1, and SELL, which were highly expressed in the original progenitors of HPCs but were restrained during the differentiation from GAPs to MEPs. C, Feature plots showing the expression of LTB, TCF7, and MAFB. D, Violin plots illustrating the expression of receptor genes, including CSF1R, IL7R, and IL2RB, which were specifically expressed in the redefined progenitor subgroups. GAP, genes-controlled progenitors; MEPs, megakaryocytic–erythroid progenitors; HPCs, hematopoietic progenitor cells. GAPs, gene-controlled progenitors; MEPs, megakaryocytic–erythroid progenitors; HPCs, hematopoietic progenitor cells

Table 1. Selected specific marker genes and genes associated with fate decisions in all redefined subgroups.

STAR Methods

Sample collection and HPC enrichment

Overall, 8 mL of peripheral blood was obtained from each donor. Before blood collection, mobilized donors received 300 μg of subcutaneous recombinant human G-CSF at 12 h intervals for 5 days. After 5 days of continuous treatment, 10 mL of peripheral blood was collected in heparinized tubes from the stem cell donor; PBMCs and total white blood cells were obtained according to standard protocols. The protocol for PBMC isolation was based on separation via density gradient centrifugation with Ficoll® (Cytiva, GE), a highly branched polysaccharide. Lineage (Lin−, CD3−CD19−CD56−)-negative cells were removed using STEMCELL reagents (15272). The self-designed method was applied to enrich the HPCs. Ethical approval was obtained from the ethics committee of the Shenzhen People's Hospital in adherence with the ethical standards recommended in the 1964 Declaration of Helsinki. Informed consent was obtained from each study participant.

ScRNA preparation and sequencing

RNA sequencing was performed on human enriched HPCs (Lin−) purified from three donors, including PBMC progenitor cells mobilized with G-CSF. A single-cell suspension was loaded, captured using chromium microfluidic chips, and barcoded with a 10× Chromium Controller (10X Genomics). Barcoded scRNA was reverse-transcribed, and sequencing libraries were constructed using reagents from Chromium Next GEM Single Cell 5′ reagent kits v2, according to the manufacturer's instructions (10X Genomics). Sequencing was performed using an Illumina Nova 6000 (PE150).

ScRNA sequencing (scRNA-seq) data processing

The raw sequencing reads of all samples were demultiplexed and aligned to the reference genome hg38 via the Cellranger pipeline. All downstream single-cell unsupervised dimensionality reduction and clustering functions were performed using Seurat [54]. GO enrichment analysis of marker genes was performed using the cluster

profiler R package. Furthermore, trajectory analysis was performed using Monocle 3 [55]. TF regulation inference and gene regulatory network analyses were performed using Single-Cell rEgulatory Network Inference and Clustering (SCENIC) [56]. Gene set variation enrichment analysis was performed using a non-parametric and unsupervised method of gene set variation analysis [57].

Integrated analysis with an online single-cell dataset

The online scRNA datasets GSE120221 [58], GSE117498 [4], GSE145531 [59], GSE181989 [60], and GSE96583 [61] were downloaded from the Genomic Spatial Event (GSE) database. Integrated analysis was performed using Seurat. The bone marrow scRNA datasets GSE120221, GSE117498, GSE145531, and GSE181989 were merged. Similarly, the PBMC scRNA datasets GSE120221 and GSE96583 were merged.

Cytokine-based primary culture of HPCs

From human cord blood, $CD34^+$ mononuclear cells were separated using Ficoll® (Cytiva, GE) and enriched using the CD34 Kit (StemCell, 15026). Purified $CD34^+$ cells were induced to generate megakaryocytic–erythroid-lineage cells through cytokine-based primary culture. Purified $CD34^+$ cells were cultured in StemSpan SFEM (StemCell) for 21 d at 37 °C with 5% fetal bovine serum under two different conditions, as follows: (1) 10 ng/ml human SCF (PeproTech), 20 ng/ml EPO (MCE), and 10 ng/ml IL3 (MCE); (2) 10 ng/ml human SCF, 20 ng/ml EPO, 10 ng/ml IL5 (MCE). Cultures were supplemented with 1× insulin–transferrin–selenium–ethanolamine (ProCell) and penicillin–streptomycin–glutamine (Life Technologies). After 2 weeks, human IL-3 and SCF were withdrawn from the culture medium, and enucleation culture was performed for each group.

Flow cytometric analysis of erythrocyte cells

Cultured cells were counted using a cytometer (Count Star) before and after culture.

Flow cytometric analysis of surface markers was conducted each week and was performed by staining cells with anti-human CD34-phycoerythrin (1:100), anti-human CD235a-fluorescein isothiocyanate, anti-human SLC40A1–Alexa Fluor 647, and anti-human CD71-adenomatous polyposis coli (BioLegend, 1:100) antibodies, followed by flow cytometric analysis (Beckman). Antibodies are described in Supplementary Table 2. The results were analyzed using CytExpert software.

References

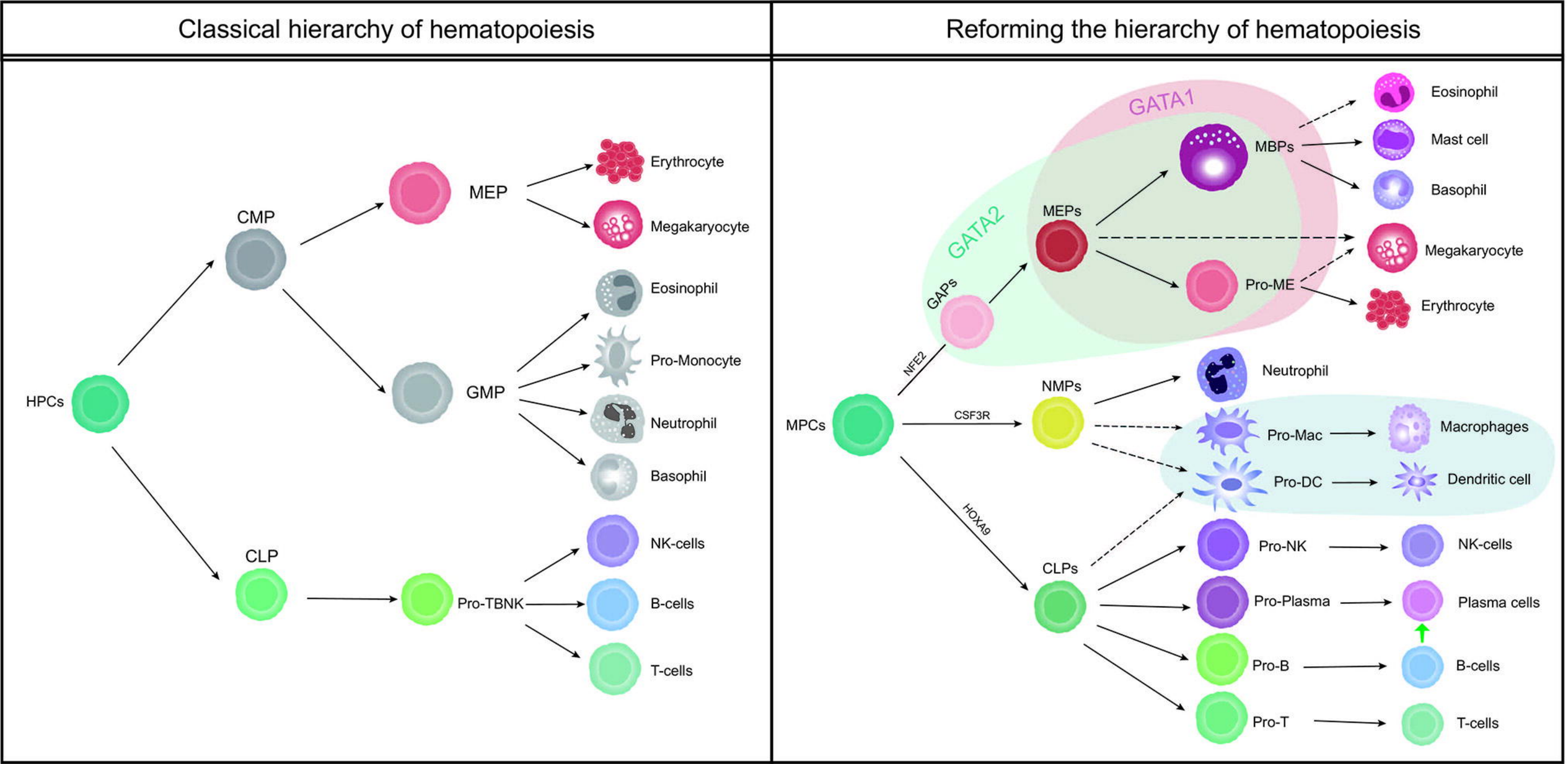
1. Cheng H, Zheng Z, Cheng T: **New paradigms on hematopoietic stem cell differentiation**. *Protein & cell* 2020, **11**(1):34-44.
2. Xie X, Liu M, Zhang Y, Wang B, Zhu C, Wang C, Li Q, Huo Y, Guo J, Xu C *et al*: **Single-cell transcriptomic landscape of human blood cells**. *Natl Sci Rev* 2021, **8**(3):nwaa180.
3. CD-REFRanzoni AM, Tangherloni A, Berest I, Riva SG, Myers B, Strzelecka PM, Xu J, Panada E, Mohorianu I, Zaugg JB *et al*: **Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis**. *Cell stem cell* 2021, **28**(3):472-487 e477.
4. Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, Biffi A, Klein AM, Biasco L: **A comprehensive single cell transcriptional landscape of human hematopoietic progenitors**. *Nature communications* 2019, **10**(1):2395.
5. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D *et al*: **Human haematopoietic stem cell lineage commitment is a continuous process**. *Nat Cell Biol* 2017, **19**(4):271-281.
6. Sugimura R, Jha DK, Han A, Soria-Valles C, da Rocha EL, Lu YF, Goettel JA, Serrao E, Rowe RG, Malleshaiah M *et al*: **Haematopoietic stem and progenitor cells from human pluripotent stem cells**. *Nature* 2017, **545**(7655):432-438.
7. Zhu J, Emerson SG: **Hematopoietic cytokines, transcription factors and lineage commitment**. *Oncogene* 2002, **21**(21):3295-3313.
8. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A *et al*: **Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors**. *Cell* 2015, **163**(7):1663-1677.
9. de Morree A, Rando TA: **Regulation of adult stem cell quiescence and its functions in the maintenance of tissue integrity**. *Nature Reviews Molecular Cell Biology* 2023, **24**(5):334-354.
10. Jacobsen SEW, Nerlov C: **Haematopoiesis in the era of advanced single-cell technologies**. *Nat Cell Biol* 2019, **21**(1):2-8.
11. Haas S, Trumpp A, Milsom MD: **Causes and Consequences of Hematopoietic Stem Cell Heterogeneity**. *Cell stem cell* 2018, **22**(5):627-638.
12. Manz MG, Miyamoto T, Akashi K, Weissman IL: **Prospective isolation of human clonogenic common myeloid progenitors**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(18):11872-11877.
13. Akashi K, Traver D, Miyamoto T, Weissman IL: **A clonogenic common myeloid progenitor that gives rise to all myeloid lineages**. *Nature* 2000, **404**(6774):193-197.
14. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY,

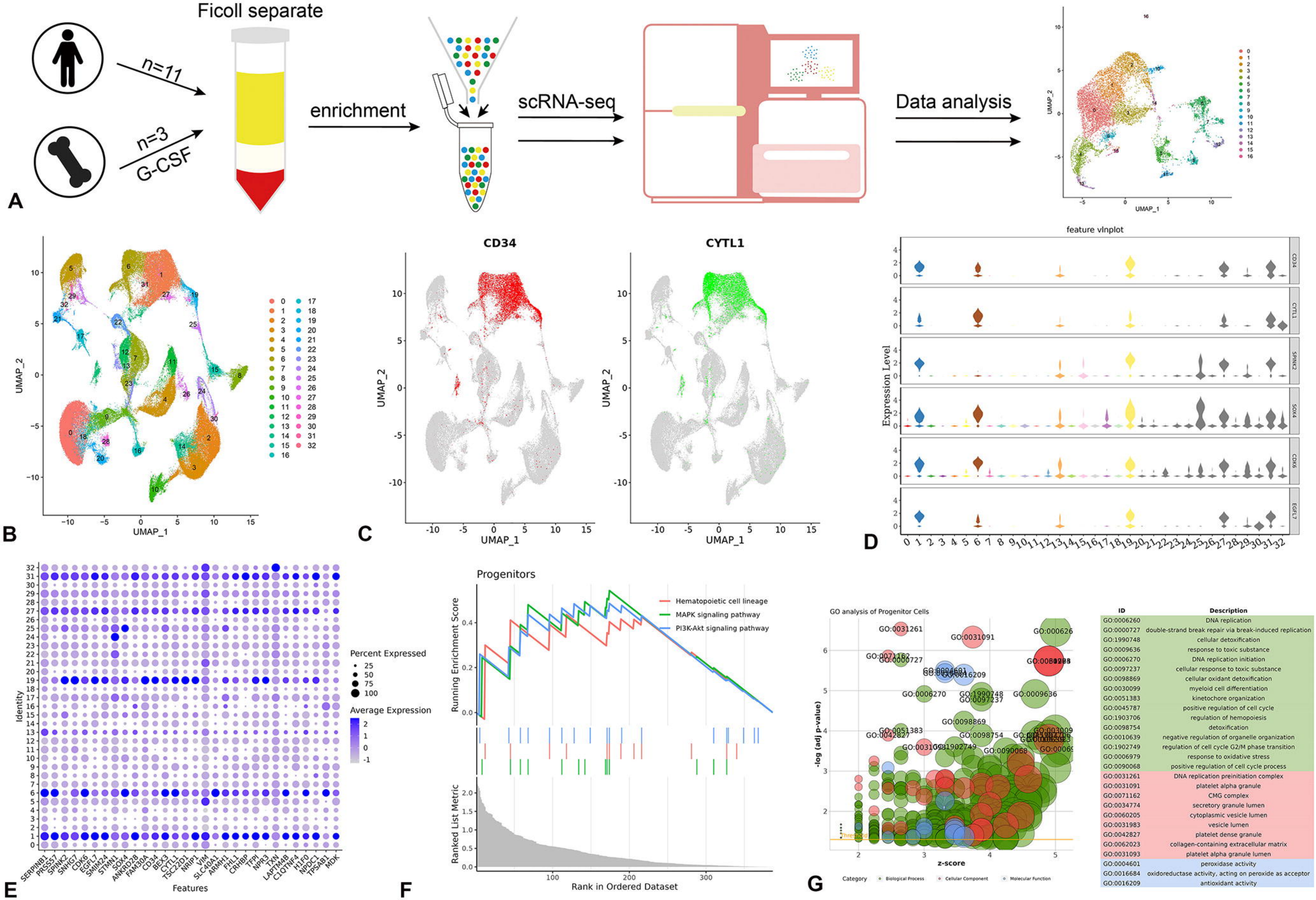
964 Greenleaf WJ: **Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of**
965 **Human Hematopoietic Differentiation**. *Cell* 2018, **173**(6):1535-1548 e1516.
966 15. Laurenti E, Gottgens B: **From haematopoietic stem cells to complex differentiation**
967 **landscapes**. *Nature* 2018, **553**(7689):418-426.
968 16. Beaudin AE, Boyer SW, Forsberg EC: **Flk2/Flt3 promotes both myeloid and lymphoid**
969 **development by expanding non–self-renewing multipotent hematopoietic progenitor cells**.
970 *Experimental hematology* 2014, **42**(3):218-229.e214.
971 17. Metzler KD, Fuchs TA, Nauseef WM, Reumaux D, Roesler J, Schulze I, Wahn V,
972 Papayannopoulos V, Zychlinsky A: **Myeloperoxidase is required for neutrophil extracellular**
973 **trap formation: implications for innate immunity**. *Blood* 2011, **117**(3):953-959.
974 18. Schultz J, Kaminker K: **Myeloperoxidase of the leucocyte of normal human blood. I. Content**
975 **and localization**. *Archives of Biochemistry and Biophysics* 1962, **96**(3):465-467.
976 19. Huang Y, Sitwala K, Bronstein J, Sanders D, Dandekar M, Collins C, Robertson G, MacDonald J,
977 Cezard T, Bilenky M *et al*: **Identification and characterization of Hoxa9 binding sites in**
978 **hematopoietic cells**. *Blood* 2012, **119**(2):388-398.
979 20. Hölig K, Kramer M, Kroschinsky F, Bornhäuser M, Mengling T, Schmidt AH, Rutt C, Ehninger G:
980 **Safety and efficacy of hematopoietic stem cell collection from mobilized peripheral blood in**
981 **unrelated volunteers: 12 years of single-center experience in 3928 donors**. *Blood* 2009,
982 **114**(18):3757-3763.
983 21. Li D, Xue W, Li M, Dong M, Wang J, Wang X, Li X, Chen K, Zhang W, Wu S *et al*: **VCAM-1(+)**
984 **macrophages guide the homing of HSPCs to a vascular niche**. *Nature* 2018,
985 **564**(7734):119-124.
986 22. Peterson TR, Laplante M, Thoreen CC, Sancak Y, Kang SA, Kuehl WM, Gray NS, Sabatini DM:
987 **DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and**
988 **required for their survival**. *Cell* 2009, **137**(5):873-886.
989 23. Totzke G, Essmann F, Pohlmann S, Lindenblatt C, Janicke RU, Schulze-Osthoff K: **A novel**
990 **member of the IkappaB family, human IkappaB-zeta, inhibits transactivation of p65 and its**
991 **DNA binding**. *The Journal of biological chemistry* 2006, **281**(18):12645-12654.
992 24. Blank U, Karlsson S: **TGF-β signaling in the control of hematopoietic stem cells**. *Blood* 2015,
993 **125**(23):3542-3550.
994 25. Ohishi K, Varnum-Finney B, Flowers D, Anasetti C, Myerson D, Bernstein ID: **Monocytes**
995 **express high amounts of Notch and undergo cytokine specific apoptosis following**
996 **interaction with the Notch ligand, Delta-1**. *Blood* 2000, **95**(9):2847-2854.
997 26. Staal FJT, Luis TC, Tiemessen MM: **WNT signalling in the immune system: WNT is spreading**
998 **its wings**. *Nature Reviews Immunology* 2008, **8**(8):581-593.
999 27. Zhou W, Gao F, Romero-Wolf M, Jo S, Rothenberg EV: **Single-cell deletion analyses show**
1000 **control of pro–T cell developmental speed and pathways by Tcf7, Spi1, Gata3, Bcl11a, Erg,**
1001 **and Bcl11b**. *Science immunology* 2022, **7**(71).
1002 28. van de Wetering M, Oosterwegel M, Dooijes D, Clevers H: **Identification and cloning of TCF-1,**
1003 **a T lymphocyte-specific transcription factor containing a sequence-specific HMG box**. *The*
1004 *EMBO journal* 1991, **10**(1):123-132.
1005 29. Shapiro-Shelef M, Calame K: **Regulation of plasma-cell development**. *Nature Reviews*
1006 *Immunology* 2005, **5**(3):230-242.
1007 30. Drissen R, Buza-Vidas N, Woll P, Thongjuea S, Gambardella A, Giustacchini A, Mancini E, Zriwil

1008 A, Lutteropp M, Grover A *et al*: **Distinct myeloid progenitor-differentiation pathways**
1009 **identified through single-cell RNA sequencing**. *Nature immunology* 2016, **17**(6):666-676.
1010 31. Siatecka M, Bieker JJ: **The multifunctional role of EKLF/KLF1 during erythropoiesis**. *Blood*
1011 2011, **118**(8):2044-2054.
1012 32. Nei Y, Obata-Ninomiya K, Tsutsui H, Ishiwata K, Miyasaka M, Matsumoto K, Nakae S, Kanuka H,
1013 Inase N, Karasuyama H: **GATA-1 regulates the generation and function of basophils**.
1014 *Proceedings of the National Academy of Sciences* 2013, **110**(46):18620-18625.
1015 33. Brugger W, Kreutz M, Andreesen R: **Macrophage Colony-Stimulating Factor Is Required for**
1016 **Human Monocyte Survival and Acts as a Cofactor for Their Terminal Differentiation to**
1017 **Macrophages In Vitro**. *Journal of leukocyte biology* 1991, **49**(5):483-488.
1018 34. Hiraoka A, Yano Ki K, Kagami N, Takeshige K, Mio H, Anazawa H, Sugimoto S: **Stem cell growth**
1019 **factor: in situ hybridization analysis on the gene expression, molecular characterization and**
1020 **in vitro proliferative activity of a recombinant preparation on primitive hematopoietic**
1021 **progenitor cells**. *The hematology journal : the official journal of the European Haematology*
1022 *Association* 2001, **2**(5):307-315.
1023 35. Hiraoka A, Ohkubo T, Fukuda M: **Production of human hematopoietic survival and growth**
1024 **factor by a myeloid leukemia cell line (KPB-M15) and placenta as detected by a monoclonal**
1025 **antibody**. *Cancer research* 1987, **47**(19):5025-5030.
1026 36. Pulugulla SH, Workman R, Rutter NW, Yang Z, Adamik J, Lupish B, Macar DA, el Abdouni S,
1027 Esposito EX, Galson DL *et al*: **A combined computational and experimental approach reveals**
1028 **the structure of a C/EBPβ–Spi1 interaction required for IL1B gene transcription**. *Journal of*
1029 *Biological Chemistry* 2018, **293**(52):19942-19956.
1030 37. Chute JP, Muramoto GG, Whitesides J, Colvin M, Safi R, Chao NJ, McDonnell DP: **Inhibition of**
1031 **aldehyde dehydrogenase and retinoid signaling induces the expansion of human**
1032 **hematopoietic stem cells**. *Proceedings of the National Academy of Sciences of the United*
1033 *States of America* 2006, **103**(31):11707-11712.
1034 38. Ley K, Hoffman HM, Kubes P, Cassatella MA, Zychlinsky A, Hedrick CC, Catz SD: **Neutrophils:**
1035 **New insights and open questions**. *Science immunology* 2018, **3**(30).
1036 39. Jiang J, Faiz A, Berg M, Carpaij OA, Vermeulen CJ, Brouwer S, Hesse L, Teichmann SA, Ten
1037 Hacken NHT, Timens W *et al*: **Gene signatures from scRNA-seq accurately quantify mast cells**
1038 **in biopsies in asthma**. *Clinical and experimental allergy : journal of the British Society for*
1039 *Allergy and Clinical Immunology* 2020, **50**(12):1428-1431.
1040 40. Klimchenko O, Mori M, Distefano A, Langlois T, Larbret F, Lecluse Y, Feraud O, Vainchenker W,
1041 Norol F, Debili N: **A common bipotent progenitor generates the erythroid and**
1042 **megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis**. *Blood*
1043 2009, **114**(8):1506-1517.
1044 41. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, Endele M,
1045 Filipczyk A, Gambardella A, Ahmed N *et al*: **Early myeloid lineage choice is not initiated by**
1046 **random PU.1 to GATA1 protein ratios**. *Nature* 2016, **535**(7611):299-302.
1047 42. Bergiers I, Andrews T, Vargel Bolukbasi O, Buness A, Janosz E, Lopez-Anguita N, Ganter K,
1048 Kosim K, Celen C, Itir Percin G *et al*: **Single-cell transcriptomics reveals a new dynamical**
1049 **function of transcription factors during embryonic hematopoiesis**. *eLife* 2018, **7**.
1050 43. erythroid abnormalitiesGasiorek JJ, Blank V: **Regulation and function of the NFE2**
1051 **transcription factor in hematopoietic and non-hematopoietic cells**. *Cellular and molecular*

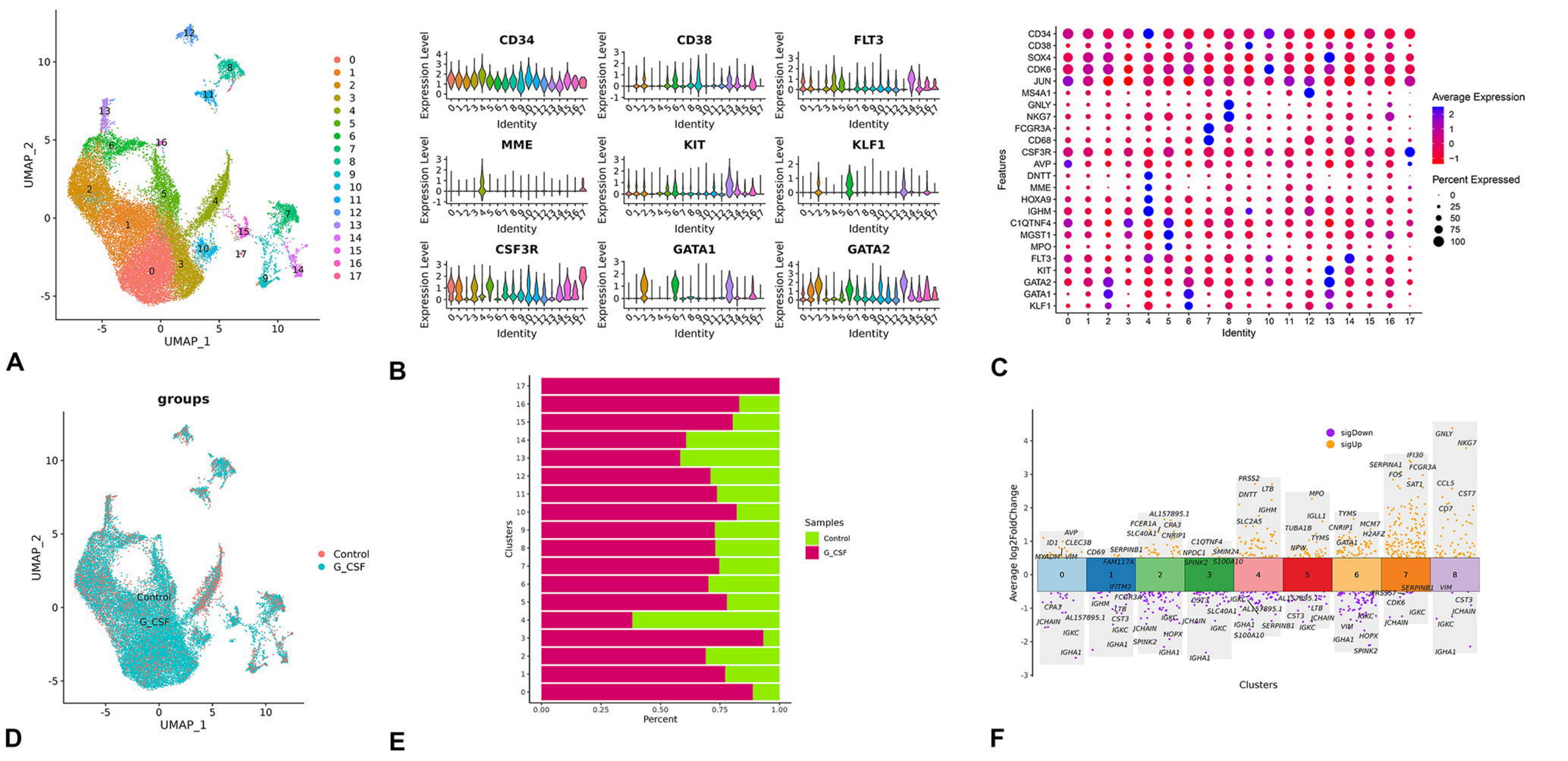
1052 *life sciences : CMLS* 2015, **72**(12):2323-2335.
1053 44. Chan WY, Follows GA, Lacaud G, Pimanda JE, Landry JR, Kinston S, Knezevic K, Piltz S,
1054 Donaldson IJ, Gambardella L *et al*: **The paralogous hematopoietic regulators Lyl1 and Scl are**
1055 **coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype**.
1056 *Blood* 2007, **109**(5):1908-1916.
1057 45. Laurenti E, Doulatov S, Zandi S, Plumb I, Chen J, April C, Fan JB, Dick JE: **The transcriptional**
1058 **architecture of early human hematopoiesis identifies multilevel control of lymphoid**
1059 **commitment**. *Nature immunology* 2013, **14**(7):756-763.
1060 46. Takahashi K, Yamanaka S: **Induction of Pluripotent Stem Cells from Mouse Embryonic and**
1061 **Adult Fibroblast Cultures by Defined Factors**. *Cell* 2006, **126**(4):663-676.
1062 47. Capellera-Garcia S, Pulecio J, Dhulipala K, Siva K, Rayon-Estrada V, Singbrant S, Sommarin
1063 Mikael NE, Walkley Carl R, Soneji S, Karlsson G *et al*: **Defining the Minimal Factors Required**
1064 **for Erythropoiesis through Direct Lineage Conversion**. *Cell reports* 2016, **15**(11):2550-2562.
1065 48. Moreau T, Evans AL, Vasquez L, Tijssen MR, Yan Y, Trotter MW, Howard D, Colzani M,
1066 Arumugam M, Wu WH *et al*: **Large-scale production of megakaryocytes from human**
1067 **pluripotent stem cells by chemically defined forward programming**. *Nature communications*
1068 2016, **7**(1).
1069 49. Beekman R, Touw IP: **G-CSF and its receptor in myeloid malignancy**. *Blood* 2010,
1070 **115**(25):5131-5136.
1071 50. Hume DA, MacDonald KP: **Therapeutic applications of macrophage colony-stimulating**
1072 **factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling**. *Blood* 2012,
1073 **119**(8):1810-1820.
1074 51. Akashi K, Traver D, Miyamoto T, Weissman IL: **A clonogenic common myeloid progenitor that**
1075 **gives rise to all myeloid lineages**. *Nature* 2000, **404**(6774):193-197.
1076 52. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X *et al*: **Clinical features of**
1077 **patients infected with 2019 novel coronavirus in Wuhan, China**. *Lancet* 2020,
1078 **395**(10223):497-506.
1079 53. Fajgenbaum DC, Longo DL, June CH: **Cytokine Storm**. *New England Journal of Medicine* 2020,
1080 **383**(23):2255-2273.
1081 54. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M,
1082 Smibert P, Satija R: **Comprehensive Integration of Single-Cell Data**. *Cell* 2019,
1083 **177**(7):1888-1902 e1821.
1084 55. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen
1085 TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by**
1086 **pseudotemporal ordering of single cells**. *Nature biotechnology* 2014, **32**(4):381-386.
1087 56. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F,
1088 Marine J-C, Geurts P, Aerts J *et al*: **SCENIC: single-cell regulatory network inference and**
1089 **clustering**. *Nature Methods* 2017, **14**(11):1083-1086.
1090 57. Hänzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and**
1091 **RNA-Seq data**. *BMC Bioinformatics* 2013, **14**(1).
1092 58. Oetjen KA, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, Dillon LW, McCoy JP, Hourigan CS:
1093 **Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow**
1094 **cytometry**. *JCI insight* 2018, **3**(23).
1095 59. Zhu C, Lian Y, Wang C, Wu P, Li X, Gao Y, Fan S, Ai L, Fang L, Pan H *et al*: **Single-cell**

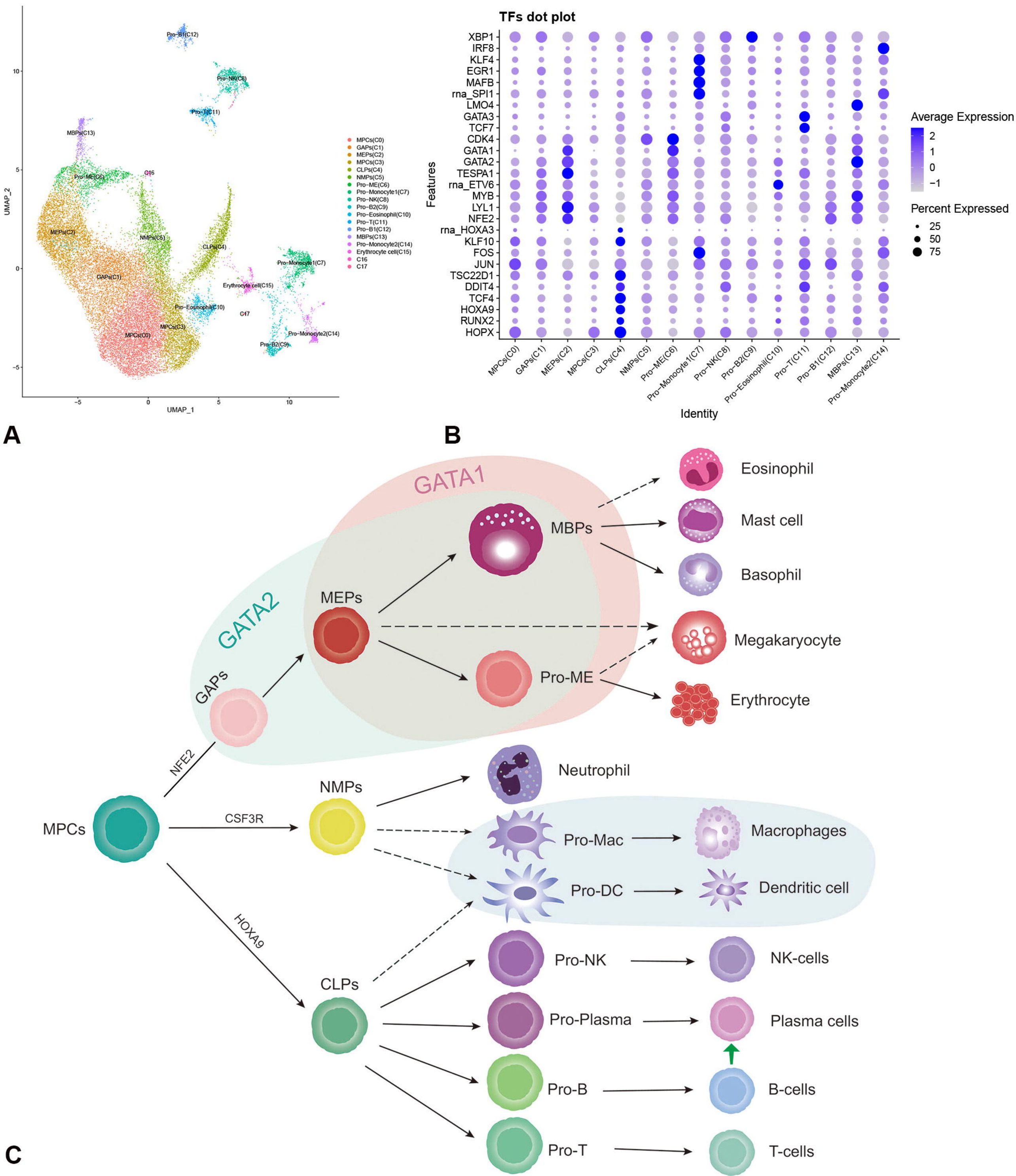
1096 **transcriptomics dissects hematopoietic cell destruction and T-cell engagement in aplastic**
1097 **anemia**. *Blood* 2021, **138**(1):23-33.
1098 60. Tonglin H, Yanna Z, Xiaoling Y, Ruilan G, Liming Y: **Single-Cell RNA-Seq of Bone Marrow Cells**
1099 **in Aplastic Anemia**. *Front Genet* 2021, **12**:745483.
1100 61. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S,
1101 Byrnes L, Lanata CM *et al*: **Multiplexed droplet single-cell RNA-sequencing using natural**
1102 **genetic variation**. *Nature biotechnology* 2017, **36**(1):89-94.
1103
1104

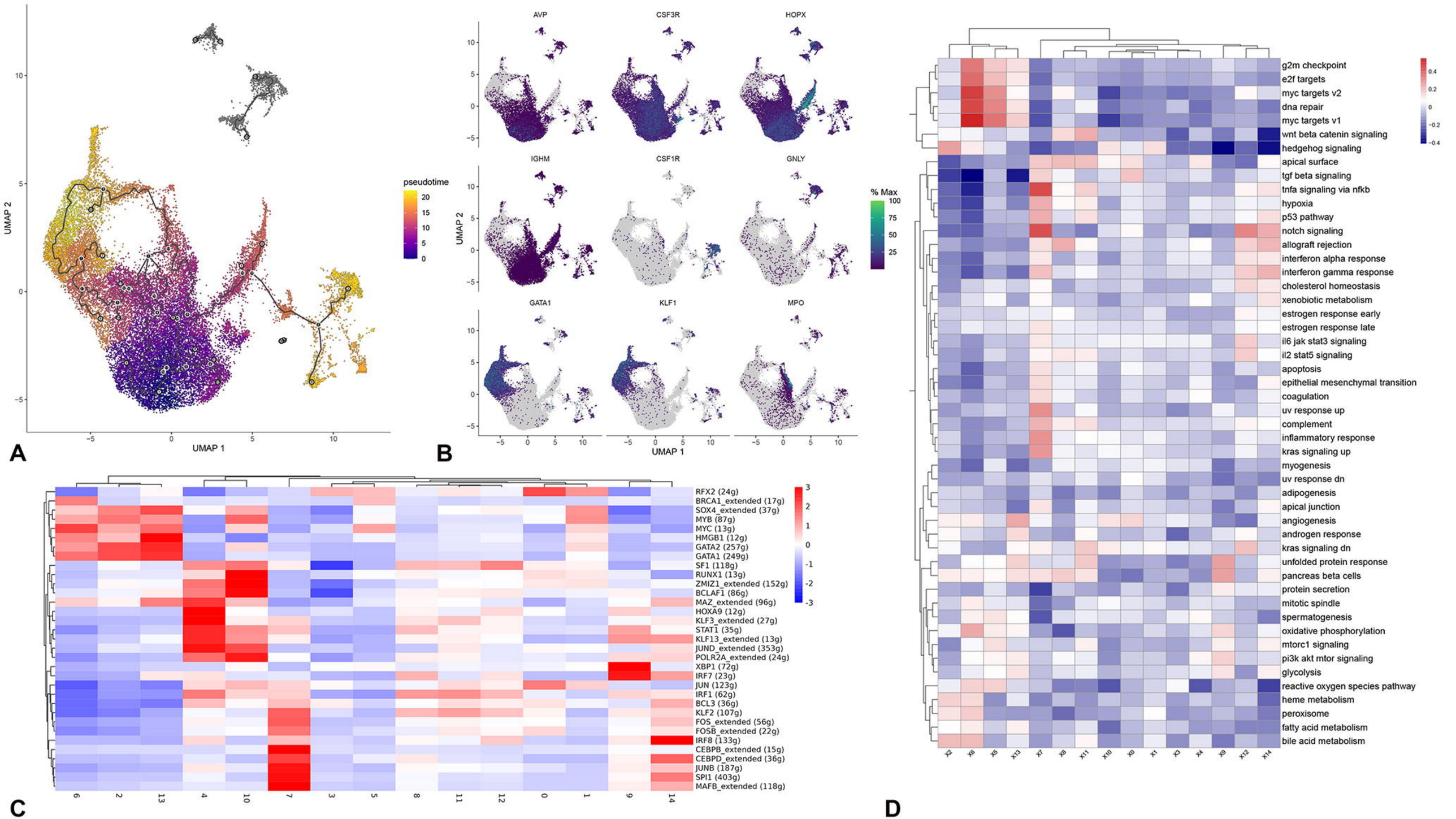

Classical hierarchy of hematopoiesis
HPCs
CMP
MEP
GMP
CLP
Pro-TBNK
Erythrocyte
Megakaryocyte
Eosinophil
Pro-Monocyte
Neutrophil
Basophil
NK-cells
B-cells
T-cells
Reforming the hierarchy of hematopoiesis
GATA1
GATA2
MPCs
NFE2
GAPs
MEPs
MBPs
Pro-ME
Eosinophil
Mast cell
Basophil
Megakaryocyte
Erythrocyte
CSF3R
NMPs
Neutrophil
Pro-Mac
Macrophages
Pro-DC
Dendritic cell
HOXA9
CLPs
Pro-NK
NK-cells
Pro-Plasma
Plasma cells
Pro-B
B-cells
Pro-T
T-cells



ID	Description
GO:0006260	DNA replication
GO:0000727	double-strand break repair via break-induced replication
GO:1990748	cellular detoxification
GO:0009636	response to toxic substance
GO:0006270	DNA replication initiation
GO:0097237	cellular response to toxic substance
GO:0098869	cellular oxidant detoxification
GO:0030099	myeloid cell differentiation
GO:0051383	kinetochore organization
GO:0045787	positive regulation of cell cycle
GO:1903706	regulation of hemopoiesis
GO:0098754	detoxification
GO:0010639	negative regulation of organelle organization
GO:1902749	regulation of cell cycle G2/M phase transition
GO:0006979	response to oxidative stress
GO:0090068	positive regulation of cell cycle process
GO:0031261	DNA replication preinitiation complex
GO:0031091	platelet alpha granule
GO:0071162	CMG complex
GO:0034774	secretory granule lumen
GO:0060205	cytoplasmic vesicle lumen
GO:0031983	vesicle lumen
GO:0042827	platelet dense granule
GO:0062023	collagen-containing extracellular matrix
GO:0031093	platelet alpha granule lumen
GO:0004601	peroxidase activity
GO:0016684	oxidoreductase activity, acting on peroxide as acceptor
GO:0016209	antioxidant activity


A
UMAP_1
UMAP_2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
B
CD34
CD38
FLT3
MME
KIT
KLF1
CSF3R
GATA1
GATA2
Expression Level
Identity
C
CD34
CD38
SOX4
CDK6
JUN
MS4A1
GNLY
NKG7
FCGR3A
CD68
CSF3R
AVP
DNTT
MME
HOXA9
IGHM
C1QTNF4
MGST1
MPO
FLT3
KIT
GATA2
GATA1
KLF1
Features
Identity
Average Expression
Percent Expressed
D
groups
Control
G_CSF
UMAP_1
UMAP_2
E
Clusters
Percent
Samples
Control
G_CSF
F
sigDown
sigUp
Average log2FoldChange
Clusters


A
UMAP_2
UMAP_1
Pro-B1(C12)
Pro-NK(C8)
Pro-T(C11)
MBPs(C13)
Pro-ME(C6)
C16
MEPs(C2)
NMPs(C5)
CLPs(C4)
Pro-Monocyte1(C7)
GAPs(C1)
Erythrocyte cell(C15)
Pro-Eosinophil(C10)
C17
MPCs(C3)
MPCs(C0)
Pro-Monocyte2(C14)
Pro-B2(C9)
MPCs(C0)
GAPs(C1)
MEPs(C2)
MPCs(C3)
CLPs(C4)
NMPs(C5)
Pro-ME(C6)
Pro-Monocyte1(C7)
Pro-NK(C8)
Pro-B2(C9)
Pro-Eosinophil(C10)
Pro-T(C11)
Pro-B1(C12)
MBPs(C13)
Pro-Monocyte2(C14)
Erythrocyte cell(C15)
C16
C17
TFs dot plot
Features
Identity
XBP1
IRF8
KLF4
EGR1
MAFB
rna_SPI1
LMO4
GATA3
TCF7
CDK4
GATA1
GATA2
TESPA1
rna_ETV6
MYB
LYL1
NFE2
rna_HOXA3
KLF10
FOS
JUN
TSC22D1
DDIT4
TCF4
HOXA9
RUNX2
HOPX
Average Expression
2
1
0
−1
Percent Expressed
25
50
75
B
GATA1
GATA2
MBPs
MEPs
Pro-ME
GAPs
NFE2
Eosinophil
Mast cell
Basophil
Megakaryocyte
Erythrocyte
Neutrophil
NMPs
MPCs
CSF3R
Pro-Mac
Macrophages
Pro-DC
Dendritic cell
HOXA9
Pro-NK
NK-cells
CLPs
Pro-Plasma
Plasma cells
Pro-B
B-cells
Pro-T
T-cells
C


A
UMAP 1
UMAP 2
pseudotime
20
15
10
5
0
B
AVP
CSF3R
HOPX
IGHM
CSF1R
GNLY
GATA1
KLF1
MPO
% Max
100
75
50
25
C
RFX2 (24g)
BRCA1_extended (17g)
SOX4_extended (37g)
MYB (87g)
MYC (13g)
HMGB1 (12g)
GATA2 (257g)
GATA1 (249g)
SF1 (118g)
RUNX1 (13g)
ZMIZ1_extended (152g)
BCLAF1 (86g)
MAZ_extended (96g)
HOXA9 (12g)
KLF3_extended (27g)
STAT1 (35g)
KLF13_extended (13g)
JUND_extended (353g)
POLR2A_extended (24g)
XBP1 (72g)
IRF7 (23g)
JUN (123g)
IRF1 (62g)
BCL3 (36g)
KLF2 (107g)
FOS_extended (56g)
FOSB_extended (22g)
IRF8 (133g)
CEBPB_extended (15g)
CEBPD_extended (36g)
JUNB (187g)
SPI1 (403g)
MAFB_extended (118g)
6
2
13
4
10
7
3
5
8
11
12
0
1
9
14
D
g2m checkpoint
e2f targets
myc targets v2
dna repair
myc targets v1
wnt beta catenin signaling
hedgehog signaling
apical surface
tgf beta signaling
tnfa signaling via nfkb
hypoxia
p53 pathway
notch signaling
allograft rejection
interferon alpha response
interferon gamma response
cholesterol homeostasis
xenobiotic metabolism
estrogen response early
estrogen response late
il6 jak stat3 signaling
il2 stat5 signaling
apoptosis
epithelial mesenchymal transition
coagulation
uv response up
complement
inflammatory response
kras signaling up
myogenesis
uv response dn
adipogenesis
apical junction
angiogenesis
androgen response
kras signaling dn
unfolded protein response
pancreas beta cells
protein secretion
mitotic spindle
spermatogenesis
oxidative phosphorylation
mtorc1 signaling
pi3k akt mtor signaling
glycolysis
reactive oxygen species pathway
heme metabolism
peroxisome
fatty acid metabolism
bile acid metabolism
X2
X6
X5
X13
X7
X8
X11
X10
X0
X1
X3
X4
X9
X12
X14

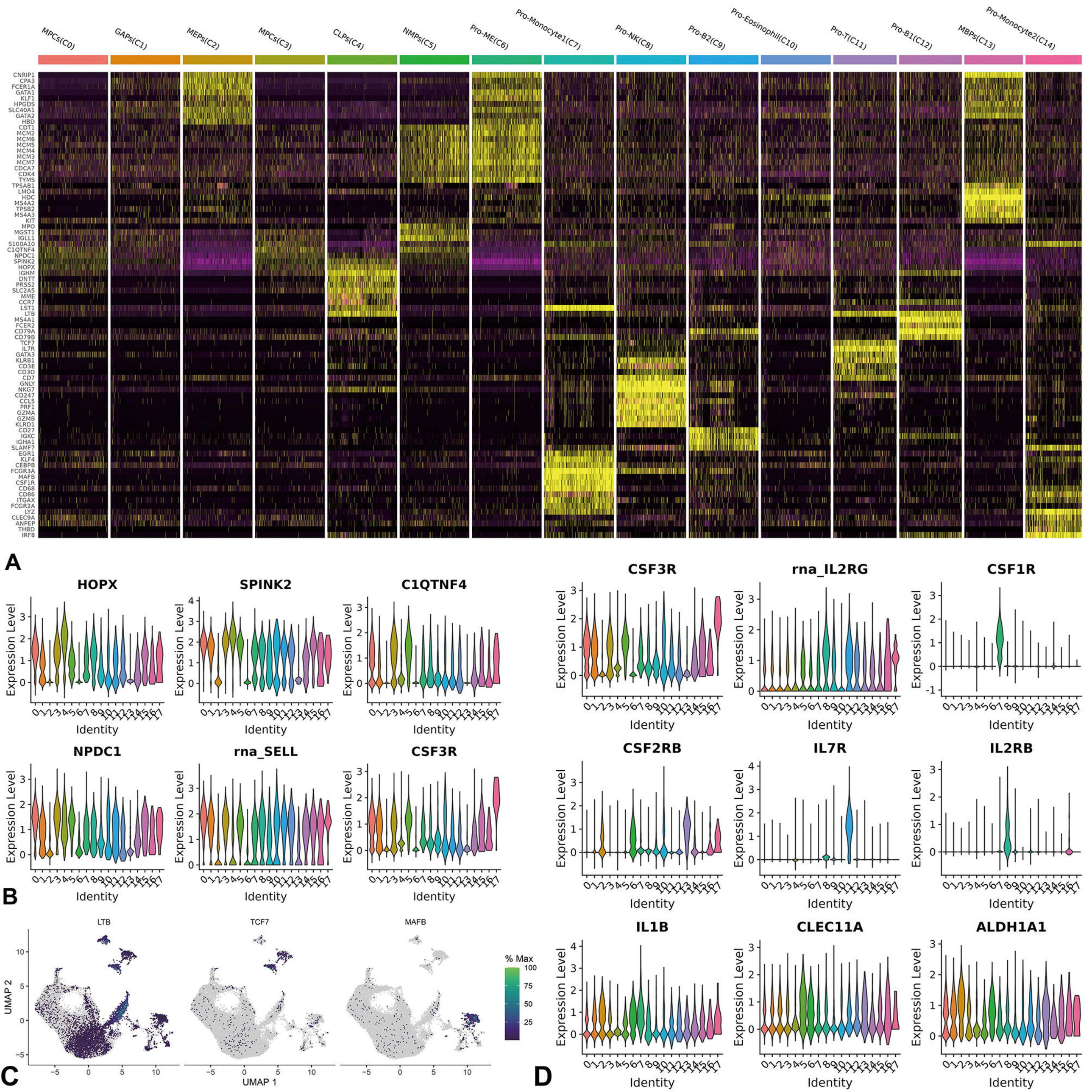

MPCs(C0)
GAPs(C1)
MEPs(C2)
MPCs(C3)
CLPs(C4)
NMPs(C5)
Pro-ME(C6)
Pro-Monocyte1(C7)
Pro-NK(C8)
Pro-B2(C9)
Pro-Eosinophil(C10)
Pro-T(C11)
Pro-B1(C12)
MBPs(C13)
Pro-Monocyte2(C14)
CNRIP1
CPA3
FCER1A
GATA1
KLF1
HPGDS
SLC40A1
GATA2
HBD
CDT1
MCM2
MCM6
MCM5
MCM4
MCM3
MCM7
CDCA7
CDK4
TYMS
TPSAB1
LMO4
HDC
MS4A2
TPSB2
MS4A3
KIT
MPO
MGST1
IGLL1
S100A10
C1QTNF4
NPDC1
SPINK2
HOPX
IGHM
DNTT
PRSS2
SLC2A5
MME
CCR7
LST1
LTB
MS4A1
FCER2
CD79A
CD79B
TCF7
IL7R
GATA3
KLRB1
CD3E
CD3D
CD7
GNLY
NKG7
CD247
CCL5
PRF1
GZMA
GZMB
KLRD1
CD27
IGKC
IGHA1
SLAMF7
EGR1
KLF4
CEBPB
FCGR3A
MAFB
CSF1R
CD68
CD86
ITGAX
FCGR2A
LYZ
CLEC9A
ANPEP
THBD
IRF8
A
HOPX
SPINK2
C1QTNF4
NPDC1
rna_SELL
CSF3R
Expression Level
Identity
B
LTB
TCF7
MAFB
% Max
100
75
50
25
UMAP 2
UMAP 1
C
CSF3R
rna_IL2RG
CSF1R
CSF2RB
IL7R
IL2RB
IL1B
CLEC11A
ALDH1A1
D

Table 1. Selected specific marker genes and genes associated with fate decisions in all redefined subgroups

Progenitor Cell Types	Marker genes and fate decision genes (partly)	Transcription factors	Negative
CLP (C4)	SPINK2, HOPX, HOXA9, RUNX2, LTB, IGHM, DNTT, PRSS2, SLC2A5, MME, CCR7, NKG7, LST1, BASP1, CD79A	HOPX,DDIT4, HOXA9, RUNX2	CNRIP1,FCER1A,GATA1, S100A10
Pro-NK (C8)	GNLY,NKG7,CD247, CCL5, FCGR3A, PRF1, GZMA, GZMB, KLRD1, KLRB1 ,CD3E,CD7. HOPX	DDIT4, HOPX	IL7R,GATA3
Pro-B2 (C9, Pro-Plasma cell)	CD27, IGKC, IGHA1, SLAMF7, CD79A, CD79B, PRDM1, IRF4, CD38, JCHAIN	PRDM1, IRF4	MS4A1, FCER2
Pro-T (C11)	TCF7, IL7R, GATA3, KLRB1, CD3E, CD3D, CD7, CD247, LTB, DDIT4	TCF7,GATA3,DDIT4	GNLY, FCGR3A, GZMA
Pro-B1 (C12)	CD19, MS4A1,FCER2,CD79A, CD79B, IGHM, LTB, IGKC, PAX5, VPREB3,CD22, CD24	PAX5	CD27
NMPs(C5)	CSF3R, MPO, MGST1, IGLL1, S100A10, C1QTNF4, NPDC1, MYB, CDK4, CDCA7	MYB, CDK4	GATA2, SLC40A1, CNRIP1,LTB
Lineages(C2,C6,C13)	CNRIP1, CPA3 ,FCER1A, GATA1, KLF1, HPGDS, SLC40A1, GATA2, ZBTB16 ,TAL1, TESPA1, MYB, CDK4, ITGA2B	GATA1, ZBTB16,TAL1, CDK4, TESPA1,KLF1	SPINK2,HOPX, C1QTNF4, CSF3R
GAPs (C1)	GATA2, NFE2, LYL1, MYB, SLC40A1, TESPA1, CSF3R	GATA2, NFE2, LYL1, MYB	GATA1, KLF1
MEPs (C2)	GATA2,NFE2,LYL1,MYB,GATA1,KLF1,CSF2RB,SLC40A1, ZBTB16,TAL1,CDK4,TESPA1	GATA1, KLF1, ZBTB16, TAL1	CSF3R
Pro-ME(C6)	HBD,CDT1, MCM2, MCM6, MCM5, MCM4, MCM3, MCM7, CDCA7, CDK4, TYMS	CDK4	MS4A2, MS4A3

MBPs (C13)	TPSAB1,LMO4, HDC, MS4A2, TPSB2, MS4A3, KIT, HPGDS, CLC, PRG2, MCM2-MCM7,APOC1	LMO4, CDK4	HBD
Eosinophil (C10)	CLC, HDC, RFLNB, ETV6	ETV6	MS4A2,TPSB2,MS4A3
Pro-Monocytes1 (C7, Macrophages)	EGR1,SPI1,KLF4,CEBPB,FCGR3A, CSF1R,CD68, CD86, ITGAX, FCGR2A , LYZ , LST1, EGR2, CEBPA, MAFB, TNF	SPI1,KLF4,CEBPB, EGR1, EGR2, CEBPA, MAFB	CLEC9A, THBD, IRF8
Pro-Monocytes2 (C14, DCs)	CLEC9A, ANPEP, THBD, IRF8, KLF4, CD68, CD86, ITGAX, LYZ, SPI1 , LST1, DDIT4, SLAMF7	SPI1,KLF4,IRF8,DDIT4	FCGR3A, CSF1R, MAFB