Redefining hematopoietic progenitor cells and reaffirming lineage commitment by transcriptional atlas

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

[1]Shenzhen Institute of Respiratory Disease, the First Affiliated Hospital of Southern University of Science and Technology, Shenzhen, Guangdong, China
[2]Shenzhen Key Laboratory of Reproductive Immunology for Peri-implantation, Fertility Center, Shenzhen Zhongshan Urology Hospital, Shenzhen, China
[3]Clinical Centre, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Shenzhen, Guangdong, China
[4]Department of hematology, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Shenzhen, Guangdong, China

*** 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

Abstract

Tracing the differentiation of the hematopoietic progenitor cells(HPCs) in human blood, particularly with the scarcity in adults, has been challenging owing to unable capture the entire differentiation stages and subgroups of HPCs. Deciphering the underlying hematopoiesis mechanism, including progenitor cells definition and fate decisions identification, is critical to understanding fundamental questions of hematopoietic lineage commitment. Often, the use of classic and traditional surface markers has been shown to be inadequate, speciosity, and ineffective in accurately defining hematopoietic progenitors. The lineage definition and commitment of hematopoietic progenitors in adult peripheral blood remains to be elucidated by a reliable transcriptional atlas. Here, single cell RNA sequencing (scRNA) was employed to define the lineage of HPCs and reaffirm the hierarchy of hematopoiesis. A total of 27,566 adult peripheral blood CD34+ cells were enriched and sequenced, capturing most differentiation stages and subgroups of HPCs. HPCs lineages were redefined by unique makers and key factor genes. We are the first to clearly depict the transcriptional atlas and define each subgroup of adult peripheral blood progenitors. We recovered the key transcript factors involved in lineage fate decision, and restructured the lineage commitment model of HPCs that can reconcile previous models and explain conflicts among functional studies. Several fate decision key factors of progenitors were identified as major switches that controlled hematopoietic lineage formation. Moreover, we are also the first to identify the cell fate decision transcriptional factors in each branch among hematopoietic hierarchy formation. Our reaffirmed lineage commitment also revealed that the erythroid lineage commitment process is a continuous process but the lymphoid lineage commitment process is more likely a discrete stepwise process that showed distinct lineage-committed populations during the hematopoiesis. With these information, an elaborate and comprehensive transcriptional atlas of all types of hematopoietic progenitors was then constructed, which enabled the whole map of the transcriptional signal and pathway among lineage commitment. Importantly, this newly created atlas can enable accurate definition, transplantation, and biological insights of hematopoietic progenitors for facilitating

new research development in cell engineering, regenerative medicine, and disease treatment, which we call Artificial Operation of HPCs. Furthermore, this opens up a wide variety of possibilities for artificial precise control progenitor cell differentiation and targeted treatment of immune disease, providing hope for untreatable diseases.

Graphical abstract

The reaffirmed lineage commitment model of HPCs

Highlights

- Classic and traditional surface markers are often inadequate, speciosity, and ineffective in accurately defining hematopoietic progenitor cells.
- Multiple hematopoietic lineages progenitor cells of adult peripheral blood were clearly depicted and redefined by unique makers and key transcript factors.
- Key fate decision transcript factors and genes in each branch among hematopoietic hierarchy formation were recovered by transcriptional atlas.
- Several key factors, 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, were identified as major regulators that controlled hematopoietic lineage formation.
- Fate decision transcription factors were identified in progenitor cells of erythrocyte (GATA1, KLF1), monocyte (MAFB, IRF8), B cells (PAX5, PRDM1), NK cells (DDIT4, HOPX), and T cells (TCF7, GATA3).
- Plasma progenitor cells were first identified and defined, and the rearrangements of IGHV, IGKV, and IGLV genes were already existed and occurred during the progenitor state of plasma cells.

- The erythroid lineage commitment process is a continuous process but the lymphoid lineage commitment process is more likely a discrete stepwise process that showed distinct lineage-committed populations during the hematopoiesis.
- The lineage commitment model of HPCs was reaffirmed that can reconcile previous models and explain the conflict among functional studies.
- Transcriptional atlas of HPCs helps to provide a well explanation of the immune reconstitution in adults and on its cytokines secretion and regulation.
- Application of this atlas can enable artificial precise control progenitor cell differentiation, in vitro accurate culture of each lineage cells, and targeted treatment of immune cell disease to be made possible for future development.

Introduction

Deciphering the hematopoiesis mechanism that underlies progenitor cells definition and fate decision are critical to understanding fundamental questions of hematopoietic lineage commitment. Toward the diversities of peripheral blood cells in the human body, it is like the multifarious stars existing in our galaxy. Although we know their existence, the exact characterization of these cell subsets remains a challenge.

Among the various cell types in peripheral blood (PB), hematopoietic stem cells (HSCs) are one of such rare cell subsets. These cells possess self-renewal capacity and can differentiate to all blood lineage cells. Such that, HSCs reside at the top of a hierarchy and differentiate to all blood lineage cells with tree-like pattern in classical hematopoietic hierarchical models [1]. Single-cell sequencing technique has been successfully implemented to explore molecular heterogeneity and hierarchy of HSCs [2-4]. The lineage commitment is affected by transcription factors (TFs) with a continuum fate decision process [5, 6]. Though *in vitro* and *in vivo* studies have reaffirmed their critical role of transcription factors in lineage commitment, the function and regulatory mechanism of HSCs is still unclear [7]. In addition, the transcriptional atlas and the differentiation among subtypes of progenitors in adult peripheral blood remain to be elucidated.

The reason for the molecular mechanisms that controlled lineage commitment and progenitor behaviors remaining an enigma is due to various reasons. Firstly, the definition of hematopoietic progenitor cells (HPCs) is not standardized. Based on the classical and traditional definition, the characterization of HPCs is based on the presence of surface markers, such as CD34, CD38, FLT3 and KIT [5]. Similarly, the combination of surface markers has been proposed for the identification toward common lymphoid progenitors (CLP) and common myeloid progenitors (CMP). These markers include CD10 (MME), FLT3, and CD38. Yet, owing to the rare and transient nature, the subtypes of CLP and CMP from HSCs have never been stringently identified and accurately captured. The capacity of surface markers is very limited in capturing the internal state of progenitor cells and reflecting the dynamic

differentiation mechanisms underlying hematopoiesis [8]. Secondly, the functional studies were also used surface markers for the characterization of HPCs. These include the selection of CD34 positive cell from bone marrow or cord blood. However, this approach can be affected by the type of culture methods and surface markers being used for the identification. Simply, the use of classical and traditional markers may be not insufficient, such as KIT, to differentiate mature subset cells. Thus, the use of combined surface markers for HPCs may cause biased and arbitrary. Thirdly, bone marrow progenitors have been widely used in functional studies and subgroups identification. However, the limitation is mainly comprised quiescent progenitors and the absence of the whole spectrum of all progenitor stages. Finally, it is still challenge to capture the entire spectrum of adult peripheral blood HPCs.

In recent years, many studies have identified the lineage by using single cell sequencing. However, there are still several limitations of the transcriptional atlas being defined. The scRNA expression data of CD34 positive cells showed two well-defined subtypes of progenitor CLP and CMP being scattered and mixed distribution in T-SNE plots. This result implies that the transcriptional atlas for these type of cells are not distinguishable by unsupervised dimensionality reduction and clustering. When compared with others mature cells, HPCs should exist as a relatively homogeneous progenitor population. The idea classification and defined subset of lineage cells should be separately distinguishable by unsupervised dimensionality reduction which carried the PCA key markers of genes as factors.

In addition, the hematopoietic hierarchical models are also inconsistent among the literature studies [1, 9, 10]. Most of the established lineage fate decision models were based on *in-vitro* / *in-vivo* functional studies or cell fate induces experiments [11, 12]. The use of cellular conditions can often lead to inaccurate interpretations of HPCs cell fate decision because cell lineages were induced by diverse experimental conditions and cannot be permissive to all possible branches of lineage commitment [12]. Similarly, single cell studies would have modified some aspects of classical hematopoiesis hierarchical model, as the hematopoietic progenitors are changing continuously among the progenitors differentiation process [5, 13]. In addition,

150 progenitor cells were defined by classical markers and selected by surface markers [5,
151 6]. As each type of progenitors is very rare in whole blood, the captured cell numbers
152 are often very low with less than one hundred cells[5]. This implies that some key
153 progenitor subgroups may be missed. Most of the bone marrow HPCs was
154 quiescent state progenitors. scRNA Sequencing of bone marrow progenitors can't
155 clearly depict the whole lineage commitment of hematopoietic progenitors. Thus, this
156 would affect the reliability of results. As the classic definition of hematopoietic
157 progenitors remains inefficient and inaccurate, this can cause the disorganization
158 among the subgroups and without a reliable identification of the boundary among
159 them [2]. Moreover, unsupervised dimensionality reduction and clustering of lineage
160 cells demonstrated farraginous model [4]. The measurement of lineage definition, fate
161 decision, and establishing the reliable lineage commitment between them are still a
162 challenge [9, 14].

163 Herein, we redefined the adult hematopoietic progenitors, and recovered the key
164 transcript factors in lineages fate decision, and reaffirmed the lineage commitment
165 model of HPCs by single-cell sequencing. Our HPCs atlas defined the cellular biology
166 of differentia stage progenitors among hematopoietic lineages generation. Our study
167 can help to open new perspectives for the characterization, definition, and
168 classification of HPCs.

169

170

171 **Results**

172 ***Single-cell sequencing identified subtypes of progenitor cells in adults peripheral***
173 ***blood.***

174 In order to enrich the whole spectrum of HPCs and intermediate state progenitors,
175 matured lymphocyte and granulocyte were removed by surface marker panel
176 (Lineage−) and progenitor cells were enriched by negative deletion (Fig. 1A). 10X
177 genomics scRNA was conducted and unsupervised dimensionality reduction and
178 clustering analysis showed 32 clusters by Seurat and UMAP (Fig. 1B). A total of
179 159603 cells were captured and sequenced. Our single cell expression atlas level

180 showed heterogeneity of several CD34 positive progenitors clusters. We further
181 analyzed the significantly identified genes among progenitor clusters with the other
182 clusters. Genes of *SXO4, CDK6, CYTL1,* and *EGFL7* were identified to be highly
183 expressed in progenitor clusters (Fig. 1C, D), which consistent with bone marrow
184 progenitors (Fig. S1). It is very obviously that the gene of *EGFL7* and *CYTL1* are very
185 high and specific expressed in progenitors clusters, so that they are likely key
186 functional gene in hematopoietic progenitor cells lineage differentiation and mature.
187 The top 30 highly expression genes of progenitor cells were presented (Fig. 1E). The
188 gene set enrichment of differentially expressed genes showed hematopoietic cell
189 lineage, MAPk, and PI3K-Akt pathways were enriched in progenitor cells but not
190 significant (Fig. 1F). Additionally, GO analysis displayed the biological process
191 enriched in DNA replication, cell differentiation, and so on (Fig. 1G). Based on the
192 surface markers and the above hematopoietic stem cells key functional genes, we can
193 confirm that the identified CD34+ clusters are hematopoietic progenitor cells.
194

195 ***The effective and accuracy of classic and traditional surface markers for***
196 ***progenitors classification***

197 To assess diversity, a re-clustering of analysis was conducted on 27566 CD34+
198 cells, resulting in 17 clusters, including 21244 mobilization progenitors and 6322
199 quiescent progenitors respectively (Fig. 2A). Most clusters contained a large number
200 of progenitor cells (Table. S1). The identified progenitor subgroup cells present
201 category and functional heterogeneity. Based on our finding, the subgroup of cells
202 aggregated in three main differentiation directions. Traditionally, membrane markers
203 of progenitor CMP and GMP (Granulocyte-monocyte progenitors) were defined with
204 CD38+MME-FLT3+, while megakaryocytic–erythroid progenitors (MEP) were
205 defined with CD38+MME-FLT3-. However, progenitor clusters based on these
206 markers were not distinguishable among the subgroups. KIT (CD117) and FLT3 were
207 expressed in different clusters (Fig. 2B). This indicates that the progenitors of
208 CD34+KIT and CD34+FLT3 belonging to two different differentiation directions. As
209 the expressions of CD38 and FLT3 were almost contradictory, this demonstrated most

FLT3 positive cells were negative of CD38 expression in clusters (Fig 2B). This implies that surface markers of CD38, FLT3, KIT, and MME can be used to distinguish myeloid progenitors and lymphoid progenitor differentiation directions, but not suitable to use in subtypes and intermediate stages classification.

Our study showed the canonical markers of CD49f and THY1 were almost not found in HPCs progenitors. The expression of THY1 (CD90) is very rare (only 17 cells) in progenitor clusters which may cause as the progenitors were isolated from adult PBMC (Fig. S1A). In addition, our data showed that the use of CD49f, THY1, FLT3 and KIT to classify or define the progenitor cells groups were very artificial and subjective. To confirm whether the traditional classification markers THY1 and CD49f (ITGA6) expressed in progenitor cells, we analyzed four bone marrow single cell projects data of GSE120221, GSE117498, GSE145531 and GSE181989 (Fig. S1B). Total 8296 CD34 positive cells were extracted of the four projects merged data. The results showed positive expression cell numbers of THY1 is only 84 and CD49f is 824, and the expression distribution is scattered. This is similar with the other identified markers of FLT3, CD38 and KIT (Fig. S1C). The traditional classification progenitor cells (CMP or CLP) were also not distinguishable by unsupervised dimensionality reduction and clustering in bone marrow CD34+ cells. With the progenitors of CD34+FLT3+ cells defined a mixed population of progenitors, this finding may help to provide a well explanation that previously functional study of selected progenitors (such as CD34+FLT3+) presented multiple differential potential. Although the expression tendency of FLT3+ was consistent with previously definition, it remained inefficient and inaccurate to classify the differentiation stages among the myeloid lineages. Classic and traditional surface markers displayed inadequate, speciosity, and ineffective in accurately defining hematopoietic progenitor cells. The efficient and accurately of classical hematopoietic progenitors markers which be used to distinguished CMP, GMP, and CLPs, should be reconsideration.

239 ***New transcriptional atlas of adult peripheral blood progenitors***

240 Unsupervised dimensionality reduction analysis was conducted to identify new
241 efficient marker genes for progenitor clusters distinguishing. The progenitors were
242 clustering into several directions (Fig. 2A). As AVP was known to be uniquely
243 expressed in HSC, our data showed both of AVP and CSF3R were mainly expressed
244 in cluster 0 and 3 (Fig. 2B, C). The other is myeloid progenitors comprised with
245 cluster 1, cluster 2, cluster 6, and cluster 5, with the gradual up-regulation of CD38,
246 GATA1, and KIT (Fig 2B). Our results demonstrated that one direction lineage
247 clusters of 1, 2, and 6 can be clearly distinguished by the expression of transcription
248 factors KLF1, GATA1 and GATA2 (Fig 2B). That indicated these clusters are
249 megakaryocyte or erythrocyte progenitors. Cluster 5 was presented with markers
250 genes of MPO and CSF3R (Fig. 2C). As myeloperoxidase (MPO) is the most
251 abundant and uniquely expressed protein in human neutrophil and plays an important
252 role in neutrophil antimicrobial responses[15, 16]. The other marker gene PRSS57 of
253 neutrophils was also presented and up-regulated expression in the cluster [17]. This
254 indicated that the progenitor branches were most likely the granulocyte-monocyte
255 progenitors, shown to be in a completely different direction with megakaryocyte or
256 erythrocyte progenitors. Our results showed the expression of IGHM, MME, and
257 HOXA9 [18] was presented in cluster 4, which confirmed the cluster was lymphoid
258 progenitor cells (Fig. 2C). The expression level is increasing of the transcription
259 factors HOXA9 from cluster 0 to cluster 4. Transcription factors presented enormous
260 capacity to fully distinguish the intermediate state and to faithfully reflect the
261 mechanisms underlying hematopoietic lineage commitment. Importantly, this
262 demonstrated that the expression and activation of transcription factors combined with
263 unique marker genes would be more efficient for lineage definition of HPCs.

264 To examine whether any different among the mobilization and quiescent
265 (unperturbed) progenitors, cell distribution analysis was conducted in the G-CSF
266 mobilization peripheral blood progenitors. Our results indicated that the total number
267 of all types of progenitors were remarkable increased. All kinds of subtype
268 progenitors were shown to be existed in the quiescent state when compared before

269 and after mobilization (Fig. 2D). This signifies that all types of progenitors were
270 existed in the quiescent state of peripheral blood, which also consistent with the
271 cluster distributions of mobilization state. This may also indicated that the
272 unperturbed adult hematopoiesis depends on replenishment from multiple potential
273 progenitors. Since the progenitor was shown to gradually differentiate to mature
274 stages in the peripheral blood, it would be expected fold changes of the more
275 primitive progenitors to be higher than mature progenitors. Based on our analysis, the
276 highest cell numbers and folder changed among the clusters were found to be in
277 clusters 0 and 3 (Fig. 2E). Those clusters were the most original multi-potent clusters
278 named as MPCs, which can prime to all types of lineages. The expression
279 characterization of CSF3R was consistent with G-CSF mobilization in peripheral
280 blood HSC transplantation.

281 To identify the common myeloid progenitor, re-clustering was conducted of
282 myeloid clusters. Interestingly, the common intermediate progenitor stage was not
283 observed among myeloid lineages indicating that the common myeloid progenitor did
284 not exist (Fig. S2). Furthermore, the expression of myeloid representative genes
285 gradually accumulated during differentiation, revealing distinct differentiation
286 characteristics. That indicated the myeloid progenitor hierarchical organization in
287 hematopoiesis is likely a continuum process. In summary, there are three clusters
288 which nearby the original progenitor of MPCs (Fig. S2A). Common myeloid
289 progenitors were not identified during the lineage commitment. The lymphoid
290 progenitors, neutrophils progenitors, and megakaryocytic–erythroid progenitors were
291 presented at the same time point at the first step of HPCs priming process. It is clearly
292 shown that during the HPCs priming and differentiation procedures, the original
293 MPCs priming into three progenitor cell types coinstantaneous and later maturing to
294 diversity hematopoietic lineages. Self-renewal capacity of MPCs is generally reduced
295 once differentiation into next stages. Taken together, based on the successful capture
296 the whole spectrum of adult peripheral blood progenitors, including both mobilization
297 and quiescent progenitor states, we are the first to clearly depict the transcriptional
298 atlas of adult peripheral blood progenitors.

299

300

301 ***Re-definition of hematopoietic progenitor cells and reaffirming the hierarchy of***
302 ***hematopoiesis***

303 To clearly re-define the subgroups of progenitors hematopoietic, the
304 characterization of all subtype progenitors cells were re-analyzed. Our result showed
305 the selected unique and specific marker genes and key transcript factors of each
306 lineage progenitors (Table 1, Fig. 2F). Additionally, the negatively expressed genes
307 were listed to distinguish each subgroup and clearly define the lineages progenitors
308 (Table 1). As previously shown, cluster 0 and 3 were the most original progenitors in
309 the peripheral blood that was identified and designated as MPCs. The transcriptional
310 atlas between cluster 3 and cluster 0 is very similar, with cluster 3 presenting an early
311 tendency towards lymphoid progenitors. All of the later mature progenitors are
312 originate from those initial cells. Based on this rationale, we re-defined the three
313 directions with common lymphoid progenitors (CLPs), GAPs (GATA genes controlled
314 progenitors), and neutrophilic and monocyte progenitors (NMPs), which cannot be
315 characterized by CD38, FLT3, and KIT (Fig. 3A, Table 1). Most of the top five
316 specific marker genes for each lineage progenitor were canonical marker genes of
317 classically defined hemopoietic cells, indicating the reliable characterization of
318 subgroups (Fig. 2F).

319

320 The expression of unique genes gradually accumulated during the differentiation
321 process among megakaryocytic–erythroid lineages, such as GATA2, CPA3, CNRIP1,
322 and SLC40A1. The differentiation of GAPs lineages was independently with the other
323 two branches. The shape of each branch was controlled by the expression of key
324 regulated genes which was shown to function like a gradually activate switch. During
325 the myeloid progenitor differentiation process, the expression of GATA2 and CSF3R
326 genes is the first branch being formed and presented (Fig. 3). This was shown to be
327 followed by the down-regulation of CSF3R genes and up-regulation of GATA2,
328 which implies the progenitors were priming to oligo-potent progenitor GAPs. The

329 expression of GATA1 and KLF1 were presented and gradually up-regulated at later
330 stage (Fig. 2B). This subgroup was GAPs mature stage defined as
331 megakaryocytic–erythroid progenitors (MEPs). The expression of CSF3R was mainly
332 showed in the original progenitors and almost disappeared in the later stage of MEPs.
333 The cytokine receptors will co-express or co-silence with the relevant lineage key
334 regulators, which indicated the progenitor lineages priming was multiple genes
335 regulate process. With the differentiation of MEPs enhancement, the second branch
336 was presented in the mature stage of MEPs. Toward the final stages of the MEPs, the
337 expression of KIT, CDK4, GATA1, KLF1 or LMO4, controlled the differentiation of
338 progenitor to megakaryocyte progenitors, erythrocyte progenitors (Pro-ME), or mast
339 cell and basophil progenitors (MBPs). Our results showed the expression of KIT,
340 MS4A2 and TPSB2 were up-regulated or unique expression in MBPs, indicating that
341 the cluster was consisted of mast cell and basophil progenitors (Fig. S1A). In contrast
342 to the other branch of myeloid progenitor differentiation process, the expression of
343 GATA2 and KLF1 were restrained during differentiation trajectories when the
344 progenitors were priming to neutrophilic and monocyte progenitors (NMPs, Fig. 3A,
345 B). The expression of GATA2 and CSF3R genes was shown to be the boundary for
346 GAPs and NMPs. Our study revealed the original of basophils, eosinophils, and
347 neutrophils was completely different, though the original of eosinophils was not
348 confirmed. Through the transcription factors expression switches gradually on-or-off
349 and further gradually accumulated, generally synergistic with key receptor genes
350 expression and signals activation, the original progenitor cells will be primed and
351 differentiated to various progenitors. That provided evidence of
352 megakaryocytic–erythroid lineages hierarchical organization in hematopoiesis is a
353 continuum process controlled by the activation and suppression of fate decision genes.
354

355 During lymphoid progenitor branch priming, the critical transcription factor of
356 HOXA9 and HOPX were identified to be up-regulated during CLPs differentiation
357 (Fig. 2C). Similarly, marker genes of lymphocyte lineages were found to be specific
358 presented and stepwise up-regulated, such as CSF1R, MS4A1, GNLY, and GATA3.

359 Some genes displayed co-expression among lineage restricted progenitors and CLPs.
360 These genes were more prominently specific expressed in lineage restricted
361 progenitors, such as IGHM, FCMR, BLNK, and LTB for B progenitors; SPI1, FOS,
362 PECAM1, and LST1 for monocyte progenitors; CD79A and JCHAIN for plasma
363 progenitors; and NKG7 for NK progenitors. The expression characterization of these
364 genes also indicated a differentiation correlation of lymphoid lineage commitment.
365 Lymphoid lineage commitment process is more likely a discrete stepwise process,
366 showing distinct lineage-committed populations during the hematopoiesis. That also
367 presented in monocyte lineages differentiating into macrophages and dendritic cells
368 (DCs) progenitors. Interestingly, we identified co-expression of ETV6, LYL1, and
369 NFE2 among NMPs and monocyte lineage restricted progenitors, which was same as
370 the co-expression of SPI1, FOS, PECAM1, and LST1 among monocyte and CLPs
371 (Fig. 2C, Fig. 3B,). The differentiation of each progenitor stage is controlled by the
372 expression of key regulated genes, which function like a stepwise activation switch.
373 Through the expression switches on-or-off of transcription factors, generally
374 synergistic with key receptor genes expression and signals activation, the original
375 progenitor cells will be primed and divided into the next stages of lymphoid lineage
376 progenitors. The fate decisions and lineage commitment of progenitor cells were
377 mainly controlled by a combination of critical TFs and key receptor genes (Fig. 3).
378 These findings provide strong evidence that lymphoid and monocyte lineages
379 progenitors include a gradient of heterogeneity. This finding is the first to validate that
380 the erythroid lineage process is a continuous process, while the lymphoid lineage
381 process is more likely a discrete stepwise process showing distinct lineage-committed
382 populations during hematopoiesis, opening a new perspective on hematopoiesis (Fig.
383 3C).
384
385
386 ***Essential genes in HPCs state maintain and differentiation***
387 Several fate decision factors were identified, including NFE2, LYL1, MYB,
388 CDK4, ETV6, TESPA1, ZBTB16, and GATA2 of myeloid progenitors and RUNX2,

389 HOXA9, JUN/FOS, TCF4, DDIT4, HOPX, KLF10, HOXA3, and TSC22D1 of
390 lymphoid progenitors (Fig. 3B). Our data revealed RUNX1 monitored the
391 differentiation of most progenitors (Fig. S3), while RUNX2 mainly controlled the
392 differentiation of lymphoid progenitors. Importantly, these key factors were shown to
393 play reversely functional role during the differentiation of the HPCs. Through the
394 up-regulated combination of NFE2, LYL1, MYB, CDK4, ETV6, TESPA1, GATA2
395 and in restraining the expression of RUNX2, HOXA9, JUN/FOS,TCF4, DDIT4,
396 HOPX, KLF10, HOXA3, and TSC22D1, the HPCs will diverse to differentiate to
397 myeloid progenitors. Together with the activation of other receptors and unique
398 transcript factors in each lineages progenitor, such as CSF3R for neutrophils and
399 GATA2 for GAPs, the whole hematopoietic lineage commitment was precisely
400 controlled and gradually formed (Table 1, Fig. 3C).
401
402 To confirm the essential genes in HPCs state maintain and differentiation. We
403 compared the expressions among three single cells data sheets, including bone
404 marrow cells, control PBMC cells and our CD34+ cells subsets. Our study showed
405 ITGA4 (CD49d) was highly expressed in most of the progenitors, which is key
406 molecular regulation of HSC homing [19]. Some reported HSC functional related
407 marker genes was broadly expressed in most of the progenitor cells but not specific,
408 such as PTPRC, HMGB1, CD74, CD37, PFN1, TXN, ZFP36L2, SERPINB1,
409 TMSB4X, SMIM24, MYADM, and HSP90AA1 (Table 1 and Fig. 1E)[20-26]. As
410 most of these genes were confirmed by functional validation, our data demonstrated
411 that those high expression genes of CD34 progenitor subsets were essential genes in
412 progenitor state and differentiation. These genes were shown to mediate the
413 self-renewal, differentiation, mobilization, and maintenance of hematopoietic
414 progenitor cells. In addition, two new genes of PPIA and EIF1 were identified. These
415 genes were similarly to previous reported genes that pan-expressed in progenitors and
416 highly expressed in matured subgroups of peripheral blood cells.
417 A few progenitors' specific genes of CDK6, SOX4, CD82, EGFL7, SPINT2,
418 IMPDH2, TSC22D1, CYTL1, FOXP1, SPI1[6], LY6E (SCA-2), and ETV6 were also

419 widely identified in most progenitors[27-33] (Table 1). These genes were highly
420 expressed in bone marrow and our identified progenitor cells when compared with
421 mature PBMC cells (Fig S4). This indicates those genes are essential to regulating
422 genes in HPCs state for maintenance and differentiation. And ERG, BCL11A, LMO2,
423 and RUNX1 were also widely expressed in around half percentages of cells in each
424 progenitor subgroup (Fig. S3). Interestingly, some transcription factors of them
425 showed different expression tendency in specific progenitor subgroups, such as SPI1
426 for monocyte progenitors and ETV6 for eosinophil progenitors.
427
428
429 ***Differentiation dynamics of progenitor cells***
430 To confirm the differentiation trajectory of progenitor cells, pseudotime analysis
431 was conducted using Monocle3, which revealed three distinct differentiation
432 directions of progenitor cells. The results clearly show that along the trajectory of
433 progenitor differentiation, MPCs gradually primed into three separate progenitor cell
434 branches, and later transitioned into diverse lineage stages (Fig. 4A). Common
435 myeloid progenitors were not identified among the lineage priming process.
436 Additionally, pseudotime analysis demonstrated that lineage signature genes exhibited
437 unique expression patterns along the differentiation trajectory, such as KLF1, GATA1,
438 and CSF1R, which were separated from other lineages. The expression trajectory of
439 signature genes was highly consistent with the differentiation trajectory of lineages
440 (Fig. 4B).
441 Likewise, SCENIC assesses the activity of transcription factors and regulons
442 showed the activation and expression of key transcription factors was different among
443 different progenitor subgroups (Fig. 4C). RFX2 was activated in original progenitor
444 of MPCs. The transcription factors of MAZ, HOXA9, KLF3, STAT1, JUND,
445 POLR2A, and KLF13 were identified as the key regulators in CLP priming. Several
446 TFs were identified in eosinophilic progenitors, including SF1, RUNX1, ZMIM1, and
447 BCLAF1. And GATA1, MYC, SOX4, and HMGB1 played critical role in the lineage
448 progression of megakaryocytic–erythroid lineages (Fig. 4C). XBP1 and IRF7 were

449 activated in plasma progenitor cells. IRF8, CEBPD, JUNB/FOSB, SPI1, KLF2, and
450 CEBPB were the key TFs which play the fate decision role in moncyte lineage
451 differentiation. It is clearly that the activation characteristics of TFs were similar
452 among the same original lineages, such as megakaryocytic–erythroid lineages and
453 monocyte lineages (Fig. 4C). Interestingly, unsupervised dimensionality reduction
454 analysis showed that the two nearby CLPs and Pro-eosinophilic clusters had similar
455 TFs activation characteristics, indicating that they are possible primed from the same
456 original cells (C3). The combination of multi-omics and epigenetic analysis by
457 scRNA ATAC (assay for transposase-accessible chromatin) will provide stronger
458 evidence for the fate decision process of HPCs. Our results displayed highly
459 consistent expression and activation patterns of TFs throughout the lineage priming
460 process, reflecting the synergistic effects of key regulators depression and activation.
461 These findings suggest that the original progenitor cells can be primed and
462 differentiated into various progenitors during hematopoietic lineage commitment.
463
464
465 ***Pathways of each subgroups***
466 Diversity pathways affect and control the differentiation of HPCs. Our analysis
467 results showed that JUN/FOS (AP-1), JUNB, and JUND were enriched and highly
468 expressed in most progenitors (Fig. 2C and 3B). This indicated that the JUN related
469 signaling pathway may play a critical role in the differentiation, mobilization, and
470 maintenance of hematopoietic progenitor cells. The DEPTOR gene is highly
471 expressed in clusters of MEPs (Fig. S3) and acts as a negative regulator of the
472 mTORC1 and mTORC2 signaling pathways [34]. This meaning the mTOR signaling
473 pathway should be suppressed during MEPs mature and differentiation. NFKBIA is
474 widely expressed in over 80% of the progenitor cells, while NFKBIZ is highly
475 expressed in progenitors of monocyte and macrophages (Fig. S3). NFKBIZ is a
476 member of the NF-kappa-B inhibitor family [35], indicating that the NFKB signaling
477 pathway is suppressed in progenitors, especially during monocyte and macrophages
478 mature and differentiation.

479 Genes set enrichment analysis of clusters by GSVA revealed the pathways and
480 signaling that affect the differentiation of each branch and progenitor subgroup (Fig.
481 4D). TGF-b signaling was enriched in the original progenitor cells (C0, MPCs), which
482 plays a key role in maintaining the quiescent state of HPCs[36]. Our study showed
483 that the Notch, KRAS, and TNF-α signaling pathways were enriched and may play a
484 crucial role in maintaining monocyte progenitors (Fig. 4D). Other studies have
485 indicated the novel role of Notch in regulating monocyte survival during
486 hematopoiesis[37]. Heme metabolism and hedgehog pathway gene sets were enriched
487 in erythroid progenitors. The activation of TGF-β and hedgehog pathway gradually
488 decreased during the mature stage of erythroid progenitors. Hallmark cell
489 proliferation-related gene sets (G2M Checkpoint, E2F Targets, MYC Targets V1 and
490 V2, p53 pathway) defined by MSigDB were highly enriched among erythroid and
491 neutrophilic progenitors, but the TNF signaling gene set were restrained (Fig. 4D).
492 The Wnt beta catenin signaling was enriched in NK and T cell progenitors, which was
493 reported to regulate aspects of peripheral T-cell activation and migration[38].
494 Interferon response gene sets were enriched in B cell and monocyte progenitors. It is
495 clear that the enriched gene sets among myeloid progenitors (C2, C5, C6, C13) and
496 lymphoid progenitors were different.
497
498
499 ***Fate decision genes of each redefined subgroups***
500 Advanced analysis was conducted to clearly re-define the progenitor subgroups
501 and identify fate decision genes, particularly transcription factors. Unique and specific
502 marker genes were used to define each progenitor subgroup (Fig 5.A), while negative
503 expression genes helped distinguish similar or identical original lineages subgroups
504 (Table 1). Six genes of SPINK2, HOPX, C1QTNF4, CSF3R, NPDC1, and SELL were
505 highly expressed in the original progenitors of HPCs (MPCs), but gradually restrain
506 during differentiation from GAPs to MEPs (Fig 5.B). CSF3R, S100A10, and
507 C1QTNF4 were also gradually restrained during differentiation of lymphoid early
508 progenitors (C4 cluster, Fig 5.B). Interestingly, we identified a few genes which are

509 completely contrary and specifically expressed among erythrocyte lineages (C2,C6,13)
510 and other progenitors, such as SPINK2, HOPX, C1QTNF4, and FCER1A, GATA1,
511 SLC40A1(Fig 5.A, B). This indicated the fate decision among erythrocyte lineages
512 (C2, C6, 13) and other progenitors may be controlled by those genes during the
513 lineage commitment. Two class A basic helixloop-helix protein genes of TAL1 and
514 LYL1 were highly expressed among erythrocyte lineages. HBD was the early protein
515 in erythrocyte cell development. mini-chromosome maintenance proteins gene of
516 MCM2, MCM6, MCM5, MCM4, MCM3, and MCM7 which acts as component of
517 the MCM2-7 complex, were highly expressed in erythrocyte progenitors(Pro-ME) and
518 NMPs (Table 1). MCM complex is the replicative helicase that may play an
519 essential role in erythrocyte denucleation and neutrophilic progenitors maturation.
520 LMO4 is the most specific TF gene of MBPs.

521 Three key TFs of GATA3, TCF7, and DDIT4 were identified in T progenitor cells
522 (Fig 5.C), especially the two novel expression genes of GATA3 and TCF7. These
523 three TFs play a critical role in the fate decision of T cells lineages differentiation
524 (Table 1, Fig 3B). Previous studies have reported TCF7 and GATA3 to be important
525 TFs in T cell development and differentiation[39, 40], which is highly consistent with
526 our redefined characterization of T progenitor cells. LTB was widely expressed in
527 Pro-TBN, Pro-T, and Pro-B progenitor cells, with an up-regulation in the later stages
528 of lymphoid progenitor differentiation. Two distinct types of B progenitor cells were
529 identified and redefined as having different fate decision TFs and marker genes in C9
530 and C12 clusters. This indicates the development and differentiation of B cells occur
531 early on in the progenitor stages. Cluster 9 is the progenitor of plasma cells, whereas
532 cluster 12 represents classical B progenitor cells (Table.1). Interestingly,
533 immunoglobulin variable genes were mainly expressed in the progenitors of Cluster 9
534 cells, and rearrangements of IGHV, IGKV and IGLV genes had already occurred
535 during the progenitor state of plasma cells. It's widely accepted that plasma cell are
536 the final stage of B cell proliferation and are unable to proliferate. However, our
537 single cell data confirmed the existence of plasma progenitor cells in whole blood.
538 The two factors PRDM1 and IRF4 were specifically expressed in cluster 9, while

539 PAX5 was specifically expressed in cluster 12. XBP1 was widely expressed in both
540 clusters and up-graduated in cluster 9. These four TFs factors play a critical role in the
541 fate decision of B lineage cells. HOPX was found to be very highly expressed in the
542 progenitor and mature NK cells but negative in the erythrocyte lineages (C2, C6, 13),
543 which was one of the most important TF for NK cells (Fig 3B, Fig 5A, Table 1).

544 Two types of monocyte progenitor cell in C7 and C14 clusters were identified and
545 redefined. One of the most specially expressed genes in monocyte lineages is IFI30,
546 which is related to interferon gamma signaling pathway. SPI1 and LST1, were widely
547 expressed in most progenitors but were highly expressed in the monocyte
548 subgroups(C7), which were also highly expressed in mature states of monocyte
549 original cells, such as macrophages and dendritic cells (Data not showed). This
550 suggests that SPI1 and LST1 are critical genes for monocyte progenitors. The
551 monocyte progenitors of C7 and C14 were controlled by various TFs including KLF4,
552 EGR1, MAFB, IRF8, and DDIT4 (Table 1, Fig 5B). The two clusters of C7 and C14
553 were CD86 and CD68 positive. C7 highly expressed EGR1, CEBPB, and MAFB, but
554 were negative for IRF8 and DDIT4. C7 was FCGR3A positive, which indicates more
555 likely to be macrophage progenitors, and CSF1R was novel and highly expressed in
556 that group. The cluster C7 preferred differentiation to macrophages which expressed
557 the secreted cytokine of TNF. The monocyte progenitor of C14 was positively
558 expressed CLEC9A, ANPEP, and THBD (Table 1), which indicates it is more likely to
559 differentiate into DCs. These TFs and receptors were the main fate decision factors
560 that affect the differentiation of monocyte progenitor lineages. Although we have
561 identified the fate decision factors of B cell and monocyte progenitor, the association
562 between these cell types is still unknown in the HPCs differentiation procedure.
563 Future investigation is required to determine whether the two types of progenitor cells
564 derive from the same or independent original cells. It's widely accepted that
565 neutrophilic and monocyte progenitors originate from the same myeloid progenitors.
566 However, our results showed that fate decision TFs and characterization genes among
567 the two types of progenitors were highly different, which reminds us that the original
568 and association of these two types of lineages should be reconsider.

569 *PECAM1* (CD31) was widely expressed in more than 50% of progenitors and
570 CST3 was widely expressed in more than 60% of progenitors, both of them are
571 mainly high expression in monocyte progenitors. Previous studies defined endothelial
572 progenitor cell by the markers of CD31, CD34, and THBD (CD141). Our study
573 showed one cluster (C14) cells positively expressed those markers which were
574 re-defined and confirmed as monocyte progenitor (DCs) (Table 1, Fig. 5A). That
575 reminds us the classical definition of endothelial progenitor cell should be
576 reconsideration.
577
578
579 ***Functional validation and primary culture***
580 Previously studies already conducted many functional validations of HSCs key
581 regulated genes. We hypothesis that our HPCs definitions and hematopoiesis model
582 were right, that meanings the functional studies phenotype results would be consistent
583 or can be explanation by our reaffirmed model. And furthermore, those conflicting
584 and confusing results can be clearly interpretation by the reaffirmed hematopoiesis
585 hierarchical organization model.
586 Our analysis results were found to be consistent with other functional studies
587 results. KLF1 and GATA1 were well-known play a key role in the erythroid and
588 megakaryocytic differentiation [41, 42]. The expression of GATA1 progenitors
589 generated mast cells, eosinophils, megakaryocytes and erythroid cells, and lacking
590 expression can drive generated monocytes, neutrophils, and lymphocytes[41]. Our
591 results revealed that GATA1 was unique expressed in the later stages of MEPs and
592 negatively expressed in the monocytes, neutrophils and lymphocytes progenitors (Fig.
593 2B). The result implies that the key transcript factor of GATA1 was very critical for
594 the MEPs progenitor lineages differentiation. Other evidence is transcript factor NFE2.
595 It was reported that NFE2 acts as a critical regulator in hematopoietic cells. The
596 knockout of NFE2 in mouse model displayed mild erythroid abnormalities and major
597 defect in megakaryocyte biogenesis [43, 44]. Likewise, our results revealed NFE2
598 was up-regulated in the NMPs and GAPs lineage directions (Fig. 3B).

599 Based on findings from primary culture, M-CSF was shown to be critical for the
600 survival of macrophage, which is consistent with our results that the receptor CSF1R
601 of M-CSF only expression in monocyte progenitors and matured macrophages (Fig 5.
602 D). In addition, SCGF, also known as CLEC11A, is a secreted glycoprotein that
603 exhibits erythroid burst (BFU-E) promoting activity in the absence of SCF/c-kit
604 Ligand[45, 46]. Our results showed stem cell growth factor (SCGF) was highly
605 expression in NMP and GAP progenitors, especially in the whole lineages of MEPs
606 that can explain SCGF burst promote erythroid activity. Furthermore, SCGF
607 cooperates with GM-CSF, or G-CSF to enhance the growth of
608 granulocyte-macrophage colonies (CFU-GM)[45, 46]. Because the CSF3R receptor of
609 G-CSF was highly expression in progenitors of myeloid, SCGF cooperates with
610 G-CSF would activate and induced the progenitor differentiation to neutrophils and
611 macrophages but reduce MEPs lineages generation. The IL15 receptor of IL2RB was
612 novel expressed in NK progenitors (Fig 5. D). The common receptor beta chain
613 CSF2RB of IL-3, IL-5, and GM-CSF, was novel expressed in erythroid lineage
614 progenitors. IL7R was novel expressed in T cell progenitors which is an essential
615 driver of T cell development and homeostasis. IL2RG and IL1B were expressed
616 across most progenitor subgroups (Fig 5. D). IL18 was identified as one of the most
617 essential cytokines in early stage of progenitors (Data not shown). Our study indicated
618 that cytokine IL1B plays a critical role in hematopoiesis. The transcription of IL1B
619 was regulated by TFs of SPI1 and CEBPB[47], which exhibit high expression in
620 monocytes progenitors (Table 1). ALDH1A1 was highly expressed in erythroid
621 lineage progenitors, and its product participates in retinoic acid synthesis. It's widely
622 reported that retinoic acid is required for progenitor cell maintenance [48]. The
623 expression characterization of these receptors and genes was highly consistent with
624 previous studies, providing strong evidence that our re-definition of these progenitors
625 was reliable and reasonable.
626
627
628 **Discussion**

629 Taken together, we demonstrated the transcription atlas data of original state
630 HPCs in adult peripheral blood with large number of HPCs. We redefined the
631 characterization of progenitor lineages and identified specific TFs and marker genes
632 that are critical for their development during hematopoiesis. Our transcription atlas
633 data provided novel insights into the fate decision factors and characterization genes
634 of hematopoietic progenitor cells. Our study further reaffirmed and established a
635 comprehensive model of hematopoiesis that can reconcile previous models and
636 explain conflicts among functional studies. This helps to provide a detailed insight
637 into lineage commitment from HPCs into almost branches of human peripheral blood
638 progenitors, and revealed the mechanism of HPCs self-renewal and multipotent
639 differentiation. Our single-cell transcriptional atlas provides a integrated landscape of
640 heterogeneity, development, hierarchy of the adults peripheral blood progenitor cells.
641
642 Currently, there is without a standard consensus for the definition of all HPCs
643 lineages. This is owing to the inconsistent classification with conflict functional
644 outcomes among the current literatures [10, 14]. Our study succeed enriched the large
645 numbers and almost whole spectrum of HPCs in adult peripheral blood. This will be
646 particularly attractive for large single cells studies to establish the comprehensive
647 model of hematopoiesis with clearly HPCs definition and reliable lineage
648 commitment. All of the hematopoietic progenitors were reaffirmed by unique makers
649 and key factor genes, which can be well distinguishable. We are the first clearly
650 definition all lineages of HPCs by unsupervised and inartificial clustering, which
651 provide the unambiguous transcription atlas of each type of HPCs lineages. The
652 unique markers genes of progenitor subgroups were highly consistent with previously
653 characterization descriptions, such as KLF1 of MEPs, HPGDS of mast cell, and MPO
654 of neutrophils[15, 49, 50].
655
656 It is widely known that the HPCs would differentiate into common myeloid
657 progenitors (CMPs) and common lymphoid progenitors (CLPs). Our data showed
658 none of the MME positive cells co-expressed CD38, suggesting that classification of

659 CLPs using solely CD38+ and MME+ was not sufficient. CMPs are defined with
660 CD34+CD38+MME- which is consistent with classic definition, but our single cell
661 expression atlas showed the CD34+CD38-MME+ progenitor cells is inconsistent with
662 the traditional define of CLPs[4, 5]. Based on our study, we identified subgroups of
663 progenitors by single cell sequencing which are not consistent with classical
664 definition, and showed specific classify surface or characterization markers genes,
665 such as GATA1, MPO and KLF1. Likewise, we provided a clearly evidence that the
666 classical surface markers of CD38, KIT, THY1, CD49f and FLT3 were non-specific
667 and inefficient to identify HPCs. This remind us the traditional lineage definition of
668 HSC by above markers was imperfect and need to be improved. Evidently, the
669 combination of transcriptional factors with unique marker genes would be more
670 efficient for lineage definition of HPCs.
671
672 Our classification and definition of HPCs also help to explain the conflict and
673 mechanism of *in vitro* and *in vivo* progenitor cell studies. It is well known that KLF1
674 and GATA1 are the keys regulate genes of erythrocyte or megakaryocyte
675 differentiation [51, 52]. Functional study have been shown that the upregulated
676 transcriptional factor of GATA-1 can affect mast cells, eosinophils, megakaryocytes
677 and erythroid cells generation, and downregulation to affect monocytes, neutrophils
678 and lymphocytes generation[41]. Our finding reaffirmed the definition of GAPs
679 lineages to show GATA1 to be solely expressed in the MEPs and up-regulated during
680 maturation. Previously studies reported TAL1 and LYL1 are essential for maintenance
681 of adult HSC function [53-55]. Since the induced pluripotent stem cells (iPSCs) was
682 established through activation of four transcript factors, reprogram of pluripotent stem
683 cells was widely investigated in various somatic cells[56]. GATA1 and TAL1 have
684 been validated in the reprogram of megakaryocytes and erythroid cells[51, 57].
685 Although they finally induced mature erythroid cells based on those key transcript
686 factors, but there are still many functional capacity limited during differentiation.
687
688 The CSF3R was mainly expressed in the original clusters (MPCs), which was the

689 receptor of G-CSF [58]. The underlying molecular mechanisms of G-CSF induces
690 enhancement of hematopoietic remain unclear. Our results provide a new model of
691 hematopoietic through scRNA of unperturbed and G-CSF mobilized progenitors
692 lineages in peripheral blood. All types of the progenitors were remarkable increased,
693 especially the MPCs. G-CSF directly activates the downstream activation signals of
694 HPCs through bind with CSF3R, which converted the quiescent state HPCs to
695 activation progenitors. After the mobilization procedures, HPCs was priming and
696 differentiation to all kinds of progenitors which can differentiate to mature myeloid
697 and lymphocyte cells. As key transcriptional factors and regulators controlled the
698 whole hematopoietic lineage commitment formation, the use of primary cultures also
699 showed factors of G-CSF and SCGF to induce the lineages formation. Both of these
700 mechanisms would cooperate to play key roles during the lineage commitment
701 formation. That reminds us two questions: First, how did the key transcript factors
702 was activated during the lineage commitment procedure? Secondly, the receptors
703 activated the signaling cascade and transcript factors, or the transcript factors were
704 activated independently and factors enhance the commitment formation?
705
706 Three aspects strongly support the definition and classification of HPCs to be
707 distinctly verifiable, reliable and reasonable. In brief, the unique markers were highly
708 consistent with characterization of most progenitor subgroups. Secondly, the
709 definition can very well explanation the conflict of functional studies. Thirdly, the
710 expression of distinguished receptor of cytokines can perfectly explain the primary
711 culture phenomenon of progenitor subgroups, such as CSF3R for neutrophilic and
712 CSF1R for monocyte culture[59].
713
714 Our transcriptional atlas revealed that the transcriptional factors activation and
715 suppression play a critical role in the differentiation among lineage commitment. The
716 bias during HPCs differentiation was gradually accumulated by activation of multiple
717 lineage-specifying TFs, such as NFE2, LYL1, MYB, and GATA2 of myeloid and
718 RUNX2, HOXA9, and HOPX of lymphoid progenitors. These TFs play critical role

719 during the progenitor priming with dynamic entail quantitative expression and
720 abundances change. By the multiple TFs activation and suppression, the downstream
721 signal pathway would be activated. The stepwise or continuous controlled the
722 differentiation of each branch during HPCs lineage commitment. During the process,
723 individual progenitors would gradually acquire lineage biases along multiple
724 directions during homeostasis. Our atlas results showed that key transcriptional
725 factors priming HPCs for distinct downstream lineages, which finally generated and
726 formatted the whole lineage commitment. The expression of key TFs could be
727 considered as the relative reliable boundary among stages during progenitor mature.
728 Our findings have highlighted the importance of TFs and characterization genes in
729 hematopoietic lineage fate decision, providing additional insights into the molecular
730 mechanisms underlying the hematopoietic lineage commitment. These transcription
731 factors hold great potential for reprograming progenitor cell. Through the control of
732 TFs genes, the artificial precise committed progenitor cell differentiation can be made
733 possible.
734
735 Moreover, our transcriptional atlas revealed the major role of each progenitor
736 subgroups in hematopoiesis and the fate decision of downstream various
737 hematopoietic cells. Through the expression changes of the four key transcription
738 factors of KIT, KLF1, LMO4, and GATA1, the later stage of MEPs can differentiate
739 into two directions. The expression of transcriptional factors, both in original and
740 mature stage of progenitors, revealed a basic split between myeloid-lymphocyte
741 development and megakaryocyte-erythroid development. This provides strong
742 evidence that the transcription factors were the key regulators in each stages of the
743 progenitor's differentiation. Generally, progenitor cells can be priming by key
744 transcription factor activation, which derived progenitor differentiated to downstream
745 specific mature lineages [60]. But previously study showed that transcriptional
746 programs were not strictly lineage affiliated, and expression changes between HSCs
747 and early progenitors occur independently of lineage choice[61]. That may cause by
748 not reliable classification and definition of the subgroups and different stage of

749 progenitor among lineage commitment, which based on the traditional cell surface
750 markers in most of those investigations. We demonstrated the unique markers or
751 transcriptional factors which control the progenitors differentiate to all kind of blood
752 lineages in a continuum or stepwise differentiation process. Our results strongly
753 supported that in an ideally cell fate decision model, the activation of transcription
754 factors and expression of progenitor lineage unique markers demonstrated a consistent
755 lineage affiliated with relatively obvious subgroups boundary.
756
757 It was reported the HSC lineage commitment is a continuous process in recently
758 single cell sequencing studies [5, 13]. But HSC heterogeneity and unclearly definition
759 of HSC subgroups caused the mechanism of hierarchy formation still an enigma. The
760 established hematopoietic lineage models were inconsistent and incompatible. This
761 was due to the limitation of captured cell numbers and inefficient definition by
762 identification using surface markers. Based on our clear definition and atlas of HPCs,
763 it revealed that a gradient of hematopoietic progenitors existed during the lineage
764 commitment and can be distinguished by the expression of unique TFs. Those
765 progenitors gradually acquired lineage biases, divergence and prime to each branch.
766 Our elaborate and comprehensive transcriptional atlas of HPCs can provide accurate
767 definition, hierarchy, and biological insights of hematopoietic progenitors. As our
768 study further explained how the diversified progenitor cells were primed, induced,
769 and differentiated to the whole lineages of immune system cells, this enable the
770 transcriptional atlas of HPCs to provide new perspective and well explanation about
771 how the immune was reconstituted in adult. Recently, cytokines storm was reported in
772 severe infection and coronavirus[62, 63]. Our atlas can clearly demonstrate the
773 expression cytokines receptor in specific progenitor. By controlling the original stage
774 and generation of cytokines, especially suppress the initial cytokines activation of the
775 cascade activation, this could inhibit or reduce the effect of the follow-up cascade
776 activation. Technically, this would be helpful in controlling the excessive activation
777 by cytokines and preventing the occurrence of cytokines storm as a new therapeutic
778 strategy. Our reaffirmed taxonomy will enhance more accurate prevention and

779 treatment as well as immune monitoring in health and disease.

780

781 There were some limitations in our current study. Firstly, our study did not define
782 the original progenitors with HSCs. Even though the characterization of these
783 progenitors was most likely originated with stem cells, the markers of CD49f and
784 THY1 were found to be rarely or negatively expressed. Monocyte lineage was
785 priming from myeloid progenitors, generally original from the same progenitor with
786 granulocyte. However, our study showed progenitor cells of NMPs and CLPs
787 displayed co-expression characterization among monocyte lineage restricted
788 progenitors. We speculated that another possibility that monocyte progenitor may also
789 can original from CLPs. Monocyte progenitors have two types of state progenitors,
790 which located far away from the main cell populations. We speculated two
791 possibilities to cause this phenomenon. One may be from the limitation of our
792 enrichment method. This can cause the loss of major state progenitors from these two
793 types of cells. This will have an effect to cause only the capture from the later stages.
794 Secondly, the original source of monocyte progenitors was found to be independent
795 and distinct with all of other progenitor cell types. As the whole atlas of our definition
796 was established by unsupervised dimensionality reduction and clustering from
797 statistical approach, the accuracy and efficiency should be considered. Artificial
798 intelligence and deep learning would provide a more accurate, reliable, and
799 distinguishable definition for each state and all progenitor subgroups during the whole
800 procedure of hematopoiesis.

801

802 Importantly, our study helped to re-defined hematopoietic progenitor cells and
803 reaffirming lineage commitment by transcriptional atlas. We are the first to clearly
804 depict the transcriptional atlas and defined each subgroups of adult peripheral blood
805 progenitor. Our results will enable the whole map of the transcriptional signal and
806 pathway among lineage commitment. Moreover, we are also the first to identify the
807 cell fate decision transcriptional factors in each branch among hematopoietic
808 hierarchy formation. The precise control on the hematopoietic progenitor cells

809 priming and differentiation can then be interfere by the key transcriptional factors and
810 receptors signals, which we call Artificial Operation of HPCs.
811
812 In summary, our transcriptional atlas of HPCs established a comprehensive key
813 regulators controlled model of the human adult peripheral blood hematopoiesis.
814 Toward this transcriptional atlas, it provides accurate definition, hierarchy, and
815 biological insights of hematopoiesis and lineage commitment of HPCs. Importantly,
816 this can enable new research development in cell engineering, regenerative medicine,
817 and disease treatment. Also, this can have critical insight and implications for the
818 etiology and treatment of hematologic disorders. Moreover, this opens up a wide
819 variety of possibilities for artificial precise control progenitor cells differentiation and
820 targeted treatment of immune disease, providing hope for untreatable diseases.
821

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 thanks to the laboratory support from the Key Laboratory of Shenzhen Respiratory Diseases (ZDSYS201504301616234). We also thank Chi-Wai Man of CUHK for manuscript revise.

Author Contributions

Y.Y conceived the original idea, designed the study, and performed experiments on HPC's enrichment. Y.Y and L.C conducted the bioinformatics and statistical analysis of sequencing data. Y.Y and Q.S wrote the manuscript and revised the manuscript. All authors approved the final draft of the manuscript. Q. H and G. L conducted the sample collection, and S. C draws the schematic diagram.

Declaration of Interests

The authors have declared no competing financial interest.

Declaration

The correspondence author YJ Yue would like to clarify that we already identified most marker genes, receptors, and transcript factor genes of each redefined subgroups based on the transcriptional atlas, which we didn't and can't show all of them in that version. Based on the redefined marker genes, all of the other receptors and fate decision factors can be easily identify by our transcriptional atlas of progenitors. We will present these results and reference data in the future. We welcome collaborations and funding support.

Funding

The correspondence author Dr. Yongjian Yue did not even get one coin supported from any funding or organizations in the study.

852

853 **Data Availability Statement**

854 The datasets used or analyzed during the current study are available from the
855 corresponding author on reasonable request. Anyone who want validate the
856 expression of genes in our datasets, please contact with Correspondence author.

857

858

859 **Methods**

860 **Sample collection and HPCs enrichment**

861 Totally, 8mL peripheral blood was draw from each donor. Prior to blood collection,
862 mobilizated donors will received subcutaneous recombinant human granulocyte
863 colony-stimulating factor with 300ug in each 12 hours for five days. After five days of
864 continuous treatment, a total of 10mL of peripheral blood was collected in heparinized
865 tubes from the stem cell donor, and kept at room temperature. PBMC and total white
866 blood cells were obtained according to standard protocols. The protocol of peripheral
867 blood mononuclear cell (PBMC) isolation was based on separation by density
868 gradient centrifugation with Ficoll® (Cytiva, GE), a highly branched polysaccharide.
869 Lineages (Lin−, CD3-CD19-CD56-) negative cells were removed by STEMCELL
870 reagents (15272). Self-design method was applied to enrich the HPCs. Ethical
871 approval was obtained from the ethics committee of the Shenzhen People's Hospital,
872 as per the ethical standards recommended in the 1964 Declaration of Helsinki.
873 Informed consent was obtained from the subjects in the study.

874

875

876 **Single cell RNA preparation and sequencing**

877 RNA-sequencing was also performed on human enriched HPCs (Lin−) purified from
878 donors, including three mobilization PBMC progenitor cells by G-CSF. The single
879 cells suspension was loaded and captured by Chromium microfluidic chips and
880 barcoded with a 10× Chromium Controller (10X Genomics). Barcoded single cells
881 RNA was subsequently reverse-transcribed and sequencing libraries constructed with
882 reagents from a Chromium Next GEM Single Cell 5' reagent kits v2 according to the

883 manufacturer's instructions (10X Genomics). Sequencing was conducted by Illumina
884 Nova 6000 (PE150).
885

886 **Single-cell RNA-seq data processing**

887 The raw sequencing reads of all samples were demultiplexed and aligned to the
888 reference genome hg38 by Cellranger pipeline. All downstream single-cell
889 unsupervised dimensionality reduction and clustering was conducted by Seurat[64].
890 Gene Ontology (GO) enrichment analysis of marker genes was implemented by the
891 cluster Profiler R package. Trajectory analysis was conducted by Monocle3 [65].
892 Transcription factors regulation inference and gene regulatory networks analyses
893 were performed using SCENIC [66]. Genes set variation enrichment analysis was
894 conducted by a non-parametric, unsupervised method of GSVA[67].
895

896 **Integrated analysis with online single cell dataset**

897 The online scRNA data of GSE120221[68], GSE117498[4], GSE145531[69], and
898 GSE181989[70], GSE96583[71] were downloaded from GSE database. Integrated
899 analysis was performed by Seurat. Bone marrow scRNA datasets was merged by
900 datasets of GSE120221, GSE117498, GSE145531 and GSE181989. PBMC scRNA
901 datasets were merged using data from GSE120221 and GSE96583.
902

903 **References**

904 1. Cheng H, Zheng Z, Cheng T: **New paradigms on hematopoietic stem cell differentiation**.
905 *Protein & cell* 2020, **11**(1):34-44.
906 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**
907 **transcriptomic landscape of human blood cells**. *Natl Sci Rev* 2021, **8**(3):nwaa180.
908 3. Ranzoni AM, Tangherloni A, Berest I, Riva SG, Myers B, Strzelecka PM, Xu J, Panada E,
909 Mohorianu I, Zaugg JB *et al*: **Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of**
910 **Human Developmental Hematopoiesis**. *Cell stem cell* 2021, **28**(3):472-487 e477.
911 4. Pellin D, Loperfido M, Baricordi C, Wolock SL, Montepeloso A, Weinberg OK, Biffi A, Klein AM,
912 Biasco L: **A comprehensive single cell transcriptional landscape of human hematopoietic**
913 **progenitors**. *Nature communications* 2019, **10**(1):2395.
914 5. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak
915 D *et al*: **Human haematopoietic stem cell lineage commitment is a continuous process**. *Nat*
916 *Cell Biol* 2017, **19**(4):271-281.

917 6. Sugimura R, Jha DK, Han A, Soria-Valles C, da Rocha EL, Lu YF, Goettel JA, Serrao E, Rowe RG,
918 Malleshaiah M *et al*: **Haematopoietic stem and progenitor cells from human pluripotent**
919 **stem cells**. *Nature* 2017, **545**(7655):432-438.
920 7. Zhu J, Emerson SG: **Hematopoietic cytokines, transcription factors and lineage commitment**.
921 *Oncogene* 2002, **21**(21):3295-3313.
922 8. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury
923 M, Weiner A *et al*: **Transcriptional Heterogeneity and Lineage Commitment in Myeloid**
924 **Progenitors**. *Cell* 2015, **163**(7):1663-1677.
925 9. Jacobsen SEW, Nerlov C: **Haematopoiesis in the era of advanced single-cell technologies**.
926 *Nat Cell Biol* 2019, **21**(1):2-8.
927 10. Haas S, Trumpp A, Milsom MD: **Causes and Consequences of Hematopoietic Stem Cell**
928 **Heterogeneity**. *Cell stem cell* 2018, **22**(5):627-638.
929 11. Manz MG, Miyamoto T, Akashi K, Weissman IL: **Prospective isolation of human clonogenic**
930 **common myeloid progenitors**. *Proceedings of the National Academy of Sciences of the United*
931 *States of America* 2002, **99**(18):11872-11877.
932 12. Akashi K, Traver D, Miyamoto T, Weissman IL: **A clonogenic common myeloid progenitor that**
933 **gives rise to all myeloid lineages**. *Nature* 2000, **404**(6774):193-197.
934 13. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, Majeti R, Chang HY,
935 Greenleaf WJ: **Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of**
936 **Human Hematopoietic Differentiation**. *Cell* 2018, **173**(6):1535-1548 e1516.
937 14. Laurenti E, Gottgens B: **From haematopoietic stem cells to complex differentiation**
938 **landscapes**. *Nature* 2018, **553**(7689):418-426.
939 15. Metzler KD, Fuchs TA, Nauseef WM, Reumaux D, Roesler J, Schulze I, Wahn V,
940 Papayannopoulos V, Zychlinsky A: **Myeloperoxidase is required for neutrophil extracellular**
941 **trap formation: implications for innate immunity**. *Blood* 2011, **117**(3):953-959.
942 16. Schultz J, Kaminker K: **Myeloperoxidase of the leucocyte of normal human blood. I. Content**
943 **and localization**. *Archives of Biochemistry and Biophysics* 1962, **96**(3):465-467.
944 17. Perera NC, Wiesmuller KH, Larsen MT, Schacher B, Eickholz P, Borregaard N, Jenne DE: **NSP4**
945 **is stored in azurophil granules and released by activated neutrophils as active**
946 **endoprotease with restricted specificity**. *Journal of immunology* 2013, **191**(5):2700-2707.
947 18. Huang Y, Sitwala K, Bronstein J, Sanders D, Dandekar M, Collins C, Robertson G, MacDonald J,
948 Cezard T, Bilenky M *et al*: **Identification and characterization of Hoxa9 binding sites in**
949 **hematopoietic cells**. *Blood* 2012, **119**(2):388-398.
950 19. Li D, Xue W, Li M, Dong M, Wang J, Wang X, Li X, Chen K, Zhang W, Wu S *et al*: **VCAM-1(+)**
951 **macrophages guide the homing of HSPCs to a vascular niche**. *Nature* 2018,
952 **564**(7734):119-124.
953 20. Becker-Herman S, Rozenberg M, Hillel-Karniel C, Gil-Yarom N, Kramer MP, Barak A, Sever L,
954 David K, Radomir L, Lewinsky H *et al*: **CD74 is a regulator of hematopoietic stem cell**
955 **maintenance**. *PLoS biology* 2021, **19**(3):e3001121.
956 21. Sundaramoorthy P, Wang Q, Zheng Z, Jiao Y, Chen BJ, Doan PL, Chao NJ, Kang Y: **Thioredoxin**
957 **mitigates radiation-induced hematopoietic stem cell injury in mice**. *Stem cell research &*
958 *therapy* 2017, **8**(1):263.
959 22. Zheng J, Lu Z, Kocabas F, Bottcher RT, Costell M, Kang X, Liu X, Deberardinis RJ, Wang Q, Chen
960 GQ *et al*: **Profilin 1 is essential for retention and metabolism of mouse hematopoietic stem**

961 **cells in bone marrow**. *Blood* 2014, **123**(7):992-1001.
962 23. Zhang L, Prak L, Rayon-Estrada V, Thiru P, Flygare J, Lim B, Lodish HF: **ZFP36L2 is required for**
963 **self-renewal of early burst-forming unit erythroid progenitors**. *Nature* 2013,
964 **499**(7456):92-96.
965 24. Benarafa C, LeCuyer TE, Baumann M, Stolley JM, Cremona TP, Remold-O'Donnell E: **SerpinB1**
966 **protects the mature neutrophil reserve in the bone marrow**. *Journal of leukocyte biology*
967 2011, **90**(1):21-29.
968 25. Altmann S, Lange S, Pommerencke J, Murua Escobar H, Bullerdiek J, Nolte I, Freund M,
969 Junghanss C: **High Mobility Group Box 1-Protein expression in canine haematopoietic cells**
970 **and influence on canine peripheral blood mononuclear cell proliferative activity**. *Veterinary*
971 *immunology and immunopathology* 2008, **126**(3-4):367-372.
972 26. Kruta M, Sunshine MJ, Chua BA, Fu Y, Chawla A, Dillingham CH, Hidalgo San Jose L, De Jong B,
973 Zhou FJ, Signer RAJ: **Hsf1 promotes hematopoietic stem cell fitness and proteostasis in**
974 **response to ex vivo culture stress and aging**. *Cell stem cell* 2021, **28**(11):1950-1965 e1956.
975 27. Saito-Reis CA, Balise VD, Pascetti EM, Jiminez M, Gillette JM: **Tetraspanin CD82 regulates**
976 **S1PR1-mediated hematopoietic stem and progenitor cell mobilization**. *Stem cell reports*
977 2021, **16**(10):2422-2431.
978 28. Karigane D, Kobayashi H, Morikawa T, Ootomo Y, Sakai M, Nagamatsu G, Kubota Y, Goda N,
979 Matsumoto M, Nishimura EK *et al*: **p38alpha Activates Purine Metabolism to Initiate**
980 **Hematopoietic Stem/Progenitor Cell Cycling in Response to Stress**. *Cell stem cell* 2016,
981 **19**(2):192-204.
982 29. Scheicher R, Hoelbl-Kovacic A, Bellutti F, Tigan AS, Prchal-Murphy M, Heller G,
983 Schneckenleithner C, Salazar-Roa M, Zochbauer-Muller S, Zuber J *et al*: **CDK6 as a key**
984 **regulator of hematopoietic and leukemic stem cell activation**. *Blood* 2015, **125**(1):90-101.
985 30. Doulatov S, Vo LT, Chou SS, Kim PG, Arora N, Li H, Hadland BK, Bernstein ID, Collins JJ, Zon LI
986 *et al*: **Induction of multipotential hematopoietic progenitors from human pluripotent stem**
987 **cells via respecification of lineage-restricted precursors**. *Cell stem cell* 2013, **13**(4):459-470.
988 31. Calvanese V, Capellera-Garcia S, Ma F, Fares I, Liebscher S, Ng ES, Ekstrand S,
989 Aguade-Gorgorio J, Vavilina A, Lefaudeux D *et al*: **Mapping human haematopoietic stem cells**
990 **from haemogenic endothelium to birth**. *Nature* 2022, **604**(7906):534-540.
991 32. Naudin C, Hattabi A, Michelet F, Miri-Nezhad A, Benyoucef A, Pflumio F, Guillonneau F,
992 Fichelson S, Vigon I, Dusanter-Fourt I *et al*: **PUMILIO/FOXP1 signaling drives expansion of**
993 **hematopoietic stem/progenitor and leukemia cells**. *Blood* 2017, **129**(18):2493-2506.
994 33. Yu J, Ershler M, Yu L, Wei M, Hackanson B, Yokohama A, Mitsui T, Liu C, Mao H, Liu S *et al*:
995 **TSC-22 contributes to hematopoietic precursor cell proliferation and repopulation and is**
996 **epigenetically silenced in large granular lymphocyte leukemia**. *Blood* 2009,
997 **113**(22):5558-5567.
998 34. Peterson TR, Laplante M, Thoreen CC, Sancak Y, Kang SA, Kuehl WM, Gray NS, Sabatini DM:
999 **DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and**
1000 **required for their survival**. *Cell* 2009, **137**(5):873-886.
1001 35. Totzke G, Essmann F, Pohlmann S, Lindenblatt C, Janicke RU, Schulze-Osthoff K: **A novel**
1002 **member of the IkappaB family, human IkappaB-zeta, inhibits transactivation of p65 and its**
1003 **DNA binding**. *The Journal of biological chemistry* 2006, **281**(18):12645-12654.
1004 36. Blank U, Karlsson S: **TGF-β signaling in the control of hematopoietic stem cells**. *Blood* 2015,

1005 **125**(23):3542-3550.
1006 37. Ohishi K, Varnum-Finney B, Flowers D, Anasetti C, Myerson D, Bernstein ID: **Monocytes**
1007 **express high amounts of Notch and undergo cytokine specific apoptosis following**
1008 **interaction with the Notch ligand, Delta-1**. *Blood* 2000, **95**(9):2847-2854.
1009 38. Staal FJT, Luis TC, Tiemessen MM: **WNT signalling in the immune system: WNT is spreading**
1010 **its wings**. *Nature Reviews Immunology* 2008, **8**(8):581-593.
1011 39. Zhou W, Gao F, Romero-Wolf M, Jo S, Rothenberg EV: **Single-cell deletion analyses show**
1012 **control of pro–T cell developmental speed and pathways by Tcf7, Spi1, Gata3, Bcl11a, Erg,**
1013 **and Bcl11b**. *Science immunology* 2022, **7**(71).
1014 40. van de Wetering M, Oosterwegel M, Dooijes D, Clevers H: **Identification and cloning of TCF-1,**
1015 **a T lymphocyte-specific transcription factor containing a sequence-specific HMG box**. *The*
1016 *EMBO journal* 1991, **10**(1):123-132.
1017 41. Drissen R, Buza-Vidas N, Woll P, Thongjuea S, Gambardella A, Giustacchini A, Mancini E, Zriwil
1018 A, Lutteropp M, Grover A *et al*: **Distinct myeloid progenitor-differentiation pathways**
1019 **identified through single-cell RNA sequencing**. *Nature immunology* 2016, **17**(6):666-676.
1020 42. Siatecka M, Bieker JJ: **The multifunctional role of EKLF/KLF1 during erythropoiesis**. *Blood*
1021 2011, **118**(8):2044-2054.
1022 43. erythroid abnormalitiesGasiorek JJ, Blank V: **Regulation and function of the NFE2**
1023 **transcription factor in hematopoietic and non-hematopoietic cells**. *Cellular and molecular*
1024 *life sciences : CMLS* 2015, **72**(12):2323-2335.
1025 44. Lecine P, Villeval JL, Vyas P, Swencki B, Xu Y, Shivdasani RA: **Mice lacking transcription factor**
1026 **NF-E2 provide in vivo validation of the proplatelet model of thrombocytopoiesis and show**
1027 **a platelet production defect that is intrinsic to megakaryocytes**. *Blood* 1998,
1028 **92**(5):1608-1616.
1029 45. Hiraoka A, Yano Ki K, Kagami N, Takeshige K, Mio H, Anazawa H, Sugimoto S: **Stem cell growth**
1030 **factor: in situ hybridization analysis on the gene expression, molecular characterization and**
1031 **in vitro proliferative activity of a recombinant preparation on primitive hematopoietic**
1032 **progenitor cells**. *The hematology journal : the official journal of the European Haematology*
1033 *Association* 2001, **2**(5):307-315.
1034 46. Hiraoka A, Ohkubo T, Fukuda M: **Production of human hematopoietic survival and growth**
1035 **factor by a myeloid leukemia cell line (KPB-M15) and placenta as detected by a monoclonal**
1036 **antibody**. *Cancer research* 1987, **47**(19):5025-5030.
1037 47. Pulugulla SH, Workman R, Rutter NW, Yang Z, Adamik J, Lupish B, Macar DA, el Abdouni S,
1038 Esposito EX, Galson DL *et al*: **A combined computational and experimental approach reveals**
1039 **the structure of a C/EBPβ–Spi1 interaction required for IL1B gene transcription**. *Journal of*
1040 *Biological Chemistry* 2018, **293**(52):19942-19956.
1041 48. Chute JP, Muramoto GG, Whitesides J, Colvin M, Safi R, Chao NJ, McDonnell DP: **Inhibition of**
1042 **aldehyde dehydrogenase and retinoid signaling induces the expansion of human**
1043 **hematopoietic stem cells**. *Proceedings of the National Academy of Sciences of the United*
1044 *States of America* 2006, **103**(31):11707-11712.
1045 49. Jiang J, Faiz A, Berg M, Carpaij OA, Vermeulen CJ, Brouwer S, Hesse L, Teichmann SA, Ten
1046 Hacken NHT, Timens W *et al*: **Gene signatures from scRNA-seq accurately quantify mast cells**
1047 **in biopsies in asthma**. *Clinical and experimental allergy : journal of the British Society for*
1048 *Allergy and Clinical Immunology* 2020, **50**(12):1428-1431.

1049 50. Klimchenko O, Mori M, Distefano A, Langlois T, Larbret F, Lecluse Y, Feraud O, Vainchenker W,
1050 Norol F, Debili N: **A common bipotent progenitor generates the erythroid and**
1051 **megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis**. *Blood*
1052 2009, **114**(8):1506-1517.
1053 51. Capellera-Garcia S, Pulecio J, Dhulipala K, Siva K, Rayon-Estrada V, Singbrant S, Sommarin
1054 Mikael NE, Walkley Carl R, Soneji S, Karlsson G *et al*: **Defining the Minimal Factors Required**
1055 **for Erythropoiesis through Direct Lineage Conversion**. *Cell reports* 2016, **15**(11):2550-2562.
1056 52. Wontakal SN, Guo X, Smith C, MacCarthy T, Bresnick EH, Bergman A, Snyder MP, Weissman
1057 SM, Zheng D, Skoultchi AI: **A core erythroid transcriptional network is repressed by a master**
1058 **regulator of myelo-lymphoid differentiation**. *Proceedings of the National Academy of*
1059 *Sciences* 2012, **109**(10):3832-3837.
1060 53. Bergiers I, Andrews T, Vargel Bolukbasi O, Buness A, Janosz E, Lopez-Anguita N, Ganter K,
1061 Kosim K, Celen C, Itir Percin G *et al*: **Single-cell transcriptomics reveals a new dynamical**
1062 **function of transcription factors during embryonic hematopoiesis**. *eLife* 2018, **7**.
1063 54. Souroullas GP, Salmon JM, Sablitzky F, Curtis DJ, Goodell MA: **Adult hematopoietic stem and**
1064 **progenitor cells require either Lyl1 or Scl for survival**. *Cell stem cell* 2009, **4**(2):180-186.
1065 55. Chan WY, Follows GA, Lacaud G, Pimanda JE, Landry JR, Kinston S, Knezevic K, Piltz S,
1066 Donaldson IJ, Gambardella L *et al*: **The paralogous hematopoietic regulators Lyl1 and Scl are**
1067 **coregulated by Ets and GATA factors, but Lyl1 cannot rescue the early Scl-/- phenotype**.
1068 *Blood* 2007, **109**(5):1908-1916.
1069 56. Takahashi K, Yamanaka S: **Induction of Pluripotent Stem Cells from Mouse Embryonic and**
1070 **Adult Fibroblast Cultures by Defined Factors**. *Cell* 2006, **126**(4):663-676.
1071 57. Moreau T, Evans AL, Vasquez L, Tijssen MR, Yan Y, Trotter MW, Howard D, Colzani M,
1072 Arumugam M, Wu WH *et al*: **Large-scale production of megakaryocytes from human**
1073 **pluripotent stem cells by chemically defined forward programming**. *Nature communications*
1074 2016, **7**(1).
1075 58. Beekman R, Touw IP: **G-CSF and its receptor in myeloid malignancy**. *Blood* 2010,
1076 **115**(25):5131-5136.
1077 59. Hume DA, MacDonald KP: **Therapeutic applications of macrophage colony-stimulating**
1078 **factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling**. *Blood* 2012,
1079 **119**(8):1810-1820.
1080 60. Hoppe PS, Schwarzfischer M, Loeffler D, Kokkaliaris KD, Hilsenbeck O, Moritz N, Endele M,
1081 Filipczyk A, Gambardella A, Ahmed N *et al*: **Early myeloid lineage choice is not initiated by**
1082 **random PU.1 to GATA1 protein ratios**. *Nature* 2016, **535**(7611):299-302.
1083 61. Laurenti E, Doulatov S, Zandi S, Plumb I, Chen J, April C, Fan JB, Dick JE: **The transcriptional**
1084 **architecture of early human hematopoiesis identifies multilevel control of lymphoid**
1085 **commitment**. *Nature immunology* 2013, **14**(7):756-763.
1086 62. 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**
1087 **patients infected with 2019 novel coronavirus in Wuhan, China**. *Lancet* 2020,
1088 **395**(10223):497-506.
1089 63. Fajgenbaum DC, Longo DL, June CH: **Cytokine Storm**. *New England Journal of Medicine* 2020,
1090 **383**(23):2255-2273.
1091 64. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M,
1092 Smibert P, Satija R: **Comprehensive Integration of Single-Cell Data**. *Cell* 2019,

1093 **177**(7):1888-1902 e1821.
1094 65. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen
1095 TS, Rinn JL: **The dynamics and regulators of cell fate decisions are revealed by**
1096 **pseudotemporal ordering of single cells**. *Nature biotechnology* 2014, **32**(4):381-386.
1097 66. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F,
1098 Marine J-C, Geurts P, Aerts J *et al*: **SCENIC: single-cell regulatory network inference and**
1099 **clustering**. *Nature Methods* 2017, **14**(11):1083-1086.
1100 67. Hänzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and**
1101 **RNA-Seq data**. *BMC Bioinformatics* 2013, **14**(1).
1102 68. Oetjen KA, Lindblad KE, Goswami M, Gui G, Dagur PK, Lai C, Dillon LW, McCoy JP, Hourigan CS:
1103 **Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow**
1104 **cytometry**. *JCI insight* 2018, **3**(23).
1105 69. Zhu C, Lian Y, Wang C, Wu P, Li X, Gao Y, Fan S, Ai L, Fang L, Pan H *et al*: **Single-cell**
1106 **transcriptomics dissects hematopoietic cell destruction and T-cell engagement in aplastic**
1107 **anemia**. *Blood* 2021, **138**(1):23-33.
1108 70. Tonglin H, Yanna Z, Xiaoling Y, Ruilan G, Liming Y: **Single-Cell RNA-Seq of Bone Marrow Cells**
1109 **in Aplastic Anemia**. *Front Genet* 2021, **12**:745483.
1110 71. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S,
1111 Byrnes L, Lanata CM *et al*: **Multiplexed droplet single-cell RNA-sequencing using natural**
1112 **genetic variation**. *Nature biotechnology* 2017, **36**(1):89-94.
1113
1114

1115 **List of Table and Legends**

1116 **Table 1. The selected specific marker genes and fate decision genes (partly) of all**

1117 **re-defined subgroups**

Progenitor Cell Types	Marker genes and fate decision genes (partly)	Transcription factors	Negative
Pan-Progenitors	EIF1,PPIA, HMGB1, CD74, PFN1, TXN, ZFP36L2, CD37, SERPINB1,HSP90AA1, TMSB4X, SMIM24		
Progenitors(Specific)	CD34, CDK6, SOX4, CD49d, CD82, EGFL7, CYTL1, IMPDH2, TSC22D1, SPI1(PU.1), FOXP1, EVT6, XBP1	CDK6, SOX4, IMPDH2, TSC22D1, SPI1(PU.1), EVT6, XBP1	
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
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, PRG2, CLC, 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

1118
1119

Figure Legends

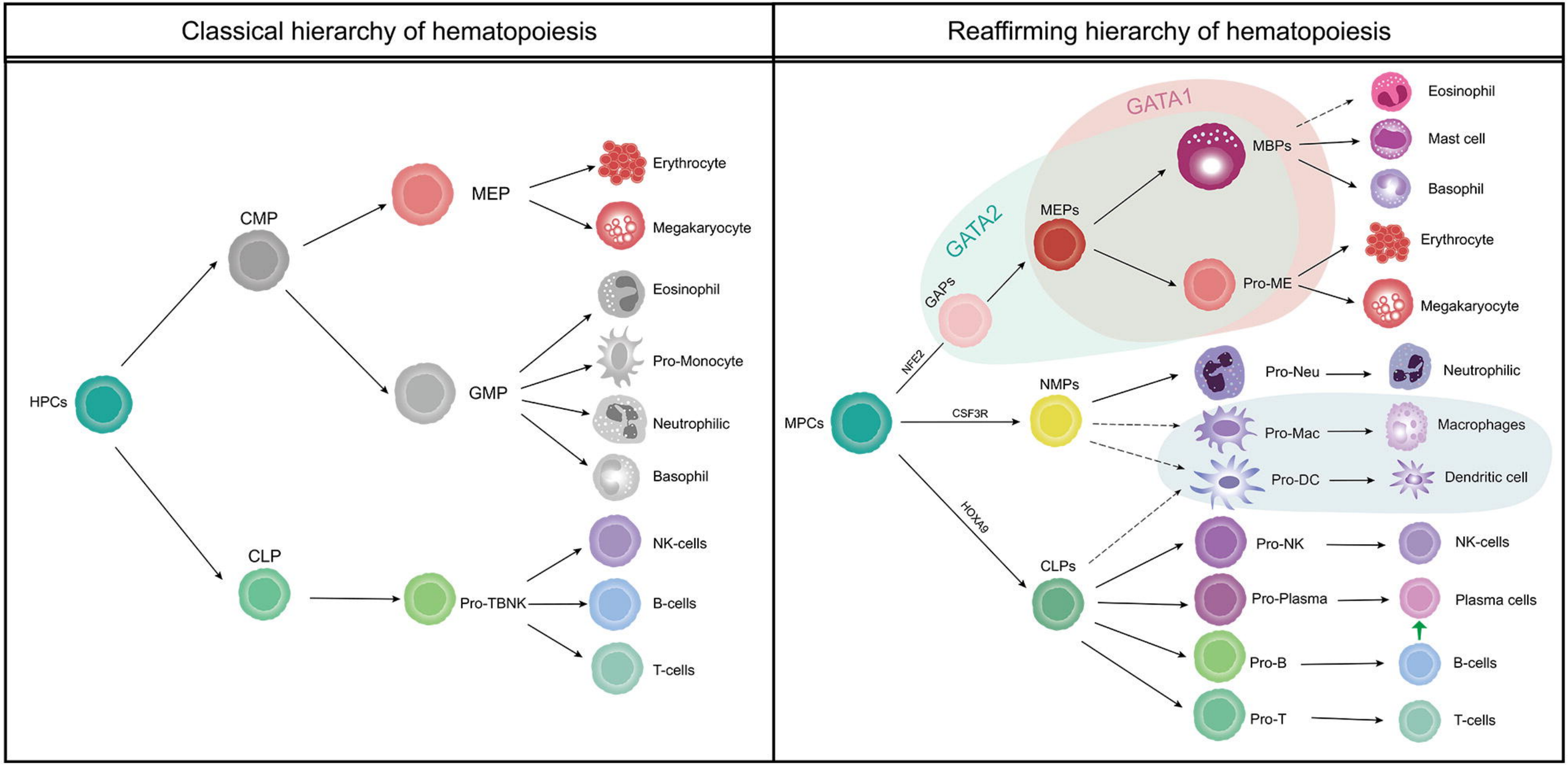
Graphical abstract: The reaffirmed lineage commitment model of HPCs

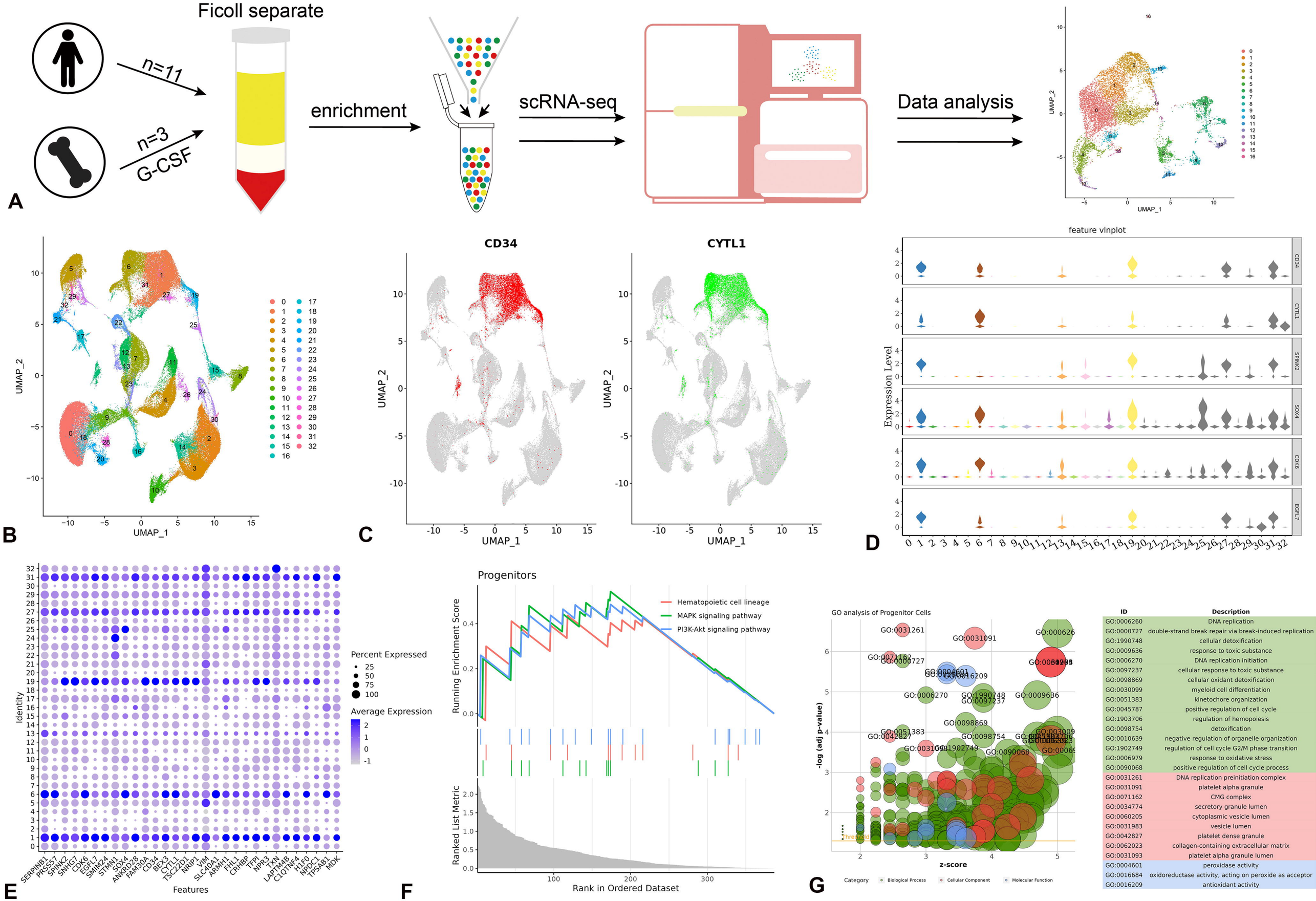
Figure 1. Progenitor subtypes were identified by unsupervised dimensionality reduction and clustering analysis in adults peripheral blood. A, Schematic showing the progenitor cells enrichment and sequencing workflow. B, Integrated 32 clusters were identified by UMAP analysis. C, Expression feature plots of progenitor markers CD34 and CYTL1. D, Violin plot presenting the expression of newly identified and known progenitor cell marker genes of CDK6, EGFL7, CYTL1, and SOX4. E, Dotplot results illustrating the expression of progenitor cells related marker genes in all clusters. F, Enrichment results of three signaling pathways in progenitors cells. G, Dotplot displaying the enriched categories of GO analysis.

Figure 2. Characterization of progenitor subtypes in quiescent and activated states. A, UMAP illustrating 17 clusters of integrated CD34 positive progenitor cells. B, Violin plots presenting the expression of traditional surface markers and unique marker genes, which demonstrated GATA2, CSF3R, and MME were more efficient and accurate markers than traditional surface markers KIT and FLT3 for myeloid, granulocyte, and lymphoid progenitor classification. C, Dot plots presenting the expression of representative marker genes 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 percentage changes of all progenitor clusters. The primitive cluster 0 and 3 exhibit the most obviously changes in cell numbers and folder change. F, Volcano plot showing the top five fold change genes in progenitor clusters.

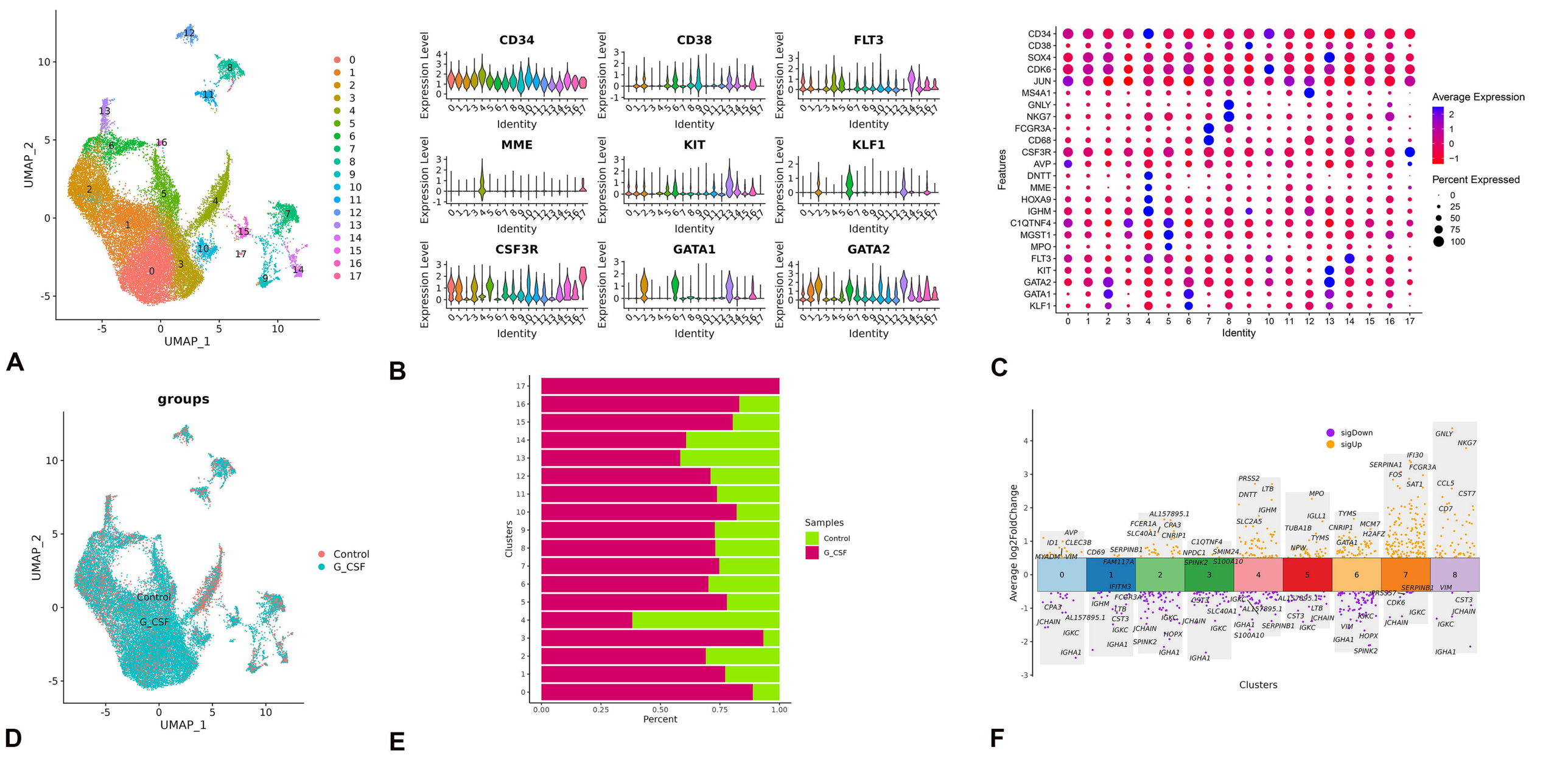
1151 **Figure 3. Newly reaffirmed hematopoietic hierarchy of hematopoietic progenitor**
1152 **cells differentiation.** A, UMAP plots demonstrating the re-defined progenitor clusters.
1153 B, Dotplots presenting the expression patterns of representative fate decision
1154 transcription factor genes of each subgroup progenitors. C, Reaffirmed lineage
1155 commitment model through scRNA transcriptional atlas. In the new hematopoietic
1156 hierarchy, the MPCs are primed into three progenitor cell types simultaneously and
1157 later mature into diverse hematopoietic lineages. The first progenitor cell type is
1158 common lymphoid progenitors (CLPs), which stepwise differentiate into T cell, B cell
1159 and NK progenitor cells. The second progenitor cell type is GAPs, which
1160 continuously differentiate into the intermediate stage of megakaryocytic–erythroid
1161 progenitors (MEPs) first, controlled by GATA2 and GATA1. In the later stages, MEPs
1162 differentiate into two directions controlled by KIT, LMO4, and KLF1. The KIT and
1163 LMO4 high expression branches will be derived to mast cell or basophil, eosinophil
1164 progenitors (MBPs). The other branch differentiates into megakaryocytic and
1165 erythroid pro-progenitors (Pro-ME). The third progenitor cell type is neutrophilic and
1166 monocyte progenitors (NMPs), which differentiate into neutrophils and monocytes.
1167 Dashed lines meaning partial lineage connections that lack strong direct evidence of
1168 differentiation relationships. The shaded area indicates the GATA2 and GATA1
1169 expression domains.
1170
1171
1172 **Figure 4. Dynamics of lineage unique marker genes and** activity **of lineage**
1173 **priming transcription factors** regulons**.** A, UMAP visualizing the differentiation
1174 trajectory of progenitor cells using pseudotime analysis with monocle3. B, UMAP
1175 displays feature plots of transcription activities for lineage signature genes related to
1176 differentiation trajectory. C, Heatmap results depicting the different activation of
1177 transcription factors regulons among all progenitor subgroups. D, Heatmap plot
1178 showing gene set enrichment analysis results of clusters by GSVA.
1179
1180

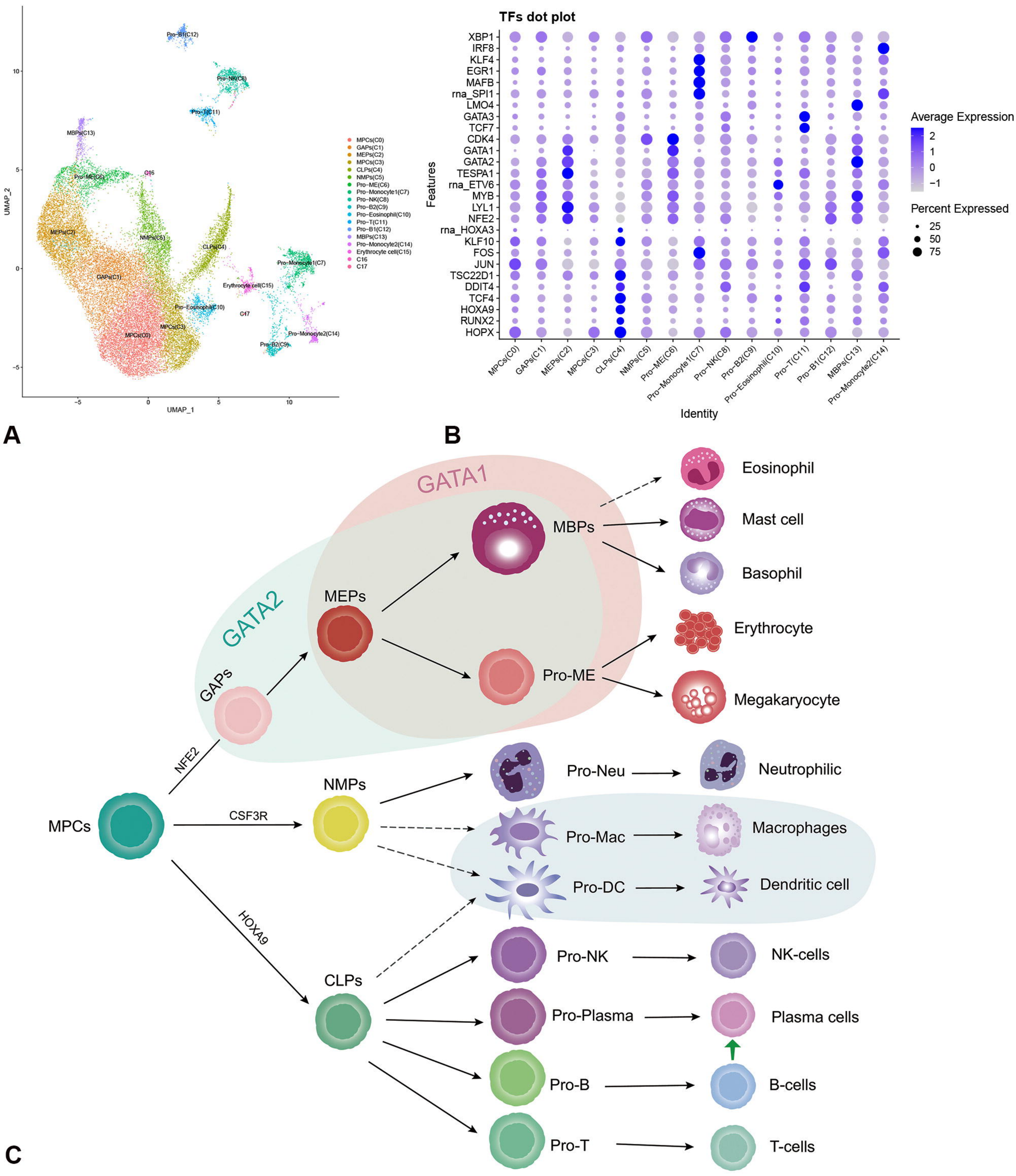
1181 **Figure 5. Expression of markers genes and fate decision genes (partly) of**
1182 **redefinition hematopoietic progenitor cells.** A, Heatmap showing the expression of
1183 marker genes and fate decision genes of re-defined progenitor clusters. B, Violin plot
1184 illustrating the expression of feature markers genes of SPINK2, HOPX, C1QTNF4,
1185 CSF3R, NPDC1, and SELL, which were highly expressed in the original progenitors
1186 of HPCs but restrained during the differentiation from GAPs to MEPs. C, Feature plot
1187 displaying the expression of LTB, TCF7, and MAFB. D, Violin plot demonstrating the
1188 expression of receptor genes, including CSF1R, IL7R, IL2RB, and others, which were
1189 specifically expressed in the redefined progenitor subgroups.

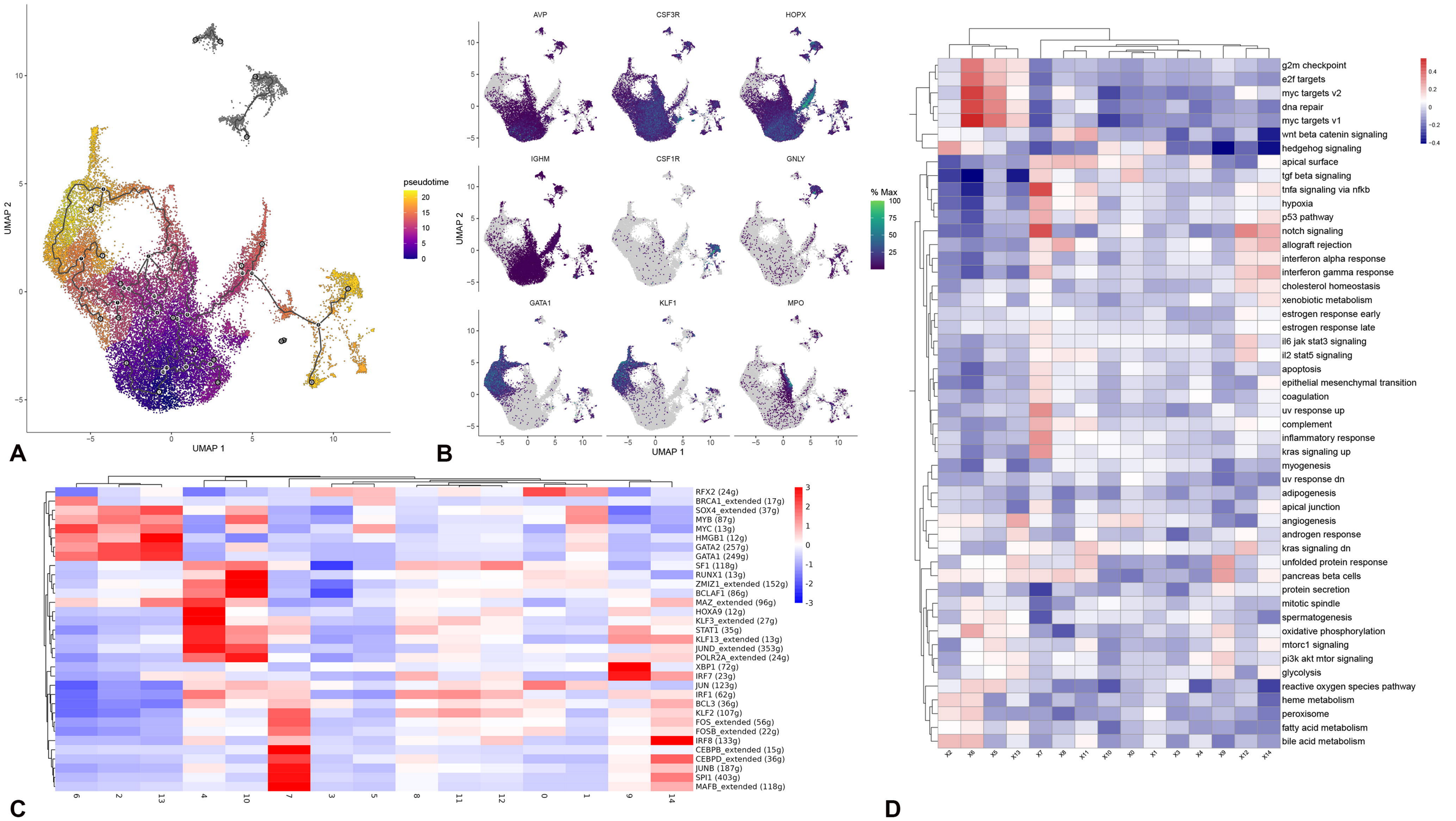

Classical hierarchy of hematopoiesis
HPCs
CMP
CLP
MEP
GMP
Pro-TBNK
Erythrocyte
Megakaryocyte
Eosinophil
Pro-Monocyte
Neutrophilic
Basophil
NK-cells
B-cells
T-cells
Reaffirming hierarchy of hematopoiesis
MPCs
NFE2
CSF3R
HOXA9
GAPs
GATA2
GATA1
MEPs
MBPs
Pro-ME
NMPs
CLPs
Eosinophil
Mast cell
Basophil
Erythrocyte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

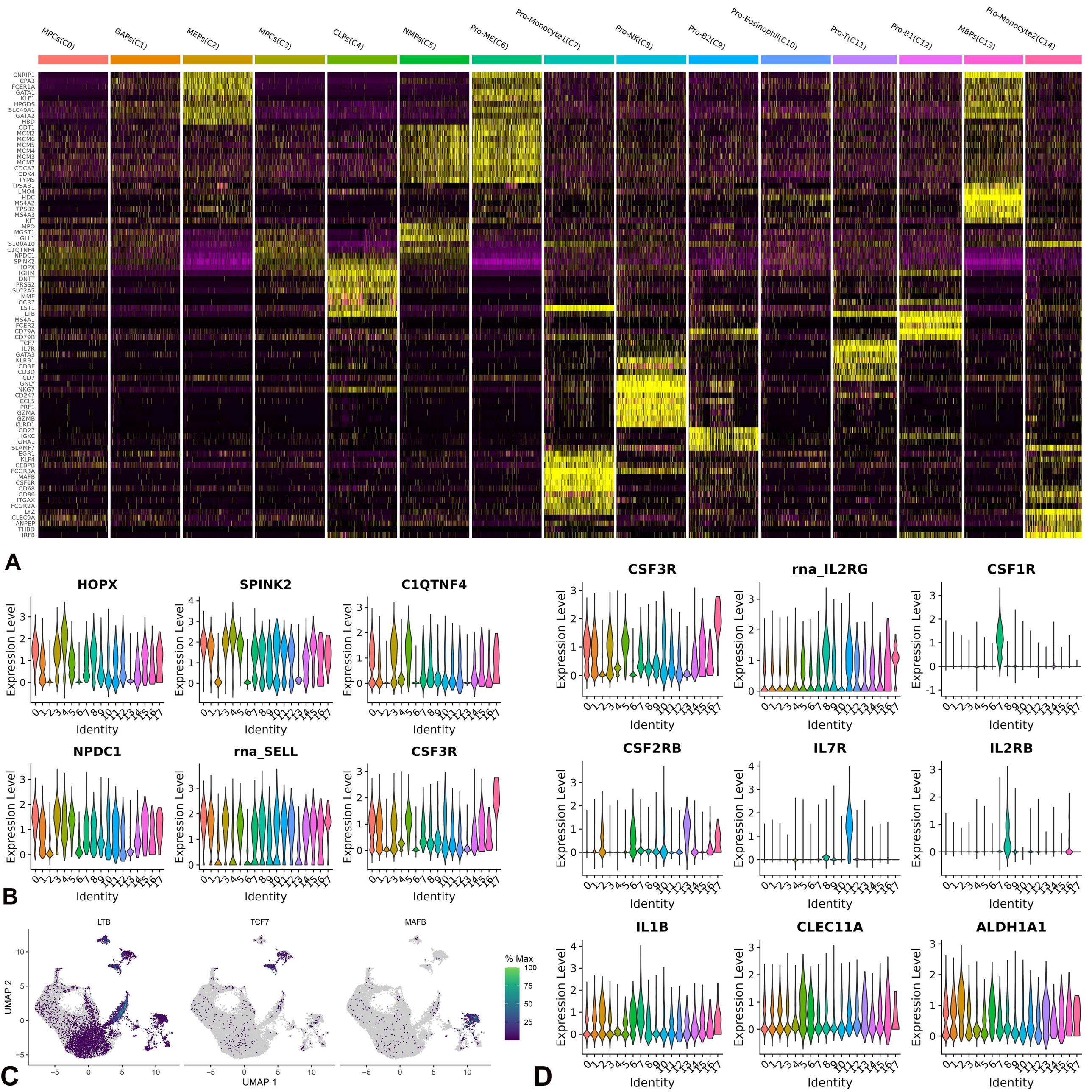


ID	Description
GO:0006260	DNA replication
GO:0000727	double-strand break repair via break-induced replication
GO:1990748	cellular detoxification
GO:0009636	response to toxic substance
GO:0006270	DNA replication initiation
GO:0097237	cellular response to toxic substance
GO:0098869	cellular oxidant detoxification
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_2
UMAP_1
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_2
UMAP_1
E
Clusters
Percent
Samples
Control
G_CSF
F
sigDown
sigUp
Average log2FoldChange
Clusters


TFs dot plot
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)
Identity
Features
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
A
B
GATA1
GATA2
MBPs
MEPs
Pro-ME
GAPs
NFE2
MPCs
CSF3R
NMPs
HOXA9
CLPs
Eosinophil
Mast cell
Basophil
Erythrocyte
Megakaryocyte
Pro-Neu
Neutrophilic
Pro-Mac
Macrophages
Pro-DC
Dendritic cell
Pro-NK
NK-cells
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