Functional stemness-related genes revealed by single-cell profiling of naïve and stimulated human CD34 + cells from CB and mPB

Guoyi Dong

College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049，China

Xiaojing Xu

College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049，China

Yue Li

Department of Hematology and Oncology, Shenzhen Children's Hospital, Shenzhen, China

Wenjie Ouyang

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

Weihua Zhao

Shenzhen Second People's Hospital, First Affiliated Hospital of Shenzhen University, Shenzhen, China

Ying Gu

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

Jie Li

College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049，China

Tianbin Liu

College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049，China

Xinru Zeng

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

Huilin Zou

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

Shuguang Wang

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

Sixi Liu

Department of Hematology and Oncology, Shenzhen Children's Hospital, Shenzhen, China

Hai-Xi Sun

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

Chao Liu (✉ liuchao4@genomics.cn)

China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

---

Research Article

Abstract

Hematopoietic stem cells (HSCs) from different sources show varied repopulating capacity, and HSCs lose their stemness after long-time *ex vivo* culture. However, the underlying mechanisms of the stemness differences because of the cell sources and the culture stimulation are not fully understood. Here, we applied single-cell RNA-seq (scRNA-seq) to analyze the naïve and stimulated human CD34+ cells from cord blood (CB) and (mPB). We collected over 16,000 single-cell data to construct a comprehensive trajectory inference map and characterized the HSC population on the hierarchy top, which is under quiescent state. Then we compared HSCs in CB to those in mPB and HSCs of naïve samples to those of cultured samples, and identified stemness-related genes (SRGs) associated with culture time (CT-SRGs) and cell source (CS-SRGs), respectively. Interestingly, CT-SRGs and CS-SRGs share genes enriched in the signaling pathways such as *mRNA catabolic process*, *Translational initiation, Ribonucleoprotein complex biogenesis* and *Cotranslational protein targeting to membrane*, suggesting dynamic protein translation and processing may be a common requirement for stemness maintenance. Meanwhile, CT-SRGs are enriched in pathways involved in glucocorticoid and corticosteroid response that affect HSCs homing and engraftment. In contrast, CS-SRGs specifically contain genes related purine and ATP metabolic process which is important to initiate hematopoiesis. Finally, we presented an application through a small-scale drug screening using Connectivity Map (CMap) against CT-SRGs and found a small molecule cucurbitacin I, targeting STAT3/JAK2, can efficiently expand HSCs *ex vivo* while maintaining its stemness. These results indicate SRGs revealed by scRNA-seq can provide helpful insights to understand the stemness differences under diverse circumstances, and CT-SRGs can be a valuable database to identify candidates enhancing functional HSC expansion during *ex vivo* culture.

Introduction

Hematopoietic stem cells (HSCs) are responsible for initiating hematopoiesis and maintaining the homeostasis of the hematopoietic system (Eaves, 2015), defined as cells on top of the hematopoietic hierarchy with totipotency (Notta et al., 2011). The molecular profile of human HSCs has been extensively investigated. Cell cycle senescence, glycolysis metabolism and self-renewal capacity are key biological signatures for human HSCs. Transcription factors, including *MEIZ1*, *TCF15* and *MLLT3*, and signal pathways, including Wnt, mTOR and HIF1, are involved in HSC specification, differentiation and maintenance (Calvanese et al., 2019; Rodriguez-Fraticelli et al., 2020). Based on the above understanding, strategies aiming to modulate those factors or signaling pathways are explored to expand and maintain the functional HSCs *ex vivo* (Wilkinson et al., 2020). Meanwhile, unbiased screenings using small molecule libraries have been performed. Several candidates capable of efficiently expanding HSCs, including SR1 and UM171, have been identified and validated in animal models (Boitano et al., 2010; Fares et al., 2014). However, the strategies and small molecules mentioned above have not been successfully applied under clinical settings, suggesting other important underlying mechanisms required for HSC maintenance have not yet been identified.

In human, CD34$^+$ cells isolated with magnetic beads containing the HSCs are clinically used for hematopoietic stem cell transplantation (HSCT) and hematopoietic stem cell gene therapy (HSC-GT). Increasing evidence shows that CD34$^+$ cells are a heterogeneous cell population at different levels (Crisan and Dzierzak, 2016; Haas et al., 2018). Generally, to fully explore the HSC complexity, single-cell RNA sequencing (scRNA-seq) has been intensively used in this field with two approaches: one approach is to isolate the HSCs and other progenitor cells using fluorescence assisted cell sorting (FACS) with different antibodies against cell surface markers, and then perform the scRNA-seq in those "characterized" cell populations, respectively (Pellin et al., 2019); another approach is to perform scRNA-seq on unsorted CD34$^+$ cells and then define the sub-cell populations based on bioinformatic analysis (Zheng et al., 2018). However, as the first approach still relies on the expression of surface markers, which may introduce bias in cell definition, it may not fully cover and reveal the *bona fide* cell populations when considerable cells are discarded during FACS process. Moreover, recent studies have found HSCs may also exist in other cell populations, suggesting the characterization based on cell surface markers are not sufficient. By contrast, the second approach, in theory, is able to capture all kinds of cells in an unbiased way, by which cell populations are clustered and defined by their gene expression profiles. However, when investigating cell sub-populations, particularly for HSCs, few known marker genes are reported and functional validation is lacking, making the reliability of these results questionable.

It is well-known that CD34$^+$ cells derived from cord blood (CB) or mobilized peripheral blood (mPB) have different hematopoietic regeneration capacities (Mayani et al., 2020). For example in clinical practice, the minimal number of mPB CD34$^+$ cells required for allogenic HSCT is usually above 2×10^6/kg. By contrast for CB CD34$^+$ cells, a quarter population (0.5×10^6/kg) is sufficient for allogeneic cell transplantation (Remberger et al., 2015; Wagner et al., 2002). In addition, when generating humanized mice model, the required human CD34$^+$ cells are 0.5×10^5 of CB or 1.0×10^6 of mPB cells per mouse(Wang et al., 1997), further indicating the higher stemness of CB than mPB CD34$^+$ cells. On another aspect, the capacity of HSCs to rebuild the hematopoiesis is decreased along with the *ex vivo* culture time, as HSCs lose their stemness when they undergo activation and differentiation during the culture period (Glimm et al., 2000). Therefore, comparing CD34$^+$ cells from CB and mPB sources, as well as from naive and cultured conditions would shed light on the signature genes and signaling pathways of HSCs.

Here we collected CB and mPB CD34$^+$ cells from independent individuals, conducted the scRNA-seq immediately or after 48 hours *ex vivo* culture. We characterized HSC population and identified two sets of stemness-related genes (SRGs) termed as CT-SRGs and CS-SRGs, associated with culture time and cell source, respectively. Using CT-SRGs to perform CMap searching, we found small molecule cucurbitacin I can efficiently expand HSCs *ex vivo* while maintaining its stemness. Our results demonstrate SRGs revealed by scRNA-seq can provide helpful insights for understanding the stemness maintenance of HSCs.

Results

Identification of cell populations in human CD34$^+$ cells

To characterize the human HSC population and its functional signaling pathways, we collected CB and mPB CD34$^+$ cells from fresh (termed naïve) or cultured (termed stimulated) conditions and captured their single-cell transcriptomes using a massively parallel single-cell library preparation technique, DNBelab C4 (Liu et al., 2019). With the comprehensive scRNA-seq dataset, we performed bioinformatic analysis to idendify HSC stemness related genes (SRGs) associated with culture time and cell source. Furthermore, we functionally validated the candidate small molecules predicted to regulate SRGs through *ex vivo* CD34$^+$ cell culture and following FACS and scRNA-seq analysis. The overall experimental design is shown in Fig. 1a (Fig. 1a).

We collected 2 biological replicates for each group, and after quality control we obtained 16,196 cells in total, with an average of ~ 3300 genes (~ 11,600 UMIs) per cell (Fig. S1a). Global correlation analysis revealed a strong correlation between biological replicates, and samples in the same culture conditions had higher transcriptome similarities than samples from the same source (Fig. S1b). To get an integrated single-cell transcriptome map of CD34$^+$ cells, we performed graph-based clustering of the dataset, and found that almost all cells (99.8%, 16,171 of 16,196) were CD34 positive (Fig. S1c), which was consistent with our CD34$^+$ cell sorting procedure (Fig. S1d). In summary, these results demonstrated that our transcriptome dataset had good quality for the subsequent analysis.

Next, we separated the cells into 12 populations and collected the top highly expressed marker genes to define cell types (Table. S1 and Table. S2). Based on these marker genes, we were able to assign the populations with distinct cell identities, including hematopoietic stem cells (HSCs), multipotent progenitors (MPPs), as well as myeloid, erythroid and lymphoid lineages (Fig. 1b). Many previously reported marker genes were also confirmed in our data (Fig. 1c, Fig. S2 and Table. S1), such as *AVP*, *MLLT3*, *HLF* and *CRHBP* of HSCs (Calvanese et al., 2019; Karamitros et al., 2018; Komorowska et al., 2017; Ranzoni et al., 2021; Xie et al., 2021; Zheng et al., 2018); *ZBTB16* of lymphoid-primed multipotent progenitors (LMPPs) (Karamitros et al., 2018); *TUBB*, *DUT* and *TUBA1B* of megakaryocyte-erythroid-mastcell progenitors (MEMPs) (Popescu et al., 2019); *MPO* and *LYZ* of granulocyte-monocyte progenitors (GMPs) (Karamitros et al., 2018; van Galen et al., 2019; Zheng et al., 2018); *HBD* and *GATA2* of megakaryocyte-erythroid progenitors (MEPs) (Karamitros et al., 2018; Zheng et al., 2018); *IGKC* and *MS4A1* of B cells progenitors (ProBs) (Popescu et al., 2019). To further validate the accuracy of cell type identification, we used a hypergeometric distribution test to evaluate the consistency between marker genes of cell clusters in our data and the top 500 up-regulated genes of cell types in eight published papers (Doulatov et al., 2010; Kohn et al., 2012; Laurenti et al., 2013; Laurenti et al., 2015; Lee et al., 2015; Milyavsky et al., 2010; Ng et al., 2009; Novershtern et al., 2011; Xie et al., 2021). By this analysis, we also observed a high consistency between these genes, providing further evidence supporting the cell type identification (Fig. S3).

Functional enrichment analysis of the marker genes further confirmed the characteristics of these cell types. For example, up-regulated genes in HSC/MPPs were related to cellular stress response as well as purine metabolic process, including *Response to glucocorticoid*, *Response to corticosteroid* and *Response to steroid hormone* that are reported as signature pathways enriched in HSCs in previous studies (Guo et al., 2017; Huang and Broxmeyer, 2019; Nakamura-Ishizu et al., 2020), whereas downstream progenitors were enriched for cell differentiation and cell activation related pathways in agreement with the cell development process and cell-cycle progression (Fig. 1d and Table. S3). Taken together, we obtained high-quality scRNA-seq data from 16,196 sorted CD34$^+$ cells from naïve and stimulated CB and mPB, and identified 12 cell types including HSCs consistent with previous reports, providing a comprehensive reference map for investigating the underlying mechanism of stemness maintenance of HSCs.

Differentiation trajectory of human HSCs

Previous studies found that HSCs first differentiated into MPPs, and then into LMPPs and other progenitor cells (Laurenti and Gottgens, 2018; Ng et al., 2009). Consistently, Principal Component Analysis (PCA) revealed that the cell types defined as adjacent developmental states in our map were clustered together, such as HSCs, MPPs and LMPPs which were all upstream progenitors (Fig. S4a). To further validate the accuracy of our reference map, we used Monocle (Trapnell et al., 2017) to conjecture the differentiation trajectory and checked whether our cell types exhibit similar pattern (Fig. 2a). Seven state cells were distributed along the trajectory and we found most HSCs/MPPs were located near the tips of the trajectory, while other cells were distributed amongst the six branches (Fig. 2b and Fig. S4b), in agreement with previous report. To determine the lineage affiliation of these branches, we checked the expression patterns of some marker genes. As we expected, the expression levels of *AVP, HLF* and *VIM*, all related to self-renewal potential and quiescence in HSC and MPP (Komorowska et al., 2017; Xie et al., 2021; Zheng et al., 2018), were decreased over the pseudo time. *ZBTB16* and *MZB1*, the marker genes of LMPP and MLP, were upregulated at the early stage but then dereceased later along the pseudo-time, in consistent with the position of LMPP and MLP on the trajectory. Myeloid and erythroid lineages such as MEMP, CMP, GMP and MEP were mainly situated at the end of the trajectory (State 5–7) with high expression of their lineage marker genes such as *DUT, CENPF, MPO* and *HBD* (Fig. 2c-d, Fig. S4b).

To further confirm the differentiation trajectory inferred, we used RNA velocity (La Manno et al., 2018) to project all cells into a two-dimensional UMAP space with arrows showing the direction and the speed of differentiation (Fig. 2e). When a developmental routine was finally fitted (Fig. 2f), we found that it was consistent with the current classical model of lineage determination in human hematopoietic hierarchy, thus further illustrating the accuracy of cell annotation and differentiation trajectory of our data.

The characteristics and gene regulatory networks of human HSCs

To further characterize the human HSCs in our data, we did GO term analysis of the up-regulated genes of HSCs. We found pathways associated with cell cycle as well as mitosis, such as *regulation of spindle checkpoint*, *regulation of cell cycle spindle assembly checkpoint* and *mitotic cell cycle arrest*, were present

as the most significant ones, which also includes other GO terms such as *response to glucocorticoid* and *response to corticosteroid* (Fig. 3a). The cell cycle activity of HSCs over the lifetime is dynamic, and it reflects the requirements of the organism at different developmental points (Gudmundsson et al., 2020; Lu et al., 2020). Thus, we calculated cell cycle phase scores based on canonical markers by Seurat (Stuart et al., 2019), and found that HSCs were significantly enriched in G0 and G1 phase (96.4%, *P* value = 5.62E-227, Fig. 3b and Table. S4) when compared with other cell types, indicating that most HSCs were in a resting state, in consistant with GO term analysis.

Subsequently, we wondered whether some gene regulatory networks (regulons), a collection of genes regulated by common transcript factors (TFs), specifically existed in HSCs. To achieve this, we applied SCENIC (Aibar et al., 2017) to each cell of the 12 identified cell types. SCENIC recognized 371 activated regulons whose activities were dynamically changed among different cell types (Fig. 3c and Fig. S5). Interestingly, expression of TFs in several identified regulons, like *HLF*, *MECOM* and *ELK3*, were also enriched in HSCs and MPPs (Fig. 3c-d and Table S2). In addition, the functions of these TFs and their target gene sets were significantly enriched in *Response to corticosterone* and *Response to purine-containing compound* (Fig. 3e), which were consistent with those of HSC marker genes (Fig. 3a and Fig. 1d). Thus, using regulons analysis, we confirmed previous reported TFs, such as HLF and MECOM (Maicas et al., 2017), and their target genes, were also presented in HSCs of our data (Komorowska et al., 2017; Zheng et al., 2018). Besides, we also found many novel regulons worthy of further investigations, including *FOS*, *MEF2C*, *TEAD4*, *ELK3* and *HOXA9* (Fig. 3c-d).

Identification of stemness-related genes (SRGs) probably controlling stemness of human HSCs

HSCs accounted for 9.9% of all cells profiled in our data, and we found that their composition was significantly decreased under stimulated condition in both CB and mPB samples (Fig. 4a), which was in agreement with the reduced stemness of these samples. Interestingly, we noticed that cell sources and 2 days *ex vivo* culture didn't affect the cell cycle phase scores (Fig. 4b), suggesting quiescent maintenance is a robust characteristic for all HSCs, even after short time *ex vivo* culture. To reveal SRGs responsible for stemness maintenance of HSCs, we investigated the transcriptome changes between naïve and stimulated samples. By differential expression analysis, we obtained 247 CT-SRGs (Culture Time-related SRGs), from the intersection of 716 up-regulated genes in CB CD34$^+$ naïve and 1,164 up-regulated genes in mPB CD34$^+$ naïve (*P* value < 0.05 and ln-transformed fold change > 0.25) (Fig. 4c-e and Table. S5-7). In addition, as previous studies found that CD34$^+$ cells derived from CB exhibited a higher level of stemness than mPB (Eliane et al., 2011; Martínez-Jaramillo et al., 2004; Wang et al., 2003), we also obtained 560 CS-SRGs (Cell Source-related SRGs) by comparing naïve CB and naïve mPB (*P* value < 0.05 and ln-transformed fold change > 0.25) (Fig. 4f and Table. S8). Gene set variation analysis (GSVA) using HSC data sets from two published papers (Milyavsky et al., 2010; Novershtern et al., 2011) revealed that the overall expression levels of both CT-SRGs and CS-SRGs were significantly higher than HSC lineage distinct genes (HSC-LDGs) (Fig. 4g-h), further demonstrating that both CT-SRGs and CS-SRGs could better reflect the stemness of HSCs.

Then we asked the commonalities and differences between CT-SRGs and CS-SRGs. Interestingly, a large fraction of the intersection genes in CT-SRGs and CS-SRGs are enriched in GO terms related to *Protein targeting to membrane* (P value (CT-SRGs) = 1.94E-50, P value (CR-SRGs) = 2.25E-34) and *Protein targeting to ER* (P value (CT-SRGs) = 1.48E-48, P value (CR-SRGs) = 3.26E-31) (Fig. 4i), which are signatures for protein synthesis process and translation of ribosomal coding genes. Next, we It is worth noting that nine genes specifically present in CT-SRGs, including *AREG*, *CFLAR*, *DDIT4*, *DUSP1*, *FLT1*, *FOS*, *FOSB*, *FOXO3* and *ZFP36*, were significantly enriched in GO terms of *Response to glucocorticoid* (P value = 8.50E-06) and *Response to corticosteroid* (P value = 1.96E-05) (Fig. 4j). By contrast, we found "*Purine nucleoside triphosphate metabolic process* (P value = 4.90E-12) and *ATP metabolic process* (P value = 4.48E-10) were specifically enriched in CS-SRGs (Fig. 4k). Taken together, these results demonstrated that both CT-SRGs and CS-SRGs exhibited a better consistency with HSC characteristics than HSC marker genes, indicating their potential applications in stemness maintenance of human HSCs.

The differentiation trajectory of human CD34⁺ cells reconstructed using CT-SRGs

During HSCT and HSC-GT, the ability of CD34⁺ cells to rebuild the hematopoiesis is negatively correlated with the *ex vivo* culture time, therefore to understand the transcriptomic changes during this process, we focused on CT-SRGs in further analysis. First, we asked whether we can reconstruct the differentiation trajectory of human CD34⁺ cells using CT-SRGs (Fig. 5a). Consistent with the trajectory conjectured based on marker genes (Fig. 2), HSCs/MPPs were still located at the top of the trajectory. Additionally, we were amazed to find that lymphoid lineage (State2) and myeloid lineage (State3) were clearly separated in the developmental trajectory after HSCs/MPPs (State1) (Fig. 5b and Fig. S6a), indicating that CT-SRGs had different expression patterns when HSCs differentiate into these two lineages. We further checked the expression changes of marker genes along pseudo-time. *AVP*, *HLF*, *VIM* and *KLF6* of HSCs/MPPs were highly expressed in State1, then decreased with the pseudo-time progression. Oppositely, *MKI67* and *HBD* gradually increased with pseudo-time and reached the peak in State3, signifying myeloid and erythroid lineages may enrich in State3. State2 might be lymphoid lineage with high expression of *IGKC* and *MS4A1* (Fig. S6b). Next, we identified differentially expressed genes between the branches to further corroborate the previous results, and we got three gene clusters with different expression patterns across three States. GO Term analysis of the three gene clusters indicated related pathways were enriched in corresponding states, such as *lymphoid differentiation* in Lymphoid, *cotranslational protein targeting to membrane* in HSCs/MPPs and *neutrophil activation* in Myeloid&Erythroid, further confirming the reliability of these results (Fig. 5c). A total of 15 CT-SRGs were up-regulated in State2 (lymphoid progenitors), whereas no CT-SRG was up-regulated in State3 (Fig. 5d), indicating that lymphoid progenitors may be more close to HSCs when compared to myeloid or erythroid cells. In agreement, the 15 CT-SRGs upregulated in lymphoid were also highly expressed in MPPs and HSCs (Fig. 5e). In conclusion, CT-SRGs may be better reference genes for development trajectory construction, revealing a closer relationship between HSCs and lymphoid lineage.

Small molecules modulating CT-SRG expression promote cell proliferation and stemness maintenance of human HSCs *ex vivo*

Connectivity Map (CMap) is an online tool kit based on a perturbation-driven gene expression dataset (Lamb et al., 2006). To functionally validate the effectiveness of CT-SRGs in maintaining stemness of human HSCs, we used CMap to search for candidate small molecules that could affect the global expression levels of CT-SRGs and then performed the experimental validation process (Fig. 6a). We identified 145 candidates in total. Among them, small molecules function as protein synthesis inhibitor and glucocorticoid/corticosteroid receptor agonist were predicted to positively regulate the expression levels of CT-SRGs, consistent with the above results showing that the functions of CT-SRGs were enriched in protein synthesis process and glucocorticoid/corticosteroid responses. In addtion, small molecules that target ATPase, mTOR and MAP kinase pathways, were also characterized as candidates in our screening (Table S9). Importantly, fenretinide, a retinoid receptor agonist and identified in our CMap screening, has been reported to enhance human HSC self-renewal by modulating sphingolipid metabolism *ex vivo* (Xie et al., 2019), demonstrating small molecules predicted to modulate CT-SRGs can be candidates capable of regulating human HSC stemness.

To identify other regulators of HSC stemness, we selected NVP-BEZ235, Cucurbitacin I and Calmidazolium as these molelues were related to the signal of mTOR inhibitor, PI3K inhibitor, Protein kinase inhibitor, Inhibitor of STAT3/JAK2 signaling, JAK inhibitor, Calcium channel blocker, calmodulin antagonist (Fig. 6b). We treated CB and mPB CD34$^+$ cells *ex vivo* with those small molecule candidates, cultured them for extended time period and then measured the cell proliferation and HSCs percentage. Fenretinide was used as a positive control. When compared to fenretinide treatment and untreated control, only Cucurbitacin I, but not other tested molecules, increased the CD34$^+$ cell numbers in CB and mPB after 6 days' *ex vivo* culture (Fig. 6c-d). More important, when examining a more restricted surface markers of HSCs using FACS, the cucurbitacin I treated cells also exhibited the highest proportion of CD34$^+$CD38$^-$ cells, indicating that cucurbitacin I not only enhances the expansion but also promote human HSC maintenance of both CB and mPB sources (Fig. 6e-f). Taken together, we provided a preliminary application of CT-SRGs to identify HSC modulators and validated that small molecule cucurbitacin I could enhance cell proliferation and stemness maintenance of human HSCs.

Discussion

Many previous studies have demonstrated that using scRNA-seq transcriptomic data ,a marker-free approach, can capture detailed molecular characterization of single cells and unbiased define cell clusters during hematopoiesis (Giladi et al., 2018; Zheng et al., 2018). Using this method, many novel cell types have been found and functional identification (Villani et al., 2017). In our present study, we used scRNA-seq data from CD34$^+$ cells to accurately identify all classical human hematopoietic stem and progenitor cells (HSPCs) cell types reported in previous researches (Laurenti and Göttgens, 2018b), including HSCs, without prior specific cell surface markers sorting. It is worth mentioning that our data shows better data quality, being embodied in a strong correlation between biological replicates and a

more credible cell type identification. Based on these molecular data, we found that most HSCs stay in G0 and G1 phase, indicating that cells were in a quiescent state, those critical transcription factors for the maintenance and proliferation of HSCs, such as *AVP*, *MLLT3*, *HLF*, *MECOM*,*CD52*, are highly expressed.

HSCs from CB and mPB process distinct ability of reconstruction of hematopoiesis and loss stemness *in vitro* culture. Uncovering the molecular distinction is helpful to find out the underlying mechanisms of HSC maintenance and proliferation. We roundly compared HSCs in CB to those in mPB and HSCs of naïve samples to those of cultured samples, and identified SRGs associated with culture time (CT-SRGs) and cell source (CS-SRGs), responsible for stemness maintenance and hematopoiesis initiation, respectively. We found that naive CB CD34+ may be able to both rapidly colonize and maintain stemness, whereas mPB CD34+ may tends to be naive CB CD34 colonization state with activation on the one hand, and loses glucorticoid pathway compared to naive CB CD34+ on the other hand. Meanwhile, we verified the reliability of the method to screen stemness-related genes.

One obstacle limiting the wide application of HSCs in clinics is the lack of effective methods to culture and expand HSCs *ex vivo* while maintaining its stemness. Thus we focused on investigating CT-SRGs and identified new molecules using CMap analysis. Interestingly, we found that small molecule cucurbitacin-I could boost cell proliferation and stemness maintenance of human HSCs. Cucurbitacin I is a STAT3/JAK2 inhibitor, however, its function in human HSCs expansion and stemness maintenance has not been clarified. In this study, we suggest Cucurbitacin I may be a novel therapeutic molecule to expanding HSCs for HSCT and HSC-GT in future clinical applications.

Methods

Enrichment of CD34+ cells from human CB and mPB samples

Human CB and mPB samples were obtained with informed consent from health donor. Mononuclear cells (MNC) were obtained by centrifugation on Lymphoprep medium, and after ammonium chloride lysis MNC was enriched for CD34+ cells by positive selection with the CD34 Microbead kit and LS column purification with MACS magnet technology (Miltenyi). The sorted CD34+ cells were subject to downstream experiments.

Cell cultivation and scRNA-seq

Fresh CD34+ cells were immediately cultured in vitro or single-cell RNA-seq. For cell cultivation, CD34+ cells were resuspended in SCGM medium (Cellgenix) with following recombinant hematopoietic cytokines: Recombinant Human stem cell factor (rhSCF) 100ng/ml, Recombinant Human Thrombopoietin (rhTPO)100ng/ml, Recombinant Human Fms-related Tyrosine Kinase 3 Ligand (rhFlt3-L) 100ng/ml,, and cultured in 24-well tissue culture plates at 37°C in an atmosphere of 5% CO_2 in the air for 48 hours and collected to scRNA-seq. For scRNA-seq were performed by the DNBelab C4 platform (Liu et

al., 2019). In brief, single-cell suspensions were used for droplet generation, emulsion breakage, beads collection, reverse transcription, and cDNA amplification to generate barcoded libraries. Indexed libraries were constructed according to the manufacturer's protocol. The sequencing libraries were quantified by QubitTM ssDNA Assay Kit (Thermo Fisher Scientific, #Q10212). DNA nanoballs (DNBs) were loaded into the patterned nano arrays and sequenced on the ultra-high-throughput DIPSEQ T1 sequencer using the following read length: 30 bp for read 1, inclusive of 10 bp cell barcode 1, 10 bp cell barcode 2, and 10 bp unique molecular identifier (UMI), 100 bp of transcript sequence for read 2, and 10 bp for sample index.

Quality control of scRNA-Seq data

The iDrop Software Suite (v.1.0.0) was applied to perform sample demultiplexing, barcode processing, and single-cell 3' unique molecular identifier (UMI) counting with default parameters. Processed reads were then aligned to the complete UCSC hg38 human genome by splicing-aware aligner STAR (Dobin et al., 2013) with default parameters. Gene-cell metrics were generated for advanced analysis of valid cells that were automatically recognized according to the UMI number distribution of each cell.

The R (v.3.6.3) package Seurat (v.3.2.1) (Butler et al., 2018; Stuart et al., 2019) was used to perform the following steps: 1) Quality control of three indicators: the number of genes expressed per cell, the number of UMI (Unique Molecular Identifiers) and the proportional distribution of mitochondrial RNA to screen high-quality cells for subsequent analysis(Griffiths et al., 2018; Ilicic et al., 2016). Due to the differences among samples, for example, cryopreserved (naïve) and stimulated, the number of genes expressed varied greatly, thus we selected Tukey's test method (Kannan et al., 2015) to remove cells with abnormal gene numbers. Cells that expressed genes lower than Q1-IQR or higher than Q3 + IQR were removed. Meanwhile, cells with a mitochondrial mRNA ratio greater than 10% were also removed; 2) Doublets removal. We used an R package DoubleFinder (v.2.0.3) (McGinnis et al., 2019) to remove doublets; 3) Batch effect removal. We created an integrated data assay of all samples by identifying anchors using FindIntegrationAnchors function; 4) Data normalization was performed using NormalizeData function with scaling factor 10,000 and then log-transformed the data. 5) Detection of 4,000 highly variable genes (HVGs) by FindVariableFeatures function with "vst" method; 6) Scaling of the features by ScaleData function to get a unit variance and zero mean of all samples.

Dimensionality reduction and cell cluster

We perform PCA on the previously determined variable features after scaled, by RunPCA function with top 40 significant PCs that represent a robust compression of the dataset. Next, we applied a graph-based clustering approach to construct a shared nearest neighbor graph for a given dataset by FindNeighbors function between every cell and optimize the modularity function to determine clusters by FindClusters function with resolution equal to 0.6. Finally, we used UMAP to learn the underlying manifold of the data and place similar cells together in low-dimensional space.

Cell type annotation

We executed biologically-related pairwise differential gene expression analysis between duos of clusters by FindAllMarkers function with min.pct equal to 0.2 and logfc.threshold set to 0.25, to identify DEGs and to examine the quantitative changes in the expression levels between the clusters.

Because the marker genes of HSCs and their downstream progeny cells are still uncertain, we additionally collected nine bulk RNA-seq datasets to improve the cell definition. Detailed procedures are as follows: 1) Using GEO2R, a NCBI online tool, to calculate gene expression levels of reference datasets, respectively; 2) Selected top 500 significantly up-regulated genes as biomarkers of each cell type; 3) Performed Hypergeometric distribution test between differentially expressed genes of each dataset and current data, and assigned the cell type based on the significance of P value.

Differential gene expression analysis

We executed biologically-related pairwise differential gene expression analysis between duos of samples to identify DEGs and to examine the quantitative changes in the expression levels between the samples in HSCs. We calculated the DEGs by applying the FindMarkers function (Wilcoxon rank-sum with adjusted P values for multiple testing with the Benjamini-Hochberg correction). We filtered out the obtained DEGs by setting min.pct to 0.2, so that a gene is expressed in at least 20% of the cells in one of the two tested groups. A gene was considered significant with adjusted P < 0.05 and logFC > 0.25.

RNA velocity analysis

By distinguishing between unspliced and spliced mRNAs, RNA velocity, the time derivative of the gene expression state, can be directly estimated. Thus, we utilized velocyto (La Manno et al., 2018) to compute the rate of transcriptional alteration of each cell.

Developmental trajectory inference analysis

The Monocle2 (v.2.14.0) (Qiu et al., 2017; Trapnell et al., 2017) algorithm with the core SRGs was applied to order all cells in pseudo time. By creating an object with parameter "expressionFamily = negbinomial.size", regressed out the batch effect using the "reduceDimension" function with residualModelFormulaStr setting to exclude technology influence and with default reduction_method to achieve dimensionality reduction. Cell differentiation trajectory was determined successfully built based on the above steps.

Next, the BEAM function was used to detect genes that separate cells into the considered cell branches. We used the plot_multiple_branches_heatmap function to separate the branch-related gene set with a q-value less than or equal to $10e^{-4}$ and the num_clusters = 3.

Gene Ontology (GO) terms enrichment analysis

GO enrichment analysis was performed on the given gene set. The enrichGO function of the "clusterProfiler (Wu et al., 2021)" R package was used to do enrichment analysis. Terms with the q value < 0.05 corrected by FDR were considered statistically significant.

Prediction of transcription factor regulons

To predict transcription factor regulons, we utilized SCENIC (v.1.1.3) (Aibar et al., 2017; Van de Sande et al., 2020) with default parameters and used cisTarget database (https://resources.aertslab.org/cistarget/). Generated AUC scores of cells were applied for downstream analysis.

Connectivity Map (CMap) analysis

We used Gene Expression (L1000) in Query (https://clue.io/query, a tool of CMap (Lamb et al., 2006)) with default parameters to compare the similarity between the two SRG signature lists and the expression profile in the reference database.

HSCs signature scoring

We collected the top 250 high expressed genes as HSCs marker genes of each two published papers (Milyavsky et al., 2010; Novershtern et al., 2011), utilizing GSVA(Hanzelmann et al., 2013) to score each HSCs cell, respectively. Meanwhile, we scored each HSCs cell by HSCs marker genes in current data.

FAC analysis

Cells cultured with selected small mollecus were collected at day 6, washed in DPBS, and then incubated with antibody CD34(Biolegend), CD38(BD)at 4℃ for 30min, washed and resupended in DPBS for FAC analysis.

Abbreviations

CB
Cord blood
mPB
mobilized peripheral blood
SRGs
Stemness-related genes
CT-SRGs
SRGs associated with culture time
CS-SRGs
SRGs associated with cell source
scRNA-seq
Single-cell RNA-seq
HSCs
Hematopoietic stem cells
HSPCs
Hematopoietic stem and progenitor cells

CMap
Connectivity Map
FACS
Fluorescence assisted cell sorting
PCA
Principal Component Analysis
UMI
Unique Molecular Identifiers
GO
Gene Ontology

Declarations

Availability of data and materials

The scRNA-seq data generated in this study have been deposited in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb with accession code CNP0000978.

Ethics approval

Permission for this study was obtained from the Bioethics and Biological Safety Review Committee

of BGI-Shenzhen 【BGI-IRB 20155】. We confirm that all experiments were performed in accordance with relevant guidelines and regulations.

Consent for publication

All authors read and approved the final manuscript.

Competing interests

All the authors declare no competing interests.

Author contributions

G.D., S.L., H.S., and C.L. conceived and designed the study. W.Z., Y.L., S.W., and S.L. collected the CB and mPB samples. G.D., W.O., T.L., X.Z. and H.Z. performed the experiments. X.X., J.L. and H.S. conducted the bioinformatics analysis. Y.G. helped project design and discussions. G.D. and X.X.wrote the manuscript. H.S. and C.L. reviewed the manuscript. All authors read and approved the final manuscript.

Author information

Guoyi Dong and Xiaojing Xu have contributed equally to this work. Correspondence to Sixi Liu, Haixi Sun and Chao Liu.

References

1. Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., *et al.* (2017). SCENIC: single-cell regulatory network inference and clustering. Nat Methods *14*, 1083–1086.

2. Boitano, A.E., Wang, J., Romeo, R., Bouchez, L.C., Parker, A.E., Sutton, S.E., Walker, J.R., Flaveny, C.A., Perdew, G.H., Denison, M.S., *et al.* (2010). Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. Science *329*, 1345–1348.

3. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol *36*, 411–420.

4. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., Crooks, G.M., *et al.* (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281–286.

5. Crisan, M., and Dzierzak, E. (2016). The many faces of hematopoietic stem cell heterogeneity. Development (Cambridge, England) *143*, 4571–4581.

6. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

7. Doulatov, S., Notta, F., Eppert, K., Nguyen, L.T., Ohashi, P.S., and Dick, J.E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. Nature immunology *11*, 585–593.

8. Eaves, C.J. (2015). Hematopoietic stem cells: concepts, definitions, and the new reality. Blood *125*, 2605–2613.

9. Eliane, G., Agnes, D., Helene, E., Dominique, T., Gerard, S., Pierre, L., Scott, C., Denis, E., Joanne, K., and Judith, B. (2011). Hematopoietic reconstitution in a patient with Fanconi"s anemia by means of umbilical-cord blood from an HLA-identical sibling (1989). Cellular Therapy and Transplantation *2*.

10. Fares, I., Chagraoui, J., Gareau, Y., Gingras, S., Ruel, R., Mayotte, N., Csaszar, E., Knapp, D.J., Miller, P., Ngom, M., *et al.* (2014). Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. Science *345*, 1509–1512.

11. Giladi, A., Paul, F., Herzog, Y., Lubling, Y., Weiner, A., Yofe, I., Jaitin, D., Cabezas-Wallscheid, N., Dress, R., Ginhoux, F., *et al.* (2018). Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nat Cell Biol *20*, 836–846.

12. Glimm, H., Oh Ih Fau - Eaves, C.J., and Eaves, C.J. (2000). Human hematopoietic stem cells stimulated to proliferate in vitro lose engraftment potential during their S/G(2)/M transit and do not reenter G(0).

13. Griffiths, J.A., Scialdone, A., and Marioni, J.C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. Molecular systems biology *14*, e8046.

14. Gudmundsson, K.O., Nguyen, N., Oakley, K., Han, Y., Gudmundsdottir, B., Liu, P., Tessarollo, L., Jenkins, N.A., Copeland, N.G., and Du, Y. (2020). Prdm16 is a critical regulator of adult long-term hematopoietic stem cell quiescence. Proc Natl Acad Sci U S A *117*, 31945–31953.

15. Guo, B., Huang, X., Cooper, S., and Broxmeyer, H.E. (2017). Glucocorticoid hormone-induced chromatin remodeling enhances human hematopoietic stem cell homing and engraftment. Nat Med *23*, 424–428.

16. Haas, S., Trumpp, A., and Milsom, M.D. (2018). Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. Cell Stem Cell *22*, 627–638.

17. Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7.

18. Huang, X., and Broxmeyer, H.E. (2019). Progress towards improving homing and engraftment of hematopoietic stem cells for clinical transplantation. Curr Opin Hematol *26*, 266–272.

19. Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol *17*, 29.

20. Kannan, K.S., Manoj, K., and Arumugam, S. (2015). Labeling methods for identifying outliers. International Journal of Statistics and Systems *10*, 231–238.

21. Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., *et al.* (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. Nature immunology *19*, 85–97.

22. Kohn, L.A., Hao, Q.L., Sasidharan, R., Parekh, C., Ge, S., Zhu, Y., Mikkola, H.K., and Crooks, G.M. (2012). Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin. Nature immunology *13*, 963–971.

23. Komorowska, K., Doyle, A., Wahlestedt, M., Subramaniam, A., Debnath, S., Chen, J., Soneji, S., Van Handel, B., Mikkola, H.K.A., Miharada, K., *et al.* (2017). Hepatic Leukemia Factor Maintains Quiescence of Hematopoietic Stem Cells and Protects the Stem Cell Pool during Regeneration. Cell Rep *21*, 3514–3523.

24. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lonnerberg, P., Furlan, A., *et al.* (2018). RNA velocity of single cells. Nature *560*, 494–498.

25. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., and Ross, K.N. (2006). The Connectivity Map: using gene-expression signatures to

connect small molecules, genes, and disease. science *313*, 1929–1935.

26. Laurenti, E., Doulatov, S., Zandi, S., Plumb, I., Chen, J., April, C., Fan, J.B., and Dick, J.E. (2013). The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nature immunology *14*, 756–763.

27. Laurenti, E., Frelin, C., Xie, S., Ferrari, R., Dunant, C.F., Zandi, S., Neumann, A., Plumb, I., Doulatov, S., Chen, J., *et al.* (2015). CDK6 levels regulate quiescence exit in human hematopoietic stem cells. Cell Stem Cell *16*, 302–313.

28. Laurenti, E., and Gottgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. Nature *553*, 418–426.

29. Lee, J., Breton, G., Oliveira, T.Y., Zhou, Y.J., Aljoufi, A., Puhr, S., Cameron, M.J., Sekaly, R.P., Nussenzweig, M.C., and Liu, K. (2015). Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. The Journal of experimental medicine *212*, 385–399.

30. Liu, C., Wu, T., Fan, F., Liu, Y., wu, L., Junkin, M., Wang, Z., Yu, Y., Wang, W., Wei, W., *et al.* (2019). A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling.

31. Lu, Y., Hu, M., Zhang, Z., Qi, Y., and Wang, J. (2020). The regulation of hematopoietic stem cell fate in the context of radiation. Radiation Medicine and Protection *1*, 31–34.

32. Maicas, M., Vázquez, I., Alis, R., Marcotegui, N., Urquiza, L., Cortés-Lavaud, X., Cristóbal, I., García-Sánchez, M.A., and Odero, M.D. (2017). The MDS and EVI1 complex locus (MECOM) isoforms regulate their own transcription and have different roles in the transformation of hematopoietic stem and progenitor cells.

33. Martínez-Jaramillo, G., Flores-Guzmán, P., Montesinos, J.J., Quintana, S., Bautista, J., Sánchez-Valle, E., Nambo, M.D.J., and Mayani, H. (2004). In vitro proliferation and expansion of hematopoietic progenitors present in mobilized peripheral blood from normal subjects and cancer patients. Stem cells and development *13*, 382–389.

34. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell systems *8*, 329–337. e324.

35. Milyavsky, M., Gan, O.I., Trottier, M., Komosa, M., Tabach, O., Notta, F., Lechman, E., Hermans, K.G., Eppert, K., Konovalova, Z., *et al.* (2010). A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. Cell Stem Cell *7*, 186–197.

36. Nakamura-Ishizu, A., Ito, K., and Suda, T. (2020). Hematopoietic Stem Cell Metabolism during Development and Aging. Developmental cell *54*, 239–255.

37. Ng, S.Y., Yoshida, T., Zhang, J., and Georgopoulos, K. (2009). Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity *30*, 493–507.

38. Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., and Dick, J.E. (2011). Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science *333*, 218–221.

39. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., *et al.* (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

40. Popescu, D.M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., *et al.* (2019). Decoding human fetal liver haematopoiesis. Nature *574*, 365–371.

41. Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat Methods *14*, 309–315.

42. Ranzoni, A.M., Tangherloni, A., Berest, I., Riva, S.G., Myers, B., Strzelecka, P.M., Xu, J., Panada, E., Mohorianu, I., Zaugg, J.B., *et al.* (2021). Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. Cell Stem Cell *28*, 472–487 e477.

43. Remberger, M., Torlen, J., Ringden, O., Engstrom, M., Watz, E., Uhlin, M., and Mattsson, J. (2015). Effect of Total Nucleated and CD34(+) Cell Dose on Outcome after Allogeneic Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant *21*, 889–893.

44. Rodriguez-Fraticelli, A.E., Weinreb, C., Wang, S.W., Migueles, R.P., Jankovic, M., Usart, M., Klein, A.M., Lowell, S., and Camargo, F.D. (2020). Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nature *583*, 585–589.

45. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902 e1821.

46. Trapnell, C., Cacchiarelli, D., and Qiu, X. (2017). Monocle: Cell counting, differential expression, and trajectory analysis for single-cell RNA-Seq experiments (Bioconductor).

47. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., *et al.* (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc *15*, 2247–2276.

48. van Galen, P., Hovestadt, V., Wadsworth Ii, M.H., Hughes, T.K., Griffin, G.K., Battaglia, S., Verga, J.A., Stephansky, J., Pastika, T.J., Lombardi Story, J., *et al.* (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell *176*, 1265–1281 e1224.

49. Wagner, J.E., Barker, J.N., DeFor, T.E., Baker, K.S., Blazar, B.R., Eide, C., Goldman, A., Kersey, J., Krivit, W., MacMillan, M.L., *et al.* (2002). Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival. Blood *100*, 1611–1618.

50. Wang, J.C., Doedens M Fau - Dick, J.E., and Dick, J.E. Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay.

51. Wang, J.C., Doedens M Fau - Dick, J.E., and Dick, J.E. (1997). Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay.

52. Wang, X., Ge, S., McNamara, G., Hao, Q.-L., Crooks, G.M., and Nolta, J.A. (2003). Albumin-expressing hepatocyte-like cells develop in the livers of immune-deficient mice that received transplants of highly purified human hematopoietic stem cells. Blood, The Journal of the American Society of Hematology *101*, 4201–4208.

53. Wilkinson, A.C., Igarashi, K.J., and Nakauchi, H. (2020). Haematopoietic stem cell self-renewal in vivo and ex vivo. Nat Rev Genet *21*, 541–554.

54. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., *et al.* (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (N Y) *2*, 100141.

55. Xie, S.Z., Garcia-Prat, L., Voisin, V., Ferrari, R., Gan, O.I., Wagenblast, E., Kaufmann, K.B., Zeng, A.G.X., Takayanagi, S.I., Patel, I., *et al.* (2019). Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal. Cell Stem Cell *25*, 639–653 e637.

56. Xie, X., Liu, M., Zhang, Y., Wang, B., Zhu, C., Wang, C., Li, Q., Huo, Y., Guo, J., Xu, C., *et al.* (2021). Single-cell transcriptomic landscape of human blood cells. Natl Sci Rev *8*, nwaa180.

57. Zheng, S., Papalexi, E., Butler, A., Stephenson, W., and Satija, R. (2018). Molecular transitions in early progenitors during human cord blood hematopoiesis. Mol Syst Biol *14*, e8041.

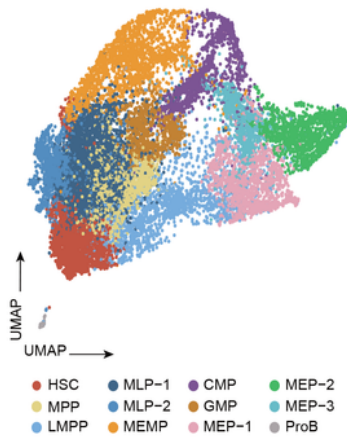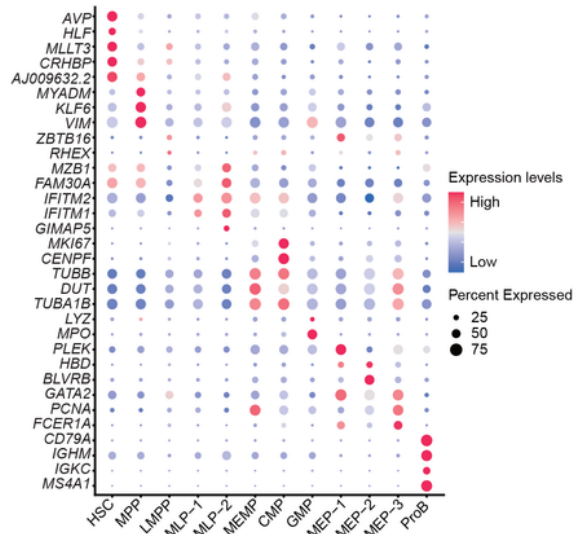# Figures

Figure 1
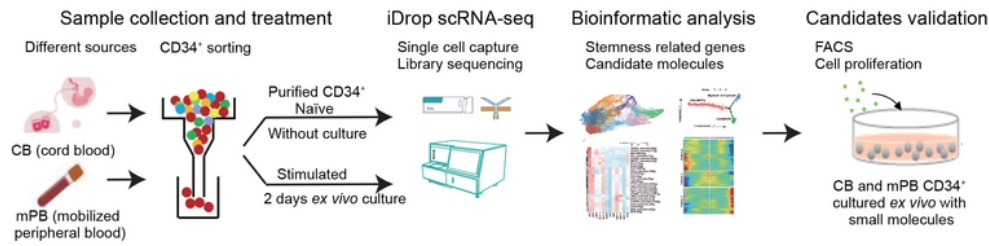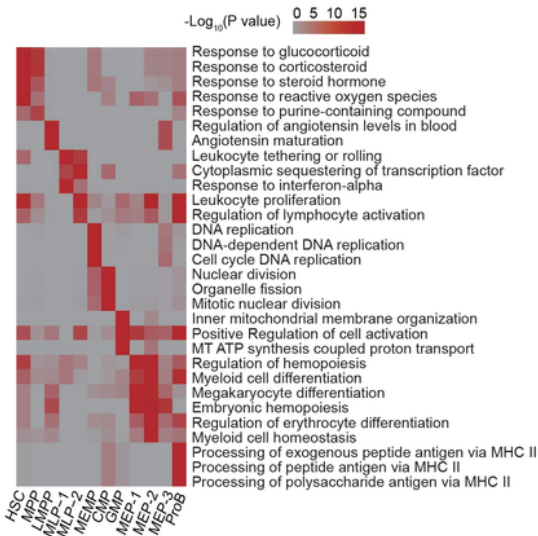
Figure 1

Single-cell transcriptome landscape of naïve and stimulated human CD34⁺ cells from cord blood (CB) and mobilized peripheral blood (mPB).

a, Schematic diagram of the overall experimental design. b, CD34⁺ cells assigned to specific lineages by k-nearest neighbor (KNN) analysis are illustrated in the same UMAP space generated from the data. c, Dot

plot displaying the expression of marker genes among of the identified 12 cell types. The node size positively correlates with the proportion of a given type of cells expressing a given marker gene. The color keys from blue to red indicate the range of relative gene expression in each cell type. d, Enrichment levels of the top enriched GO terms for marker genes of each cell type. The color keys from grey to red indicate the range of -$\log_{10}$-transformed *P* value.

Figure 2
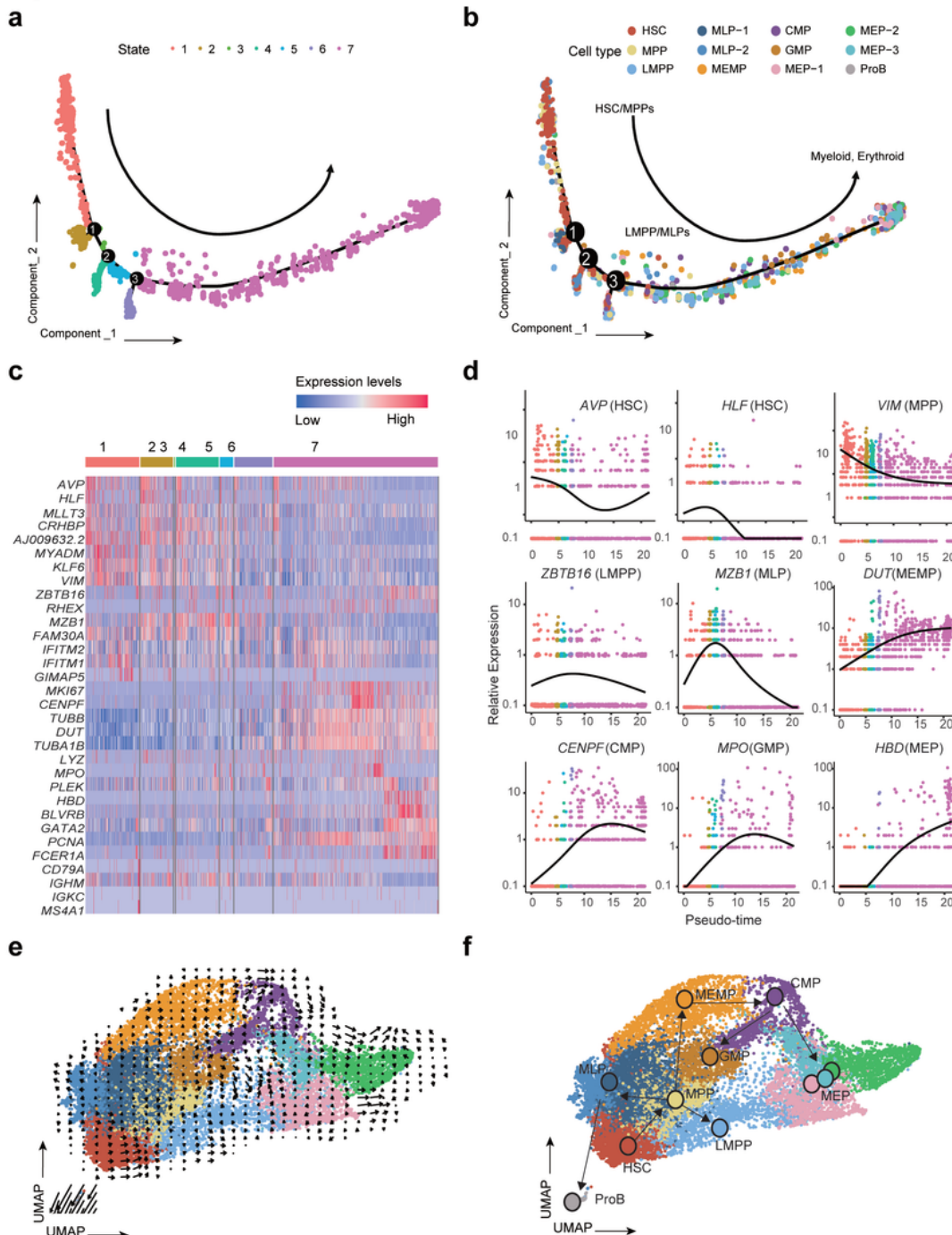
Figure 2

Developmental trajectory analysis of human HSCs.

a, Developmental trajectory showing seven states for the 16,196 CD34[+] cells. b, Cell type assignment along the developmental trajectory. HSCs/MPPs are mainly located in State1, LMPPs and MLPs lie at State2-4, myeloid and erythroid lineage cells gathered in State5-7. c, Heatmap showing the expression of marker genes (shown in Fig. 1c) in 7 developmental States. d, Expression of representative marker genes along developmental pseudo-time. e-f, amalgamated (e) and fitted (f) lineage trees showing the developmental routine from HSCs to multipotent progenitors and downstream myeloid, erythroid and lymphoid lineages. Arrows represent developmental directions, and each circle in (f) represents a cluster.

Figure 3

a

Enriched top GO terms of up-regulated genes in HSCs

-log₁₀(P value)

Response to glucocorticoid
Response to corticosteroid
Response to steroid hormone
Response to reactive oxygen species
Cellular Response to calcium ion
Neg.regulation of multicellular organism growth
Regulation of spindle checkpoint
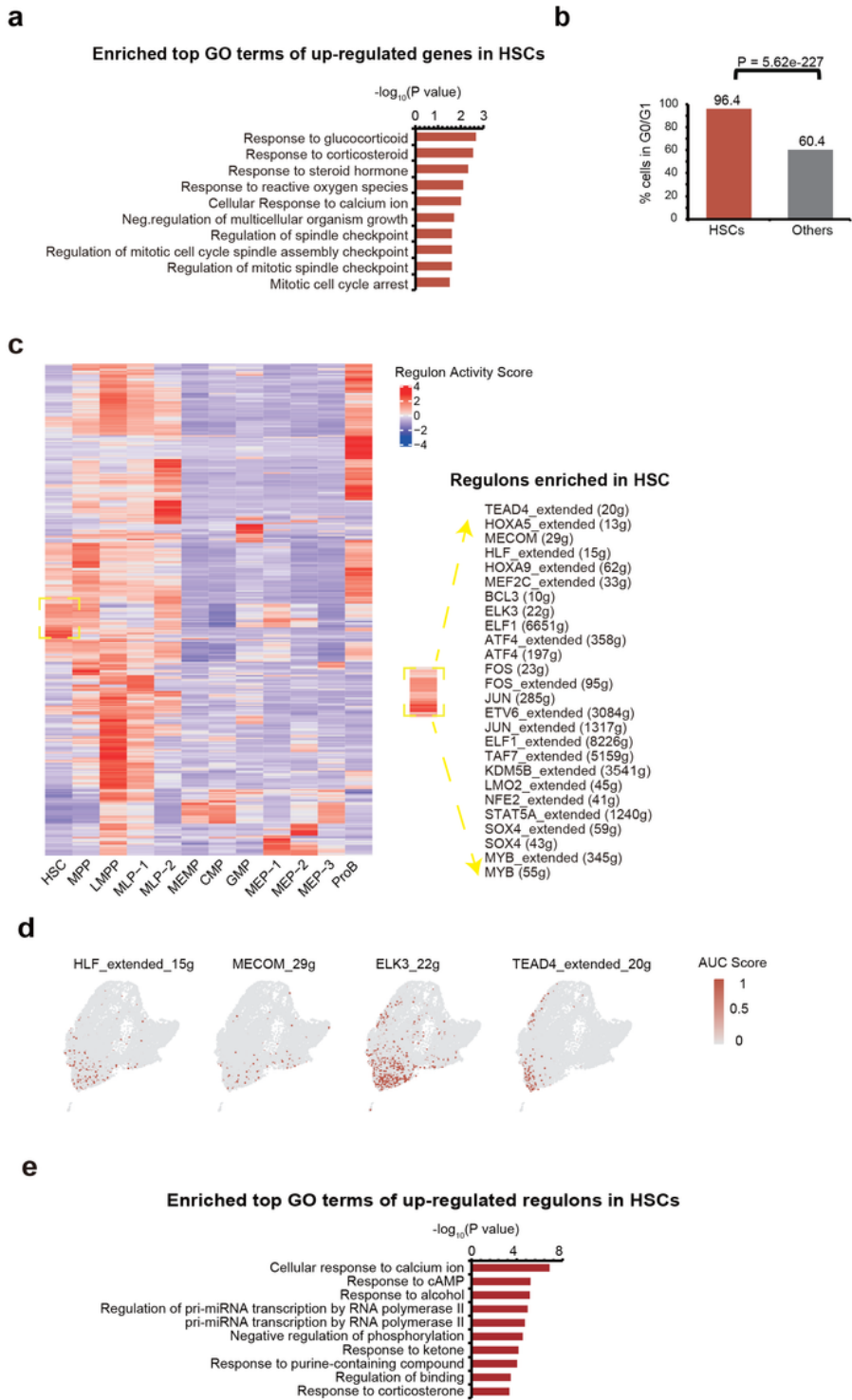Regulation of mitotic cell cycle spindle assembly checkpoint
Regulation of mitotic spindle checkpoint
Mitotic cell cycle arrest

b

P = 5.62e-227

% cells in G0/G1

96.4 HSCs
60.4 Others

c

Regulon Activity Score

Regulons enriched in HSC

TEAD4_extended (20g)
HOXA5_extended (13g)
MECOM (29g)
HLF_extended (15g)
HOXA9_extended (62g)
MEF2C_extended (33g)
BCL3 (10g)
ELK3 (22g)
ELF1 (6651g)
ATF4_extended (358g)
ATF4 (197g)
FOS (23g)
FOS_extended (95g)
JUN (285g)
ETV6_extended (3084g)
JUN_extended (1317g)
ELF1_extended (8226g)
TAF7_extended (5159g)
KDM5B_extended (3541g)
LMO2_extended (45g)
NFE2_extended (41g)
STAT5A_extended (1240g)
SOX4_extended (59g)
SOX4 (43g)
MYB_extended (345g)
MYB (55g)

HSC MPP LMPP MLP-1 MLP-2 MEMP CMP GMP MEP-1 MEP-2 MEP-3 ProB

d

HLF_extended_15g   MECOM_29g   ELK3_22g   TEAD4_extended_20g   AUC Score

e

Enriched top GO terms of up-regulated regulons in HSCs

-log₁₀(P value)

Cellular response to calcium ion
Response to cAMP
Response to alcohol
Regulation of pri-miRNA transcription by RNA polymerase II
pri-miRNA transcription by RNA polymerase II
Negative regulation of phosphorylation
Response to ketone
Response to purine-containing compound
Regulation of binding
Response to corticosterone

Figure 3

Cell cycle and gene regulatory networks of human HSCs.

a, Enriched GO terms of up-regulated genes in HSCs compared with other cell types. b, The percentage of HSCs and other cells in G0/G1 stage. *P* value was calculated by Fisher's Exact Test. c, Heatmap showing up-regulated regulons in each cell type compared to other cell types (left), up-regulated regulons in HSCs

are listed in right. The value for each regulon is the regulon activity score. The color keys from blue to red indicate low to high SCENIC regulons activity levels in each cell type. d, UMAP plot showing the representative regulons of HSCs. Cells are colored by Area Under Curve (AUC) value of each regulon. The color keys from grey to red indicate the range of SCENIC TFs AUC score. e, Enriched GO terms of up-regulated regulons in HSCs.

Figure 4

Changes in the transcriptional profiles after stimulation of HSCs during human hematopoiesis.

a, Proportion of HSCs in naïve and stimulated CD34$^+$ cells from CB and mPB. *P* value was calculated by Fisher's Exact Test. ns means not significant. b, The percentage of HSC cells in G0/G1 stage per sample. *P* value was calculated by Fisher's Exact Test, There was no difference between the two samples. c, Volcano plots showing gene expression changes of HSCs between naïve CB and stimulated CB. d, Volcano plots showing gene expression changes of HSCs between naïve mPB and stimulated mPB. e, Venn diagram showing up-regulated genes in naïve compared with stimulated samples of HSCs between CB and mPB. f, Volcano plots showing gene expression changes of HSCs between naïve CB and naïve mPB. g-h, Violin plots showing the scores of HSC marker genes (HSC-LDGs), CT-SRGs and CS-SRGs of HSCs identified in Novershtern *et al*. (g) and Milyavsky *et al.* (h). The horizontal axis represents different gene sets, and the vertical axis represents GSVA enrichment scores of signature genes of HSCs in the paper. *P* value was calculated by two-sided Wilcoxon rank-sum test. i, Common enriched GO terms of CT-SRGs and CR-SRGs. GO terms enriched in CT-SRGs are shown in red, whereas GO terms enriched in CS-SRGs are shown in blue, the colored bars indicate the range of -log$_{10}$-transformed *P* value. j, Specifically enriched GO terms of CT-SRGs compared to CS-SRGs. k, Specifically enriched GO terms of CS-SRGs compared to CT-SRGs.
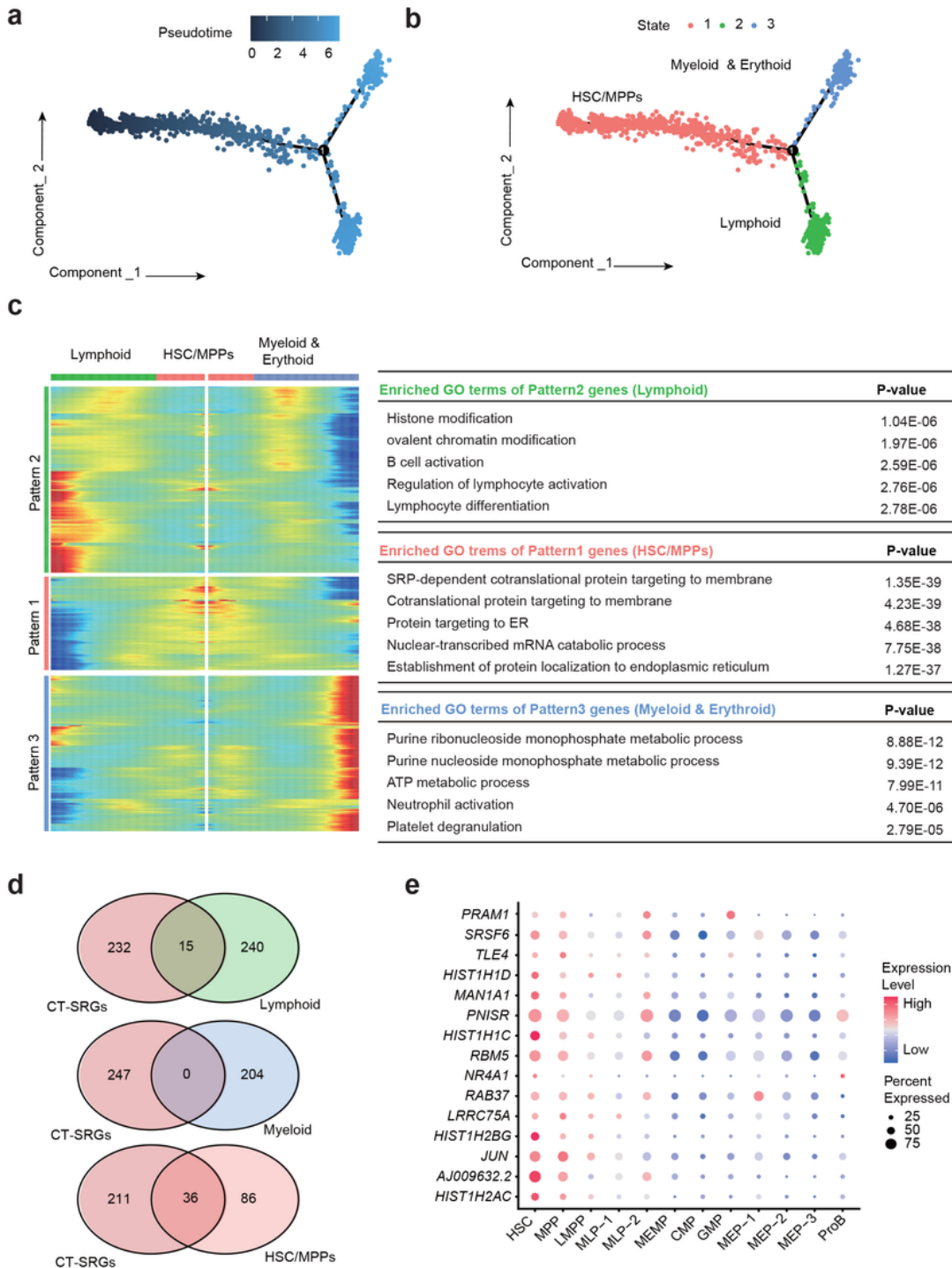
Figure 5

Figure 5

Differentiation and development trajectory based on CT-SRGs.

a, The trajectory tree based on pseudo time staining. b, The State branch recognized by Monocle2. c, Heatmap of the key genes involved in branch determination and their functions. Heatmap showing the three dynamic gene expression patterns of Lymphoid, HSCs/MPPs and Myeloid & Erythoid (left). The

specifically enriched GO terms are shown on the right. d, Venn diagram showing shared genes between CT-SRGs and three clusters (shown in c). e, Dot plot displaying the expression of 15 shared genes in CT-SRGs and cluster1 (Lymphoid). The node size represents the cell proportion that expresses the given gene. The color keys from blue to red indicate the range of relative gene expression.

Figure 6

Cucurbitacin I could enhance HSCs proliferation and stemness.

a, The small molecules screening pipeline. b, Candidate molecules affect CT-SRGs predicted by CMap. c, Proliferation kinetics of CD34$^+$ cells from CB and mPB in candidate small molecules culture. d, Fold-increase in CD34$^+$ cell number on 6 days culture ex vivo as compared to input numbers. e, Flow cytometry plots show CD34$^+$CD38$^-$ populations on 6 days culture ex vivo from CB and mPB. f, Proportion of CD34$^+$CD38$^-$ cells after different small molecule treatments detected by Flow Cytometry. P value was calculated by *t*-test, *p < 0.05; **p < 0.01; ***p < 0.001; ns, not significant. Error bars indicate standard deviations of triplicate cultures. For fenretinide and cucurbitacin I treatment, 100 nM dose was used.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tables.xlsx
- Fig.S1.pdf
- Fig.S2.pdf
- Fig.S3.pdf
- Fig.S4.pdf
- Fig.S5.pdf
- Fig.S6.pdf