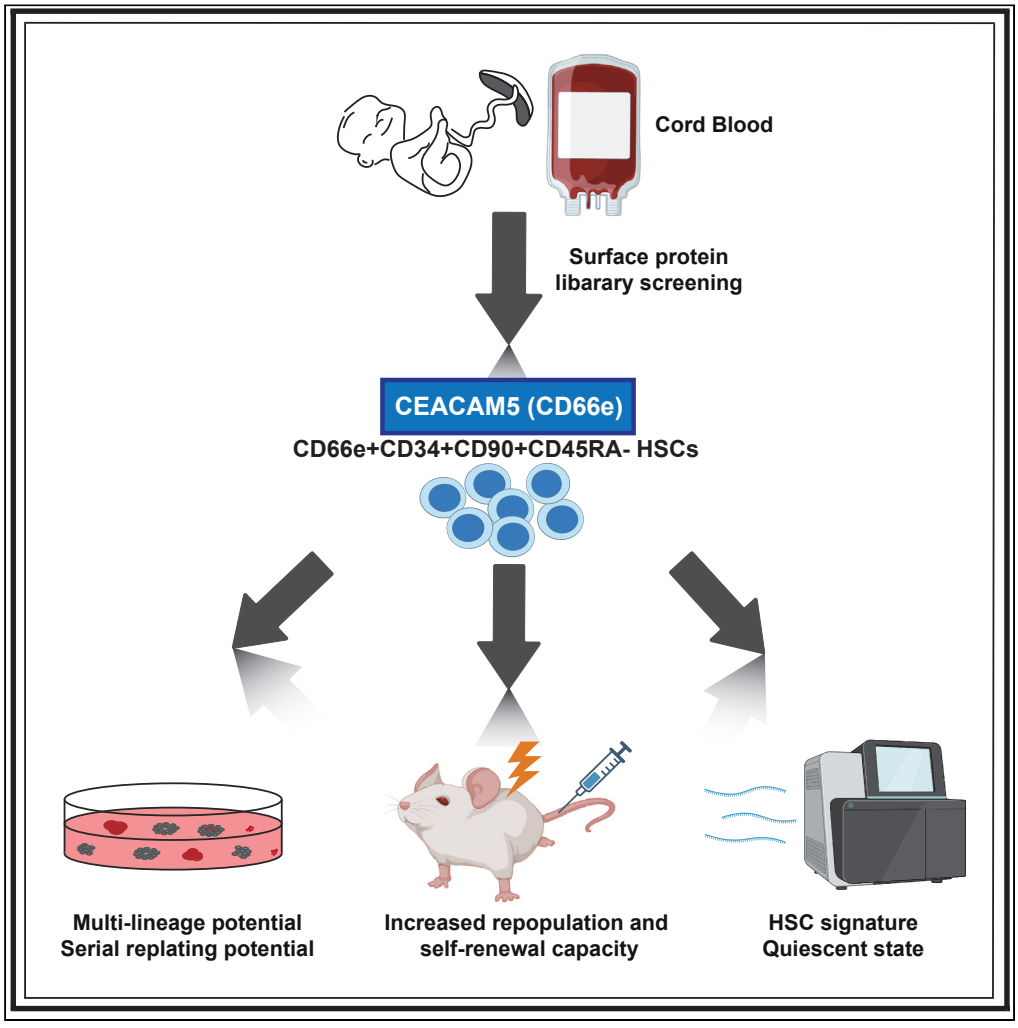
iScience

CellPress
OPEN ACCESS

Article

CEA cell adhesion molecule 5 enriches functional human hematopoietic stem cells capable of long-term multi-lineage engraftment



Kuiying Ma, Xuan Wang, Linjie Wu, ..., Wensheng Wei, Riguo Fang, Tao Cheng

hulinping@ihcams.ac.cn (L.H.)
wswei@pku.edu.cn (W.W.)
rgfang@edigene.com (R.F.)
chengtao@ihcams.ac.cn (T.C.)

Highlights

CEACAM5 is a cell surface marker that enriched functional human LT-HSCs

CEACAM5 expression is required for the repopulating activity of human HSCs

Ma et al., iScience *26*, 108561
December 15, 2023 © 2023
https://doi.org/10.1016/j.isci.2023.108561

iScience

CellPress
OPEN ACCESS

Article

CEA cell adhesion molecule 5 enriches functional human hematopoietic stem cells capable of long-term multi-lineage engraftment

Kuiying Ma,[1,7] Xuan Wang,[2,6,7] Linjie Wu,[2,6,7] Lingling Yu,[1] Jinhui Ye,[2,6] Xueling Li,[2,6] Lili Geng,[1] Zhongyu Shi,[1] Huihui Yang,[1] Xijuan Zhang,[1] Yongjian Zhang,[1] Shuchang Wu,[1] Pengfei Yuan,[1] Yingchi Zhang,[2,6] Fang Dong,[2,6] Sha Hao,[2,6] Linping Hu,[2,6,8,*] Wensheng Wei,[3,*] Riguo Fang,[1,*] and Tao Cheng[2,4,5,6,*]

SUMMARY

Hematopoietic stem cell (HSC) surface markers improve the understanding of cell identity and function. Here, we report that human HSCs can be distinguished by their expression of the CEA Cell Adhesion Molecule 5 (CEACAM5, CD66e), which serves as a marker and a regulator of HSC function. $CD66e^+$ cells exhibited a 5.5-fold enrichment for functional long term HSCs compared to $CD66e^-$ cells. $CD66e^+CD34^+CD90^+CD45RA^-$ cells displayed robust multi-lineage repopulation and serial reconstitution ability in immunodeficient mice compared to $CD66e^-CD34^+CD90^+CD45RA^-$cells. CD66e expression also identified almost all repopulating HSCs within the $CD34^+CD90^+CD45RA^-$ population. Together, these results indicated that CEACAM5 is a marker that enriches functional human hematopoietic stem cells capable of long-term multi-lineage engraftment.

INTRODUCTION

Gene-modified hematopoietic stem cell (HSC) therapy has demonstrated remarkable success in the treatment of inherited blood disorders. As the apex of the hematopoietic hierarchy, HSCs play an essential role in sustaining life-long hematopoiesis. Moreover, HSCs were identified via reliable and robust biomarkers that facilitate the development of HSC gene therapy. Several surface markers have been used to enrich long-term HSCs (LT-HSCs).[1–4]

Previous studies have shown that LT-HSCs, enriched in the $Lin^-CD34^+CD38^-CD45RA^-CD90^+CD49f^+$ population, support long-term hematopoietic reconstitution.[1] However, such as CD38, the current surface markers are unreliable during *ex vivo* culturing.[2,5] Thus, HSC characterization has been hindered by lacking bonafide biomarkers. The identification of novel HSC markers facilitated further refinement of the immunophenotype and improved the purification of HSCs, which is instrumental in gaining new insights into human HSC biology and improving both quantification and graft engineering in the clinical setting. Therefore, we performed *in vitro* screening and comprehensive functional evaluation to identify an undiscovered surface marker of human HSCs.

RESULTS

CEA cell adhesion molecule 5 expression identifies a subpopulation of human hematopoietic stem cells

We used a cell surface antigen screen panel (including 242 human cell surface markers) for initial screening and combined it with CD34 and CD90 antibodies for flow cytometry analysis to enrich human cord blood (CB) $CD34^+$ HSPCs (Figure 1A). Compared to $CD34^+$ cell population, CD66- (a, c, d, and e), CD200-, and CD48-positive cells were more enriched in $CD34^+CD90^+$ subsets (Figure 1B). Previous studies indicated that HSCs could not be maintained during *in vitro* culture without supplementation with small molecules, such as SR1 and UM171.[3,6,7] Based on this principle, we tracked these candidate surface markers. The expression of CD48, CD66c, CD66d, and CD200 was maintained or increased after 3 days of *in vitro* culture. In contrast, CD66a and CD66e expression was decreased (Figure S1A). CD66a (CEACAM1) has been identified as a human culture-compatible surface marker of expanded long-term reconstituting

[1]EdiGene Inc., Life Science Park, Changping District, Beijing 102206, China
[2]State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China
[3]Biomedical Pioneering Innovation Center, Beijing Advanced Innovation Center for Genomics, Peking-Tsinghua Center for Life Sciences, Peking University Genome Editing Research Center, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China
[4]Center for Stem Cell Medicine, Chinese Academy of Medical Sciences, Tianjin 300000, China
[5]Department of Stem Cell & Regenerative Medicine, Peking Union Medical College, Tianjin 300000, China
[6]Tianjin Institutes of Health Science, Tianjin 301600, China
[7]These authors contributed equally
[8]Lead contact
*Correspondence: hulinping@ihcams.ac.cn (L.H.), wswei@pku.edu.cn (W.W.), rgfang@edigene.com (R.F.), chengtao@ihcams.ac.cn (T.C.)
https://doi.org/10.1016/j.isci.2023.108561

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

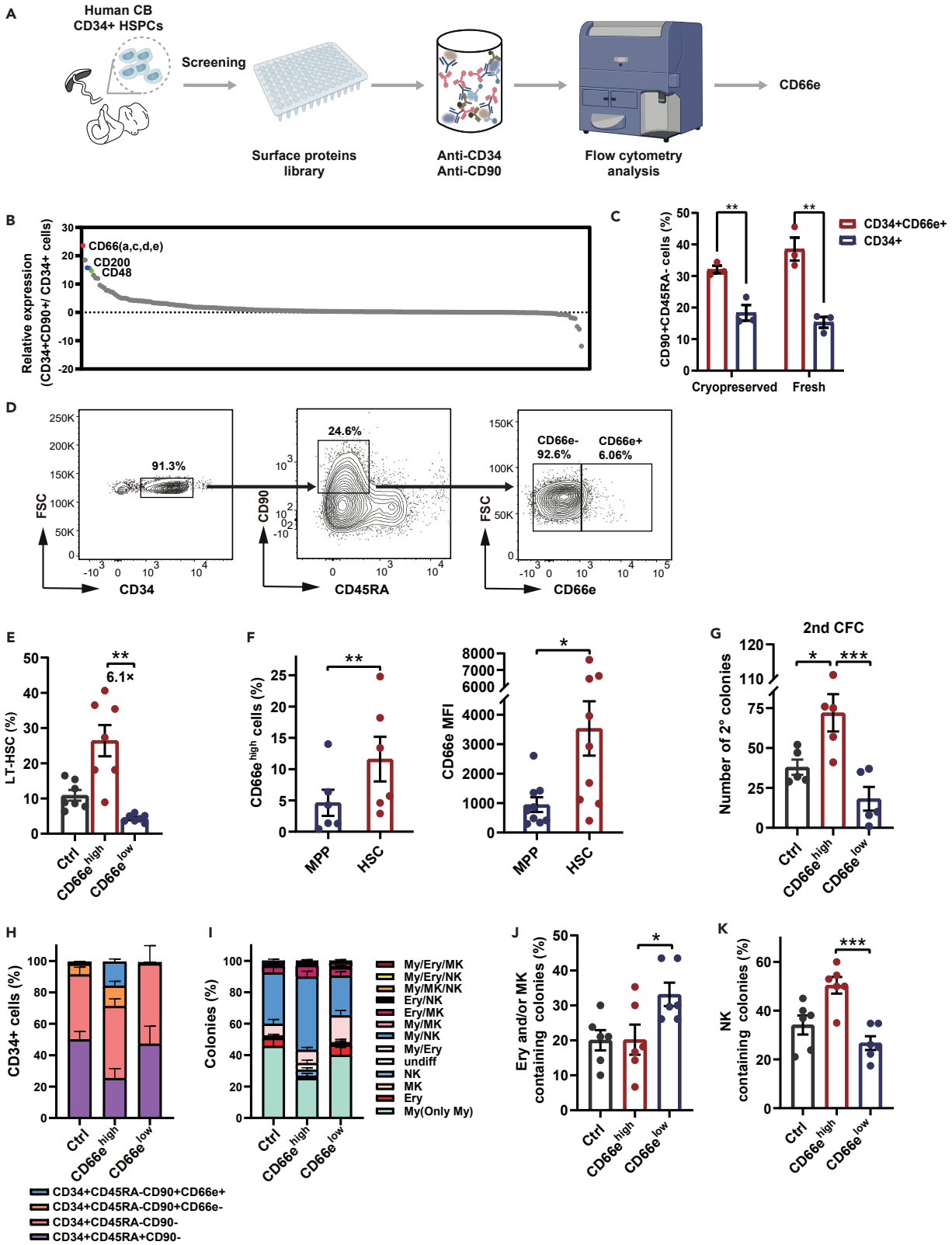

A
Human CB
CD34+ HSPCs
Screening
Surface proteins
library
Anti-CD34
Anti-CD90
Flow cytometry
analysis
CD66e
B
Relative expression
(CD34+CD90+/ CD34+ cells)
CD66(a,c,d,e)
CD200
CD48
C
CD90+CD45RA- cells (%)
CD34+CD66e+
CD34+
Cryopreserved
Fresh
D
91.3%
24.6%
CD66e-
92.6%
CD66e+
6.06%
FSC
CD34
CD90
CD45RA
FSC
CD66e
E
LT-HSC (%)
6.1×
Ctrl
CD66e high
CD66e low
F
CD66e high cells (%)
CD66e MFI
MPP
HSC
G
2nd CFC
Number of 2° colonies
H
CD34+ cells (%)
CD34+CD45RA-CD90+CD66e+
CD34+CD45RA-CD90+CD66e-
CD34+CD45RA-CD90-
CD34+CD45RA+CD90-
I
Colonies (%)
My/Ery/MK
My/Ery/NK
My/MK/NK
Ery/NK
Ery/MK
My/MK
My/NK
My/Ery
undiff
NK
MK
Ery
My(Only My)
J
Ery and/or MK
containing colonies (%)
K
NK
containing colonies (%)

Figure 1. CD66e$^+$ cells are enriched for human cord blood HSCs

(A) Schematic representation of screening strategy: a total of 242 human surface markers were co-stained with CD34 and CD90 on human CB HSPCs. Primary hints were identified by flow cytometry.

(B) Relative expression of each protein in CD34$^+$CD90$^+$ cells compared to CD34$^+$ cells from primary screening.

(C) Frequency of CD90$^+$CD45RA$^-$ cells among CD34$^+$ or CD34$^+$CD66e$^+$ cells from cryopreserved or fresh CB HSPCs (n = 3). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(D) Representative FACS plots of CD66e expression on cord blood HSCs. CB CD34$^+$ cells were stained with monoclonal antibodies against CD34, CD45RA, CD90, CD66e antigens. The representative plot was gated on living cells.

(E) Quantification of the LT-HSCs percentage of CD66e^{high} (Lin$^-$CD34$^+$CD38$^-$CD66e^{high} cells), CD66e^{low} (Lin$^-$CD34$^+$CD38$^-$CD66e^{low} cells) or control cells from CB (Lin$^-$CD34$^+$CD38$^-$ cells) (n = 7; LT-HSCs: Lin$^-$CD34$^+$CD38$^-$CD90$^+$CD45RA$^-$ cells). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(F) Frequency of CD66e^{high} cells (left) among MPP (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^-$ cells) or HSC (Lin$^-$CD34$^+$CD38$^-$CD90$^+$CD45RA$^-$ cells) (n = 6) in CB samples, and mean fluorescence index (MFI) of CD66e (right) on MPP or HSC (n = 9, CB samples). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(G) Colony counts for CD66e^{high} (Lin$^-$CD34$^+$CD66e^{high} cells), CD66e^{low} (Lin$^-$CD34$^+$CD66e^{low} cells) or control cells (Lin$^-$CD34$^+$ cells) from secondary CFC assay of CB samples (n = 5). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(H) The frequency of denoted output subpopulations after culturing CB CD66e^{low} (CD34$^+$CD90$^+$CD45RA$^-$CD66e^{low} cells, n = 6), CD66e^{high} (CD34$^+$CD90$^+$CD45RA$^-$CD66e^{high} cells, n = 4), or control cells (CD34$^+$CD90$^+$CD45RA$^-$ cells, n = 6) for 10 days. Data are represented as mean ± SEM.

(I) Proportion of different cellular outputs obtained from CB CD66e^{high} (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$CD66e^{high} cells), CD66e^{low} (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$CD66e^{low} cells) or control cells (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$ cells). Ery colonies were identified as CD45$^-$CD235a$^+$ ≥30 cells, Mk colonies as CD41$^+$ ≥10 cells, My colonies as [(CD45$^+$CD14$^+$) and (CD45$^+$CD15$^+$)] ≥30 cells, NK colonies as CD45$^+$CD14$^-$CD15$^-$CD56$^+$ ≥30 cells. Colonies were identified as undifferentiated (undiff) if CD45$^+$CD14$^-$CD15$^-$CD56$^-$ ≥30 cells, but the cells could not be identified as positive for any lineage (Ery, Mk, My, NK) using the criteria above. Data are represented as mean ± SEM.

(J) Proportion of Ery containing colonies obtained from CD66e^{high}/CD66e^{low}/control cells in the panel I (n = 6). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(K) Proportion of NK containing colonies obtained from CD66e^{high}/CD66e^{low}/control cells in the panel I (n = 6). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

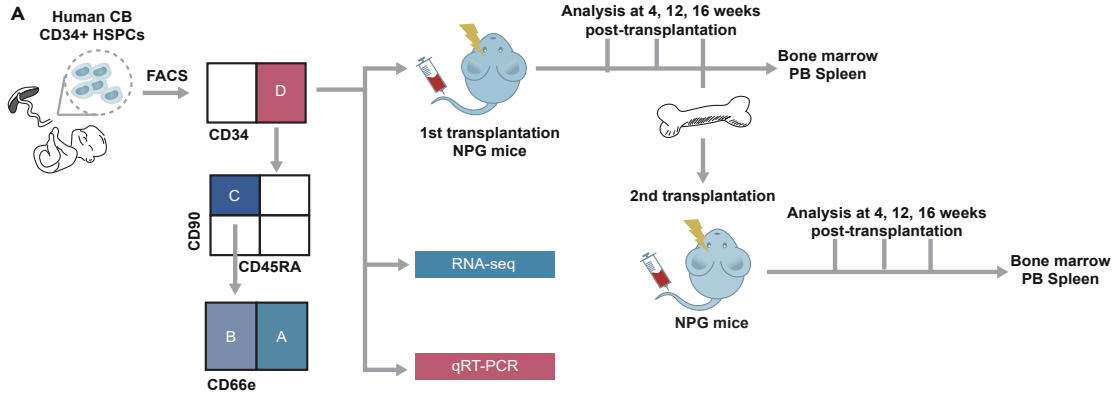
HSCs.[4] We then validated if CD66e is a marker associated with human HSCs. As expected, CD34$^+$CD66e$^+$ cells enriched more CD90$^+$CD45RA$^-$ cells compared to CD34$^+$ cells in both cryopreserved and fresh HSPCs (Figure 1C). The expression of CD66e on HSCs (CD34$^+$CD45RA$^-$CD90$^+$) is shown (Figure 1D), and we defined CD66e$^+$ and CD66e$^-$ population according to FMO staining sample (Figure S2A). To define the correlation between CD66e^{high} cells and HSCs precisely, we performed a detailed analysis by flow cytometry. CD66e^{high} cells exhibited a 6.1-fold enrichment for LT-HSCs (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$ cells) compared to CD66e^{low} cells (Figures 1E and S1B). Consistent with this finding, the percentage of LT-HSCs expressing CD66e was increased compared to multipotent progenitors (MPPs, Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^-$ cells), and the mean fluorescence intensity (MFI) of CD66e was increased on LT-HSCs compared to MPPs (Figure 1F).

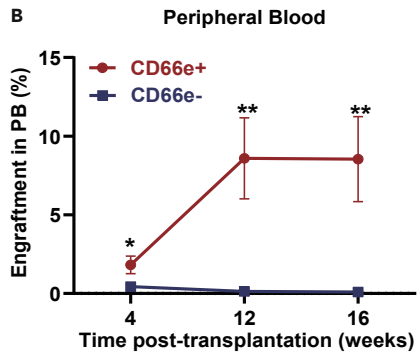
We then measured the clonogenic potential of CD66e^{high} and CD66e^{low} HSPCs in methylcellulose colony-forming assays. CD66e^{high} and CD66e^{low} HSPCs had clonogenic potential similar to controls (Ctrl), and CD66e^{low} HSPCs had a high output of erythroid (CFU-E) colonies (Figure S2B). However, CD66e^{high} HSPCs gained secondary serial replating potential, indicating increased self-renewal ability (Figure 1G). To further characterize CD66e^{high} cells functionally, the CB CD34$^+$ cells were fractionated based on CD90 and CD66e expression, and the progeny of the sorted cells were evaluated after 10 days in culture (Figure S2C). Interestingly, only CD90$^+$CD66e^{high} cells were able to reestablish the initial heterogeneity of the population in culture, giving rise to all four CD90/CD66e subsets, suggesting that the CD90$^+$CD66e^{high} population is at the apex of these subpopulations (Figure 1H).

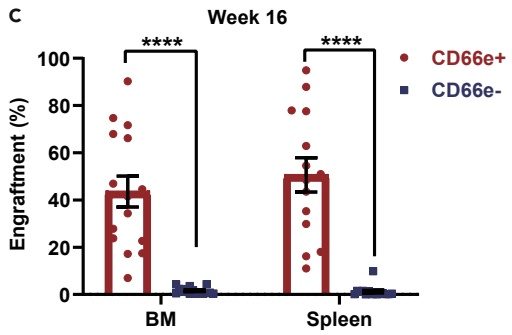
To evaluate the multi-lineage potential of CD66e^{high} and CD66e^{low} HSCs, we sorted single CD66e^{high} HSC (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$CD66e^{high}) or CD66e^{low} HSC (Lin$^-$CD34$^+$CD38$^-$CD45RA$^-$CD90$^+$CD66e^{low}), and cultured it in medium promoting simultaneous differentiation into My, Ery, Meg, and NK lineages.[8] These cells gave rise to a range of uni- or oligo-lineage colonies (Figures 1I and S2D). Notably, CD66e^{low} HSCs differentiated toward the Ery lineage, whereas CD66e^{high} HSCs had a high proportion of lymphoid (NK) lineage output (Figures 1J and 1K). This phenomenon was consistent with the principal component analysis result of bulk RNA-seq data. The lower PC1 and higher PC2 values of CD66e$^-$ HSCs indicated a higher metabolism and erythroid skewing than CD66e$^+$ HSCs (Figure S4A).

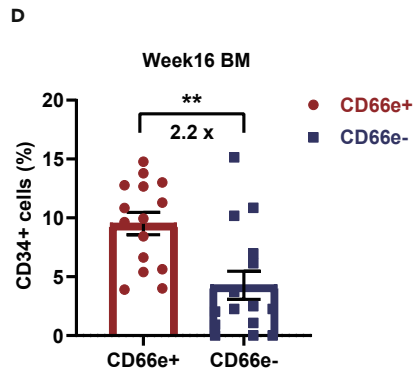
CEA cell adhesion molecule 5 is enriched in hematopoietic stem cell with increased repopulation and self-renewal capacity

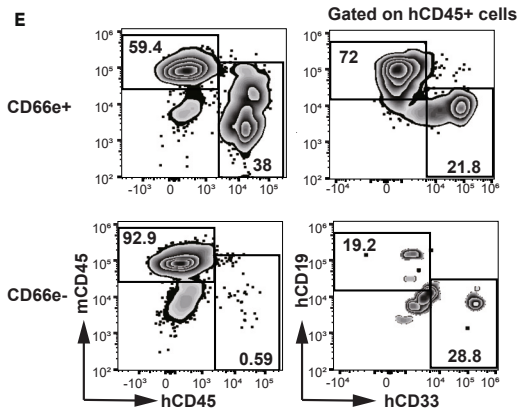
To determine if CEACAM5 expression is associated with increased long-term repopulating capacity *in vivo*, we sorted CD34$^+$CD90$^+$CD45RA$^-$CD66e$^{+/-}$ population from CB HSPCs and conducted a LDA on irradiated immunodeficiency NPG mice (Figures 2A and S3A–S3C). The repopulation kinetics of human CB HSCs were measured at different time points after transplantation. We observed that the CD66e$^+$ group (transplanted with CD34$^+$CD90$^+$CD45RA$^-$CD66e$^+$ cells) exhibited increased reconstitution over time before 12 weeks post-transplantation, whereas the CD66e$^-$ group (transplanted with CD34$^+$CD90$^+$CD45RA$^-$CD66e$^-$ cells) was only capable of reconstitution at 4 weeks post-transplantation (Figure 2B). Sixteen weeks after transplantation, the CD66e$^+$ group exhibited significantly higher reconstitution in PB, BM, and spleen in contrast to the CD66e$^-$ group (Figures 2B and 2C). Importantly, we found that the enrichment of CD34$^+$ HSPCs among the BM of the primary recipient of the CD66e$^+$ group is 2.2-fold higher than the CD66e$^-$ group (Figure 2D),

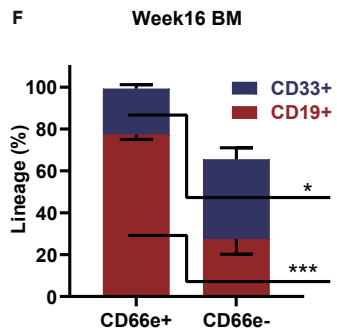

A
Human CB
CD34+ HSPCs
FACS
CD34
D
C
CD90
CD45RA
B
A
CD66e
1st transplantation
NPG mice
Analysis at 4, 12, 16 weeks
post-transplantation
Bone marrow
PB Spleen
2nd transplantation
NPG mice
Analysis at 4, 12, 16 weeks
post-transplantation
Bone marrow
PB Spleen
RNA-seq
qRT-PCR

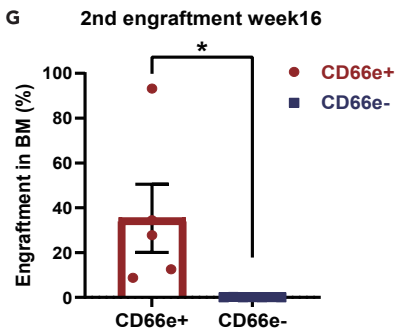

B
Peripheral Blood
CD66e+
CD66e-
Engraftment in PB (%)
Time post-transplantation (weeks)


C
Week 16
CD66e+
CD66e-
Engraftment (%)
BM
Spleen


D
Week16 BM
2.2 x
CD66e+
CD66e-
CD34+ cells (%)


E
Gated on hCD45+ cells
CD66e+
CD66e-
59.4
38
72
21.8
92.9
0.59
19.2
28.8
mCD45
hCD45
hCD19
hCD33


F
Week16 BM
CD33+
CD19+
Lineage (%)


G
2nd engraftment week16
CD66e+
CD66e-
Engraftment in BM (%)

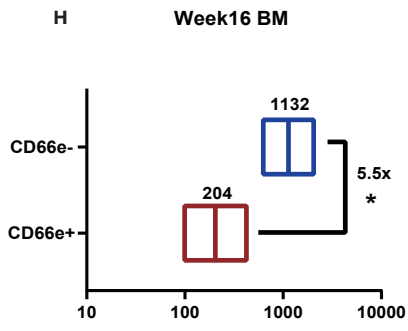

H
Week16 BM
1132
204
5.5x
CD66e-
CD66e+

Figure 2. CD66e expression enriches for HSCs with increased repopulating and self-renewal capacity

(A) Schematic representation of functional evaluations: HSPCs sorted from human CB were used in this study. A: $CD34^+CD90^+CD45RA^-CD66e^+$ cells, B: $CD34^+CD90^+CD45RA^-CD66e^-$ cells, C: $CD34^+CD90^+CD45RA^-$ cells, and D: $CD34^+$ cells. For *in vivo* assay, A and B populations were injected into irradiated NPG mice with 3000/1000/300 cells per mouse. BM cells from the primary mice were transplanted into the second recipients (n = 8). Gene expression profiles were obtained by RNA-seq and qRT-PCR using sorted A, B, C, and D populations (n = 3).

(B) PB engraftment in NPG mice during 4–16 weeks post-transplantation. Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(C) BM and spleen engraftment in NPG mice 16 weeks post-transplantation. Each dot represents one mouse. Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(D) Frequency of $CD34^+$ cells among BM cells in NPG mice 16 weeks post-transplantation. Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(E) FACS profiles of BM cells 16 weeks post-engraftment.

(F) Myeloid ($CD33^+$) and lymphoid ($CD19^+$) lineage potential of $CD66e^+$ or $CD66e^-$ cells in BM after 16 weeks post-transplantation. Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(G) Four mice with highest engraftment rate of $CD66e^+$ 3000 cells group and $CD66e^-$ 3000 cells group were selected for the secondary transplantation. Each recipient was engrafted with 40% BM cells from the primary recipients. BM engraftment in NPG mice 16 weeks after secondary transplantation. Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(H) Estimated frequency of LT-HSCs within $CD66e^+$ and $CD66e^-$ population sorted from $CD34^+CD90^+CD45RA^-$ cells. Estimated frequencies are presented as 1/number of sorted cells and 95% confidence intervals (boxes). ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

suggesting that $CD66e^+$ group contained more long-term repopulating cells than the $CD66e^-$ group. The lineage analysis indicated that $CD66e^+$ group presented a robust multi-lineage reconstitution, while the $CD66e^-$ group displayed an impaired multi-lineage differentiation pattern (Figures 2E and 2F). Moreover, $CD66e^+$ cells displayed significantly higher lymphoid potential compared with $CD66e^-$ cells, consistent with the single cell colony-forming assays and RNA-seq analysis (Figures 2F, 1K, and S4A).

Secondary transplantations remain the most commonly used strategy to assess the self-renewal ability of HSCs. Thus, we performed parallel secondary transplantations from both $CD66e^+$ and $CD66e^-$ cell primary recipients (Figure S3B). Noticeably, human cells from primary $CD66e^+$ recipients generated a significantly higher of the mean engraftment levels in the BM compared with those from the $CD66e^-$ group (Figure 2G). These results suggested that $CD66e^+$ HSCs exhibited increased self-renewal capacity and, as a result, gave rise to superior engraftment during secondary transplantation.

We next performed LDA to measure the frequency of HSCs capable of initiating engraftment in immunodeficient mice (Figure S3B). One in 1132 cells in the $CD66e^-$ group clonally initiated long term hematopoiesis in NPG mice, whereas 1 in 204 cells did so in the $CD66e^+$ group (Figures 2H and S3C). These findings revealed that the majority of cells with long-term reconstitution capacity among $CD34^+CD90^+CD45RA^-$cells were CD66e-positive.

CEA cell adhesion molecule 5^+ cells associated with hematopoietic stem cell signature

To determine the transcriptional identity of $CD66e^+$ cells, we performed mRNA sequencing experiments using $CD66e^+$ and $CD66e^-$ populations sorted from $CD34^+CD90^+CD45RA^-$ CB cells (Figure 2A). The $CD66e^+$ population showed genetic features with hematopoietic cells at the top hematopoietic hierarchy (Figure 3A).[9] The gene signature of $CD34^+CD90^+CD45RA^-CD66e^+$ cells was toward HSC compared to hematopoietic progenitors, such as multi-lymphoid progenitor (MLP), megakaryocyte-erythroid progenitor (MEP), and granulocyte-macrophage progenitor (GMP) (Figure 3B). qRT-PCR verified that CD66e expression was correlated with HSC-related genes *AVP* and *HES1* (Figure 3C).[10–12] Furthermore, $CD34^+CD90^+CD45RA^-CD66e^+$ cells have a higher expression of genes encoding HSC markers than other counterparts, such as *CD90*, *CD143*, *CD318*, *CD133*, *ITGA3*, *EPCR*, and *CEACAM1* (Figures 3D and 3E).[1–4,13–15] Furthermore, a high score of LT-HSC signature composed by published LT-HSC marker genes was determined on $CD66e^+$ subpopulation (Figure 3F).[16,17] These results suggested that the gene expression profile of $CD66e^+$ cells is associated with the HSC signature.

We also observed that the $CD66e^+$ subpopulation had a lower S score and G2M score compared to other counterparts, indicating that $CD66e^+$ HSCs were maintained in a quiescent state (Figure 3G). Cell cycle analyses showed that $CD66e^{high}$ subset ($CD34^+CD66e^{high}$) contained a significantly higher percentage of G0 phase cells than the $CD66e^{low}$ subset ($CD34^+CD66e^{low}$) (Figures S4B and S4C). Consistent with this finding, cell-cycle inhibitor *CDKN1A* was upregulated in $CD66e^+$ HSCs, which manipulated HSCs exhaustion. While *CDKN2C* was downregulated in $CD66e^+$ HSCs, which had the strongest inhibitory effect on HSC self-renewal (Figure S5A).[18–22] These results indicated that CD66e expression is associated with HSC quiescence. Previous studies had confirmed that HSCs rely on anaerobic glycolysis for energy production and shift towards mitochondrial oxidative phosphorylation upon differentiation.[23–25] Here we found that oxidative phosphorylation were significantly downregulated in the $CD66e^+$ subpopulation, which represented a feature of HSCs (Figure S5B). Moreover, CEACAM5 is a highly glycosylated protein, functionally described as an adhesion molecule.[26] GSEA and GO analysis revealed a significant enrichment of transcripts associated with cell-cell adhesion, cytokine-cytokine receptor interaction, and positive regulation of hemopoiesis in $CD66e^+$ HSCs, highlighting the potential implication of CD66e in the crosstalk with the bone marrow niche (Figures S6A–S6D).

In summary, we have identified CD66e as an HSC marker, and its expression enriches for human LT-HSCs. $CD66e^+$ cells within the $CD34^+CD90^+CD45RA^-$ population display durable multipotentiality and extensive self-renewal activity after transplantation into primary

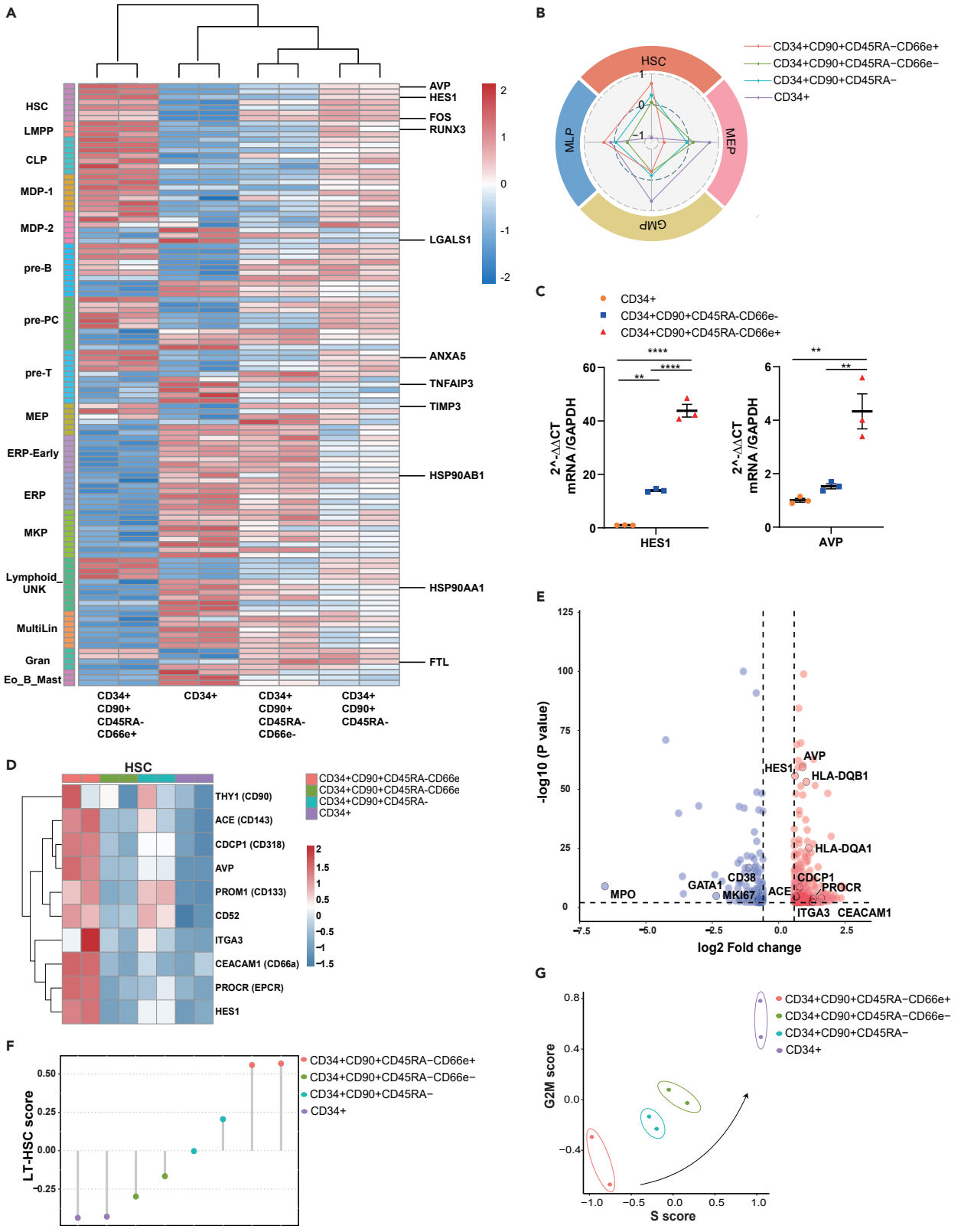

A
HSC
LMPP
CLP
MDP-1
MDP-2
pre-B
pre-PC
pre-T
MEP
ERP-Early
ERP
MKP
Lymphoid_UNK
MultiLin
Gran
Eo_B_Mast
AVP
HES1
FOS
RUNX3
LGALS1
ANXA5
TNFAIP3
TIMP3
HSP90AB1
HSP90AA1
FTL
CD34+ CD90+ CD45RA- CD66e+
CD34+
CD34+ CD90+ CD45RA- CD66e-
CD34+ CD90+ CD45RA-
B
HSC
MEP
GMP
MLP
CD34+CD90+CD45RA−CD66e+
CD34+CD90+CD45RA−CD66e−
CD34+CD90+CD45RA−
CD34+
C
CD34+
CD34+CD90+CD45RA-CD66e-
CD34+CD90+CD45RA-CD66e+
2^-ΔΔCT mRNA /GAPDH
HES1
AVP
D
HSC
THY1 (CD90)
ACE (CD143)
CDCP1 (CD318)
AVP
PROM1 (CD133)
CD52
ITGA3
CEACAM1 (CD66a)
PROCR (EPCR)
HES1
CD34+CD90+CD45RA-CD66e
CD34+CD90+CD45RA-CD66e
CD34+CD90+CD45RA-
CD34+
E
-log10 (P value)
log2 Fold change
AVP
HES1
HLA-DQB1
HLA-DQA1
CDCP1
PROCR
ACE
CD38
GATA1
MKI67
MPO
ITGA3
CEACAM1
F
LT-HSC score
CD34+CD90+CD45RA−CD66e+
CD34+CD90+CD45RA−CD66e−
CD34+CD90+CD45RA−
CD34+
G
G2M score
S score
CD34+CD90+CD45RA−CD66e+
CD34+CD90+CD45RA−CD66e−
CD34+CD90+CD45RA−
CD34+

Figure 3. CD66e$^+$ cells associated with HSC signature

(A) Heatmap shows relative expression in population frequencies between donors organized according to annotated lineages by the integrated transcriptomic analysis of CD34$^+$CD90$^+$CD45RA$^-$CD66e$^+$/CD34$^+$CD90$^+$CD45RA$^-$CD66e$^-$/CD34$^+$CD90$^+$CD45RA$^-$/CD34$^+$ cells from CB. (For related gene list, see Table S1).

(B) Radar plot showing the average HSC and lineage scores for samples with different sorting strategies. The HSC and lineage scores were an average expression of lineage_HSC$^-$score gene sets. (For related gene list, see Table S2).

(C) Gene expression ($2^{-}\Delta\Delta$Ct) of HSC-related genes performed by qRT-PCR (one-way ANOVA). Data are represented as mean ± SEM. ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05.

(D) Heatmap showing relative expression patterns of HSC-related genes obtained by the integrated transcriptomic analysis of CD34$^+$CD90$^+$CD45RA$^-$CD66e$^+$/CD34$^+$CD90$^+$CD45RA$^-$CD66e$^-$/CD34$^+$CD90$^+$CD45RA$^-$/CD34$^+$ cells from CB.

(E) Scatterplot of global transcriptional profiling shows upregulated (red) and downregulated (blue) genes in CD66e$^+$ vs. CD66e$^-$ populations among CD34$^+$CD90$^+$CD45RA$^-$ cells. Significant genes were defined as genes with adjusted p value <0.05 and $|\log_2 FC| < \log_2(1.5)$.

(F) Dotplot showing the LT-HSC scores for each sample. The LT-HSC score was the average expression of LT-HSC core gene set.

(G) Scatterplot of S and G2M scores for each sample, indicating that CD34$^+$CD90$^+$CD45RA$^-$CD66e$^+$ cells are more quiescent than other groups. The S and G2M scores indicated an average expression of relative gene sets.

and secondary mice. These characteristics have a significant potential to advance future studies on human HSC biology and improve HSC-based therapies.

DISCUSSION

HSCs are crucial for lifelong blood cell production as they reside at the apex of hematopoiesis hierarchy. Reliable surface markers to identify HSCs facilitate the development of HSC gene therapy. Therefore, we still lack bonafide surface markers to define long-term HSCs. Here, we determined the CD66e expression enriched for human HSCs with increased repopulation and self-renewal capacity.

CD66e$^+$ cells display a robust repopulation ability for both short-term and long-term engraftment in immune-deficient mice. On the other hand, CD66e$^-$ cells show a significant reduction in reconstitution (Figures 2B, 2C, and 2G). Although lymphoid and myeloid lineage were detected in the bone marrow of CD66e$^-$ recipient mice, the proportion of lymphoid lineage was significantly reduced (Figure 2F). We found significantly less CD34$^+$ HSPCs among the BM of the primary recipient of the CD66e$^-$ group (Figure 2D). These data demonstrated that CD66e$^-$ cells can differentiate into hematopoietic lineages, but could not achieve long-term repopulation. Moreover, CEACAM5 is a highly glycosylated protein and belongs to the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family. It binds to the membrane via glycosylphosphatidylinositol (GPI), which is functionally described as an adhesion molecule.[26] GSEA and GO analysis also revealed a significant enrichment of transcripts associated with cell-cell adhesion, cytokine-cytokine receptor interaction, and positive regulation of hemopoiesis in CD66e$^+$ HSCs, highlighting the potential implication of CD66e in the crosstalk with the niche (Figures S6A–S6D). The adhesive properties of CEACAM5 allow HSCs to self-renew in response to niche signals that are often within a short range. Therefore, CD66e might play a fundamental role either in the anchorage of LT-HSCs to the bone marrow niche or in the mediating signaling pathway, ensuring the maintenance of the stem cell population in a hypoxic microenvironment. Thus, CD66e$^+$ cells better maintain and further differentiated *in vivo*.

Furthermore, several evidences in our study determined that CD34$^+$CD90$^+$CD45RA$^-$CD66e$^+$ cells were associated with LT-HSCs signature. CD66e$^+$ cells displayed the gene expression of hematopoietic cells on the top hierarchy (Figure 3A). Moreover, CD66e$^+$ cells had a higher score of LT-HSC signature instead of other progenitors (Figure 3B).[16,17] The recognized human HSC-related genes *AVP* and *HES1*, were also highly expressed in CD66e$^+$ cells (Figure 3C).[10,12] The gene expression profiling further determined that CD66e$^+$ cells were with LT-HSCs characterization and confirmed our functional analysis for long-term self-renewal ability on secondary colony-forming assays and secondary transplantation assays (Figures 1G and 2G).

Moreover, we confirmed that CD66e$^+$ cells were slow cycling cells with low metabolic features. According to the expression of cell cycle related genes, we found that CD66e$^+$ cells show lower S score and G2M score compared to other counterparts (Figure 3G). Moreover, cell-cycle inhibitor *CDKN1A* was upregulated while *CDKN2C* was downregulated in CD66e$^+$ HSCs (Figure S5A). As a key molecule for cell cycle entry, HSC exhaustion could be manipulated by *CDKN1A*,[18,19] while *CDKN2C* had the strongest inhibitory effect on HSC self-renewal.[20–22] Moreover, CD66e^{high} subset exhibited more cells in the G0 phase than the CD66e^{low} subset (Figures S4B and S4C). These results indicated that CD66e expression is associated with HSC quiescence. Previous studies have confirmed that HSCs rely on anaerobic glycolysis for energy production to minimize the production of reactive oxygen species and shift toward mitochondrial oxidative phosphorylation upon differentiation.[23–25] In accordance with their HSC phenotype, previously identified stem cell markers and cell quiescence gene sets were strongly enriched in CD66e$^+$ HSCs, while those associated with progenitor activity, ribosome assembly, G2/M checkpoints, and oxidative phosphorylation were significantly downregulated (Figure S6D).

Although this study was limited by using cord blood as the only source of human HSCs, we will test it in other samples such as bone marrow or mobilized peripheral blood in the future. Altogether, our study indicated that CD66e is a reliable surface marker of human HSCs, that further refines the immunophenotype of LT-HSCs. These characteristics have potential to advance further studies on HSC biology and offer new tools to enrich HSCs for cell and gene therapies.

Limitation of the study

This work identified CD66e as a potential surface marker of HSCs using human cord blood HSCs. We plan to test CD66e in other source of HSCs such as bone marrow or mobilized peripheral blood in the future.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - *In vivo* animal models
  - Cell culture
- METHOD DETAILS
  - Human $CD34^+$ cell isolation
  - Flow cytometry and sorting
  - Colony forming assay
  - Transplantation assay
  - RNA sequencing and RT-qPCR analysis
  - RNA sequencing data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108561.

ACKNOWLEDGMENTS

The authors thank Zhiyuan Hou, Jingjing Gong (Peking university) for taking care of the animals and technical assistance with mouse experiment; and Tieshan Wang (Beijing University of Chinese Medicine) for FACS screening of surface protein library; and Chunguang Han (National Center for Protein Science) for cell sorting.

This work was supported by National Key Research and Development Program of China (2021YFA1100900, 2022YFA1103500), Science, Technology & Innovation Project of Xiongan New Area (2022XAGG0142), Haihe Laboratory of Cell Ecosystem Innovation Fund (HH22KYZX0040, 22HHXBSS00001, 22HHXBSS00039, HH22KYZX0002), the National Natural Science Foundation of China (82370117, 81970104, 81890990, 82070193, 82222004, 82370222), the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (2021-I2M-1-040), the Non-CAMS Fundamental Research Funds for Central Research Institutes (3332021093), the CAMS Fundamental Research Funds for Central Research Institutes (2022-RC180-06). This work was supported by Center for Advanced Technologies, Haihe Laboratory of Cell Ecosystem.

AUTHOR CONTRIBUTIONS

K.M. designed and performed experiments and wrote the article; X. W. designed and performed experiments; L. W. performed bioinformatics analysis; L.Y performed the experiments and analyzed the data; J. Y. and X. L. performed the experiments in part assisted by X. W.; L.G., Z.S., H.Y., X.Z. performed the experiments; Y.Z. analyzed the RNA-seq data and edited the article; S.W. analyzed the RNA-seq data; P.Y. gave advice on experimental design; Y. Z., F. D., and S. H. provided discussion of the data; W.W. conceptualized the study and reviewed the article; R.F. contributed to experimental design and interpretation of data and edited the article; L. H. designed the research, analyzed data, and wrote the article; T. C. conceptualized the study and reviewed the article; and all authors approved the final version of the article.

DECLARATION OF INTERESTS

A patent has been filed relating to the data presented. W.W. is a founder and scientific adviser for EdiGene. R.F., P.Y., K.M., L.Y., L.G., Z.S., H.Y., X.Z., Y.Z., and S.W. are full-time employees of EdiGene Inc. when these studies were conducted. All other authors declare that they have no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 25, 2023
Revised: September 28, 2023
Accepted: November 21, 2023
Published: November 24, 2023

REFERENCES

1. 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.
2. Fares, I., Chagraoui, J., Lehnertz, B., MacRae, T., Mayotte, N., Tomellini, E., Aubert, L., Roux, P.P., and Sauvageau, G. (2017). EPCR expression marks UM171-expanded CD34+ cord blood stem cells. Blood *129*, 3344–3351.
3. Tomellini, E., Fares, I., Lehnertz, B., Chagraoui, J., Mayotte, N., MacRae, T., Bordeleau, M.È., Corneau, S., Bisaillon, R., and Sauvageau, G. (2019). Integrin-α3 Is a Functional Marker of Ex Vivo Expanded Human Long-Term Hematopoietic Stem Cells. Cell Rep. *28*, 1063–1073.e5.
4. Ansari, U., Tomellini, E., Chagraoui, J., Lehnertz, B., Mayotte, N., Bordeleau, M.E., Roux, P.P., and Sauvageau, G. (2022). CEACAM1 is a novel culture-compatible surface marker of expanded long-term reconstituting hematopoietic stem cells. Blood Adv. *6*, 3626–3631.
5. Dorrell, C., Gan, O.I., Pereira, D.S., Hawley, R.G., and Dick, J.E. (2000). Expansion of human cord blood CD34+CD38- cells in ex vivo culture during retroviral transduction without a corresponding increase in SCID repopulating cell (SRC) frequency: Dissociation of SRC phenotype and function. Blood *95*, 102–110.
6. Shin, J.W., Buxboim, A., Spinler, K.R., Swift, J., Christian, D.A., Hunter, C.A., Léon, C., Gachet, C., Dingal, P.C.D.P., Ivanovska, I.L., et al. (2014). Contractile forces sustain and polarize hematopoiesis from stem and progenitor cells. Cell Stem Cell *14*, 81–93.
7. 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 Cell. *329*, 1345–1348.
8. Zhang, Y., Xie, X., Huang, Y., Liu, M., Li, Q., Luo, J., He, Y., Yin, X., Ma, S., Cao, W., et al. (2022). Temporal molecular program of human hematopoietic stem and progenitor cells after birth. Dev. Cell *57*, 2745–2760.e6.
9. Hay, S.B., Ferchen, K., Chetal, K., Grimes, H.L., and Salomonis, N. (2018). The Human Cell Atlas bone marrow single-cell interactive web portal. Exp. Hematol. *68*, 51–61.
10. Yu, X., Alder, J.K., Chun, J.H., Friedman, A.D., Heimfeld, S., Cheng, L., and Civin, C.I. (2006). HES1 Inhibits Cycling of Hematopoietic Progenitor Cells via DNA Binding. Stem Cell. *24*, 876–888.
11. Mayer, B., Németh, K., Krepuska, M., Myneni, V.D., Tisdale, J.F., Hsieh, M.M., Uchida, N., Lee, H., Michael, J., Holmbeck, K., et al. (2017). Vasopressin stimulates the proliferation and differentiation of red blood cell precursors and improves recovery from anemia. Sci. Transl. Med. *9*, 1–22.
12. 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.
13. Conze, T. (2003). CDCP is a novel marker of hematopoietic stem cells. Ann. New York Acad. Sci. *5*, 222–226.
14. Jokubaitis, V.J., Sinka, L., Driessen, R., Whitty, G., Haylock, D.N., Bertoncello, I., Smith, I., Péault, B., Tavian, M., and Simmons, P.J. (2008). Angiotensin-converting enzyme (CD143) marks hematopoietic stem cells in human embryonic, fetal, and adult hematopoietic tissues. Blood *111*, 4055–4063.
15. Majeti, R., Park, C.Y., and Weissman, I.L. (2007). Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. Cell Stem Cell *1*, 635–645.
16. 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.
17. 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. Nat. Immunol. *14*, 756–763.
18. Cheng, T., Rodrigues, N., Shen, H., Yang, Y., Dombkowski, D., Sykes, M., and Scadden, D.T. (2000). Hematopoietic stem cell quiescence maintained by p21(cip1/waf1). Science *287*, 1804–1808.
19. Hao, S., Chen, C., and Cheng, T. (2016). Cell cycle regulation of hematopoietic stem or progenitor cells. Int. J. Hematol. *103*, 487–497.
20. Yuan, Y., Shen, H., Franklin, D.S., Scadden, D.T., and Cheng, T. (2004). In vivo self-renewing divisions of haematopoietic stem cells are increased in the absence of the early G1-phase inhibitor, p18INK4C. Nat. Cell Biol. *6*, 436–442.
21. Yu, H., Yuan, Y., Shen, H., and Cheng, T. (2006). Hematopoietic stem cell exhaustion impacted by p18INK4C and p21Cip1/Waf1 in opposite manners. Blood *107*, 1200–1206.
22. Orford, K.W., and Scadden, D.T. (2008). Deconstructing stem cell self-renewal: Genetic insights into cell-cycle regulation. Nat. Rev. Genet. *9*, 115–128.
23. Simsek, T., Kocabas, F., Zheng, J., Deberardinis, R.J., Mahmoud, A.I., Olson, E.N., Schneider, J.W., Zhang, C.C., and Sadek, H.A. (2010). The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. Cell Stem Cell *7*, 380–390.
24. Hu, L., Cheng, H., Gao, Y., Shi, M., Liu, Y., Hu, Z., Xu, J., Qiu, L., Yuan, W., Leung, A.Y.H., et al. (2014). Antioxidant N-acetyl-L-cysteine increases engraftment of human hematopoietic stem cells in immune-deficient mice. Blood *124*, e45–e48.
25. Morganti, C., and Ito, K. (2021). Mitochondrial contributions to hematopoietic stem cell aging. Int. J. Mol. Sci. *22*, 11117.
26. Eidelman, F.J., Fuks, A., DeMarte, L., Taheri, M., and Stanners, C.P. (1993). Human carcinoembryonic antigen, an intercellular adhesion molecule, blocks fusion and differentiation of rat myoblasts. J. Cell Biol. *123*, 467–475.
27. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
28. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.
29. Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
30. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
31. Oh, S., Abdelnabi, J., Al-Dulaimi, R., Aggarwal, A., Ramos, M., Davis, S., Riester, M., and Waldron, L. (2020). HGNChelper: Identification and correction of invalid gene symbols for human and mouse. F1000Res. *9*, 1493.
32. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523.
33. Hu, L., Yin, X., Zhang, Y., Pang, A., Xie, X., Yang, S., Zhu, C., Li, Y., Zhang, B., Huang, Y., et al. (2021). Radiation-induced bystander effects impair transplanted human hematopoietic stem cells via oxidative DNA damage. Blood *137*, 3339–3350.
34. Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., Dong, L., Zhang, Z., Yu, C., Sun, Y., et al. (2021). The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. Genomics, Proteomics Bioinforma *19*, 578–583.
35. CNCB-NGDC Members and Partners (2023). Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. Nucleic Acids Res. *51*, D18–D28.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Brilliant Violet 510(TM) anti-human CD34	Biolegend	Cat# 343528; RRID: AB_2563856
FITC anti-human CD90 (Thy1) Antibody	Biolegend	Cat# 328108; RRID: AB_893429
APC anti-human Lineage Cocktail (CD3, CD14, CD16, CD19, CD20, CD56)	Biolegend	Cat# 348810
Brilliant Violet 510™ anti-human CD34 Antibody	Biolegend	Cat# 343528; RRID: AB_2563856
Mouse Anti-CD38 Monoclonal Antibody, PE-Cy7 Conjugated	BD Biosciences	Cat# 335790; RRID: AB_399969
APC anti-human CD90 (Thy1) Antibody	Biolegend	Cat# 328114; RRID: AB_893431
FITC anti-human CD45RA Antibody	Biolegend	Cat# 304106; RRID: AB_314410
Anti-CEACAM5 Antibody (PE), Mouse Monoclonal	Sino Biological	Cat# 11077-MM02-P; RRID: AB_2860305
APC/Cyanine7 anti-human CD45 Antibody	Biolegend	Cat# 304014; RRID: AB_314402
FITC anti-mouse CD45 Antibody	Biolegend	Cat# 103107; RRID: AB_312972
Brilliant Violet 510™ anti-human CD3 Antibody	Biolegend	Cat# 300448; RRID: AB_2563468
PE anti-human CD19 Antibody	Biolegend	Cat# 363004; RRID: AB_2564126
APC anti-human CD56 (NCAM) Antibody	Biolegend	Cat# 304610; RRID: AB_314452
Brilliant Violet 421™ anti-human CD33 Antibody	Biolegend	Cat# 303416; RRID: AB_2561690
Brilliant Violet 510(TM) anti-human CD34	Biolegend	Cat# 343528; RRID: AB_2563856
Biological samples		
Human umbilical cord blood	Ori Biotech, Hycells	N/A
Chemicals, peptides, and recombinant proteins		
Human stem cell factor	PeproTech	Cat: 300-07
Human thrombopoietin	PeproTech	Cat: 300-18
Human FLT3 ligand	PeproTech	Cat: AF-300-19
Human IL-6	PeproTech	Cat: 200-06
StemSpan SFEM II	STEMCELL Technologies	Cat: 09605
MethoCult	STEMCELL Technologies	Cat: H4034
Cell lysis buffer	Biolegend	Cat: 420301
TRIzol reagent	Invitrogen	Cat: 15596018
7-AAD Viability Staining Solution	Biolegend	Cat: 420404
Critical commercial assays		
$CD34^+$ microbead kit	Miltenyi Biotec	Cat: 130-046-702
Lyoplate screening panel	BD Bioscience	Cat: 560747
NEBNext Poly(A) mRNA Magnetic Isolation Module	NEB	Cat: E7490
NEBNext Ultra II RNA Library Prep Kit for Illumina	NEB	Cat: E7770
FastStart Essential DNA Green Master Mix	Roche	Cat: 06402712001
Deposited data		
https://ngdc.cncb.ac.cn/gsa-human	Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences	GSA-Human: HRA004527

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
NOD.Cg-Prkdc^{scid} Il2rg^{tm1Vst}/Vst (NPG) mice	Vital Star	N/A
Software and algorithms		
Flowjo v10	BD Biosciences	RRID:SCR_008520; https://www.flowjo.com
fastp v0.19.6	Chen et al.[27]	https://github.com/OpenGene/fastp
HISAT2 v2.0.5	Kim et al.[28]	https://github.com/DaehwanKimLab/hisat2
featureCounts v1.5.0-p3	Liao et al.[29]	http://subread.sourceforge.net/
DESeq2 v1.20.0	Love et al.[30]	https://bioconductor.org/packages/release/bioc/html/DESeq2.htm
HGNChelper v0.8.1	Oh et al.[31]	https://github.com/waldronlab/HGNChelper
GSEA desktop application v4.1.0	Broad Institute	http://www.broad.mit.edu/GSEA
Metascape v3.5	Zhou et al.[32]	https://metascape.org/
R v4.1.0	R Core Team	https://www.r-project.org/
Pheatmap v1.0.12	The Comprehensive R Archive Network	https://CRAN.R-project.org/package=pheatmap
GraphPad Prism v8	Graphpad Software	https://www.graphpad.com

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Linping Hu (hulinping@ihcams.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

We are committed to ensuring data transparency and will provide access to the data reported in this paper. It will be deposited in an online repository upon publication, and interested researchers can also request the data directly from the lead contact.

The raw sequence data reported in this paper have been deposited at Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2022), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA004527) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human. Details of the analysis are provided in the STAR Methods section. This paper does not use customized code and does not report original code.

Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

In vivo animal models

In vivo assay was performed in NOD.Cg-Prkdcscid Il2rgtm1Vst/Vst (NPG) mice, acquired from Vital Star (Beijing, China). Five-week-old female mice were used in the experiments. Mice were maintained at the a specific-pathogen-free facility. All experiments involving animals were approved by the institute of Hematology Animal Care and Use Committee.

Cell culture

Human umbilical cord blood (CB) was obtained from commercial sources (Ori Biotech, Hycells). Human CB CD34$^+$ hematopoietic stem and progenitor cells (HSPCs) were cultured in HSC expansion media consisting of StemSpan SFEM II (STEMCELL Technologies), supplemented with human stem cell factor (SCF), thrombopoietin (TPO), FLT3 ligand (FLT3L) and IL-6. Single-cell assays for myeloid cell (My), erythrocyte (Ery) and natural killer cell (NK) differentiation were conducted as described previously.[8] The cells were maintained at 37°C with 5% CO2.

METHOD DETAILS

Human CD34$^+$ cell isolation

Isolation of CD34$^+$ cells was performed as previously.[33] In short, mononuclear cells were isolated by Ficoll-Hypaque density gradient centrifugation, and CD34$^+$ cell enrichment was performed using a CD34$^+$ microbead kit (Miltenyi Biotec, 130-046-702).

Flow cytometry and sorting

Human cell surface markers were screened using Lyoplate screening panel (BD Biosciences, 560747) together with human antibodies CD34 BV510 and CD90 FITC. The cells were stained in fluorescence-activated cell sorting (FACS) buffer consisting of phosphate-buffered saline (PBS) supplemented with 1% bovine serum albumin (BSA) and 0.5M EDTA. Finally, the stained cells were analyzed on BD LSRFortessa SORP (BD Biosciences).

To identify the phenotype, cells were stained in FACS buffer using the following human antibodies: Human Lineage Cocktail FITC (CD3, CD14, CD16, CD19, CD20, CD56, Biolegend), CD34 BV510, CD38 PE-CY7, CD90 APC, CD45RA FITC, CD66e PE and 7AAD (please see STAR Methods for details). The stained cells were washed once and analyzed using CytoFLEX (Beckman Coulter). To isolate the cell subpopulation for *in vivo* engraftment or RNA-sequencing, stained cells were sorted using BD Aria SORP (BD Biosciences).

Colony forming assay

The frequencies of colony-forming cells were estimated by plating 150 CD66e^{high} or CD66e^{low} cells sorted from Lin$^-$CD34$^+$ populations at MethoCult (Stem Cell Technologies). The colonies were counted during 12-14 days. The colonies were harvested, and 10000 cells from each well were replated into a secondary methylcellulose culture. The secondary cultures were incubated for 2 weeks at 37°C before determination of colony numbers and types.

Transplantation assay

Five-week-old female NPG mice were sub-lethally irradiated (1.6 Gy, within 12 h before transplantation) and transplanted with human HSPCs via tail vein injection. Sixteem weeks after transplantation, the recipients were sacrificed, and the engraftment of human CB cells in the bone marrow (BM), peripheral blood (PB), and spleen was analyzed by flow cytometry. For the limiting dilution assay (LDA), cells were transplanted at three dilutions (300, 1000, and 3000 cells) in each mouse. The engraftment rate is % hCD45+ cell/ (% hCD45+ cells + % mCD45+ cells).

For secondary transplantation, 40% of total BM cells which were from primary recipients transplanted with 3000 cells were injected into secondary sub-lethally irradiated NPG mice. The BM cells of recipient mice were harvested and analyzed 16 weeks post-transplantation. Flow cytometry analysis was performed on BM, PB and spleen cells. Cells were treated with red blood cell lysis buffer (Biolegend, 420301), washed and stained with antibodies: human CD45 APC-Cy7, mouse CD45 FITC, human CD3 BV510, human CD19 PE, human CD56 APC, human CD33 BV421, and 7AAD (please see STAR Methods for details). Data were analyzed using CytoFLEX (Beckman Coulter). FlowJo was used to analyze the flow cytometry data.

RNA sequencing and RT-qPCR analysis

CD34$^+$CD90$^+$CD45RA$^-$CD66e$^{+/-}$, CD34$^+$CD90$^+$CD45RA$^-$, and CD34$^+$ cells were sorted at a density of 2-5×10^5 cells in each group from CB CD34$^+$ HSPCs. RNA was extracted using TRIzol reagent (Invitrogen, 15596018). RNA-sequencing was performed by Novogene Co. Briefly, poly(A)+ mRNA was enriched from 200 ng total RNA using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490), and the libraries were generated using NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB, E7770) following manufacturer's recommendations. Dual indexed libraries were pooled sequenced on an Illumina NovaSeq6000 platform in PE150 manner, around 20 million paired reads were generated for each sample. Replicates were included for each group to eliminate nonspecific effects. Gene Ontology (GO) enrichment analysis and Heatmap generation of differentially expressed genes were carried out by the clusterProfiler R package. Gene set enrichment analysis (GSEA) with the Kyoto Encyclopedia of Genes and Genomes (KEGG) dataset was used for cellular pathway analysis. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA004527) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human.[34,35]

RT-qPCR was performed according to the instruction for FastStart Essential DNA Green Master Mix (Roche, 06402712001) on a LightCycler 96 PCR system (Roche). The expressed values relative to control were calculated using $2^{-\Delta\Delta CT}$. *GAPDH* was used as a housekeeping gene for normalization.

RNA sequencing data analysis

The raw sequencing data was quality controlled using fastp (v0.19.6) with *–detect_adapter_for_pe -g -x -q 20 -u 25 -l 50* parameters, to remove adapter-contaminated and low-quality reads with too much N, low-quality bases or polymer tails. The filtered reads were mapped to the human reference genome (GRCh38.p12) with the corresponding gene model annotation files (Ensembl 96) using HISAT2 software (v2.0.5), and read counts were obtained from the alignment with featureCounts (v1.5.0-p3). Fragments per kilobase million (FPKM) values of each gene was calculated based on the read count and gene length to assess gene expression levels, and differential expression analysis between groups were performed using the DESeq2 R package (v1.20.0). Differentially expressed genes were selected based on significance (adjusted $p<0.05$,

two-tailed Mann-Whitney U test, adjusted by Benjamini-Hochberg procedure), their mean expression level in at least one of the comparison groups (FPKM≥0.5), and a minimum 1.5-fold expression difference. All gene names were curated using the HGNChelper R package (v0.8.1) to facilitate comparison with external datasets. GSEA analysis was carried out using the GSEA desktop application (v4.1.0, Broad Institute) and metascape online tool (v3.5, https://metascape.org/gp/index.html#/main/step1). The GSEA hallmark gene set collection and two curated HSC gene sets were acquired by standard settings. The overlap between genes associated with *EPCR* expression and published HSC signatures was determined using the intersect command in R (v4.1.0). Heat maps were generated using R function *pheatmap* (v1.0.12). Principal component analysis was performed using the R function *prcomp* and scaled FPKM values.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of all experiments was carried out using GraphPad Prism *version* 8. All data are presented as mean ± standard error of the mean (SEM). One-way analysis of variance (ANOVA) and t-test were used for sample comparison. *p*-values <0.05 indicated a significant difference. ****p<0.0001, ***p<0.001, **p<0.01, *p<0.05.