Medial *HOXA* genes demarcate haematopoietic stem cell fate during human development

Diana R. Dou\(^1,2,3,8\), Vincenzo Calvanese\(^1,2,8\), Maria I. Sierra\(^1,2\), Andrew T. Nguyen\(^1\), Arazin Minasian\(^1,2\), Pamela Saarikoski\(^1,2\), Rajkumar Sasidharan\(^1,2\), Christina M. Ramirez\(^4\), Jerome A. Zack\(^2,5,6\), Gay M. Crooks\(^1,2,7\), Zoran Galic\(^2,5\) and Hanna K. A. Mikkola\(^1,2,3,9\)

Pluripotent stem cells (PSCs) may provide a potential source of haematopoietic stem/progenitor cells (HSPCs) for transplantation; however, unknown molecular barriers prevent the self-renewal of PSC-HSPCs. Using two-step differentiation, human embryonic stem cells (hESCs) differentiated in vitro into multipotent haematopoietic cells that had the CD34\(^+\)CD38\(^-\)/CD90\(^+\)/CD45\(^+\) GPI-80\(^+\) fetal liver (FL) HSPC immunophenotype, but exhibited poor expansion potential and engrafment ability. Transcriptome analysis of immunophenotypic hESC-HSPCs revealed that, despite their molecular resemblance to FL-HSPCs, medial *HOXA* genes remained suppressed. Knockdown of *HOXA7* disrupted FL-HSPC function and caused transcriptome dysregulation that resembled hESC-derived progenitors. Overexpression of medial *HOXA* genes prolonged FL-HSPC maintenance but was insufficient to confer self-renewal to hESC-HSPCs. Stimulation of retinoic acid signalling during endothelial-to-haematopoietic transition induced the *HOXA* cluster and other HSC/definitive haemogenic endothelium genes, and prolonged HSPC maintenance in culture. Thus, medial *HOXA* gene expression induced by retinoic acid signalling marks the establishment of the definitive HSPC fate and controls HSPC identity and function.

Haematopoietic stem cells (HSCs) regenerate the blood system on transplantation, and can therefore cure inherited and acquired blood diseases. However, lack of HLA (human leukocyte antigen)-matched bone marrow or cord blood donors limits their therapeutic use\(^1\). Generation of HSCs from human embryonic stem cells (hESCs) or induced pluripotent stem cells (PSCs) could provide alternative HSC sources. Recent studies used transcription factor reprogramming to convert fibroblasts or mature blood cells into multipotent haematopoietic cells that had the CD34\(^+\)CD38\(^-\)/CD90\(^+\)/CD45\(^+\) GPI-80\(^+\) fetal liver (FL) HSPC immunophenotype, but exhibited poor expansion potential and engrafment ability. Transcriptome analysis of immunophenotypic hESC-HSPCs revealed that, despite their molecular resemblance to FL-HSPCs, medial *HOXA* genes remained suppressed. Knockdown of *HOXA7* disrupted FL-HSPC function and caused transcriptome dysregulation that resembled hESC-derived progenitors. Overexpression of medial *HOXA* genes prolonged FL-HSPC maintenance but was insufficient to confer self-renewal to hESC-HSPCs. Stimulation of retinoic acid signalling during endothelial-to-haematopoietic transition induced the *HOXA* cluster and other HSC/definitive haemogenic endothelium genes, and prolonged HSPC maintenance in culture. Thus, medial *HOXA* gene expression induced by retinoic acid signalling marks the establishment of the definitive HSPC fate and controls HSPC identity and function.

---

1. Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, California 90095, USA.
2. Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, California 90095, USA.
3. Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California 90095, USA.
4. Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, California 90095, USA.
5. Department of Medicine, Division of Hematology-Oncology, University of California, Los Angeles, Los Angeles, California 90095, USA.
6. Department of Immunology, University of California, Los Angeles, Los Angeles, California 90095, USA.
7. Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA.
8. These authors contributed equally to this work.
9. Correspondence should be addressed to H.K.A.M. (e-mail: hmikkola@mcdb.ucla.edu)

Received 22 December 2014; accepted 8 April 2016; published online 16 May 2016; DOI: 10.1038/ncb3354

© 2016 Macmillan Publishers Limited. All rights reserved
identify regulatory cues and molecular landmarks that distinguish the definitive HSC fate from the short-lived embryonic progenitors. We used a two-step hESC differentiation to generate HSPCs with human fetal HSC surface phenotype (CD45\(^+\)CD34\(^+\)CD38\(^-\)/loCD90\(^+\)GPI-80\(^+\)). Molecular profiling showed remarkable resemblance of hESC-HSPCs to FL-HSPCs, yet revealed distinct differences in HSC regulatory programs, including the HOXA genes. Knockdown and overexpression studies revealed that medial HOXA genes, in particular HOXA7, govern definitive HSC identity and function. Rescue of retinoic acid (RA) signalling during endothelial-to-haematopoietic transition induced medial HOXA genes and the definitive haematopoietic program in hESC-HSPCs. These studies uncover regulatory programs that distinguish human definitive HSC lineage from embryonic progenitors and offer a blueprint for identifying missing cues required for HSC generation.

**RESULTS**

**hESCs generate HSPCs that are unable to engraft**

To identify barriers for generating HSCs from PSCs, we compared hESC-derived HSPCs and human fetal HSCs on the basis of immunophenotypic, functional and molecular criteria. Second-trimester FL-HSCs were used as controls as they are ontologically closer to PSC-derived cells than cord blood or bone marrow. A two-step differentiation protocol was employed to generate haematopoietic cells from hESCs (Fig. 1a). To promote haemato-vascular differentiation, embryoid bodies (EBs) were cultured with BMP4 (days 4–10), FLT3L and SCF (days 4–14). Day-14 EBs had generated cells that co-expressed CD34, CD90 and CD43 (Fig. 1b), but not CD45. To promote the development of haemo-vascular precursors towards HSCs, EB-derived CD34\(^+\) cells were plated with TPO, SCF and FLT3L on OP9-M2 stroma, which supports the expansion of multipotent human HSPCs. Two-week OP9-M2 co-culture (EB-OP9) generated cells with CD34\(^+\)CD38\(^-\)/loCD90\(^+\)CD45\(^+\) HSPC immunophenotype (Fig. 1b) and some co-expressed the HSC marker GPI-80 (Fig. 1c). These data showed that two-step differentiation generates cells with the immunophenotype of human fetal HSCs.

To assess whether the hESC-HSPC maturation culture on OP9-M2 stroma confers functional properties of HSCs, CD34\(^+\) cells from EBs (isolated directly or after two-step differentiation), and FL (isolated directly or after 2-week culture) were transplanted into sublethally irradiated NOD-scid IL2R\(\gamma\)-null (NSG) mice (Supplementary Fig. 1A). Human CD45\(^+\) chimera in bone marrow was measured 12 weeks post-transplantation. Although FL-HSPCs engrafted successfully before or after OP9-M2 culture, hESC-derived cells showed minimal engraftment (Fig. 1d). Human CD45\(^+\) cells in the bone marrow of mice transplanted with FL contained HSPCs (Supplementary Fig. 1B), CD19\(^+\) B cells, CD3\(^+\) T cells and CD13\(^+\) or CD66\(^+\) myeloid cells, whereas the mice transplanted with hESC-derived cells harboured only rare human myeloid cells (Fig. 1e). These data show that hESC-HSPCs are severely impaired functionally.

**hESC-HSPCs have poor proliferative potential**

To understand the functional defects in hESC-HSPCs, hESC- and cultured FL-HSPCs (CD34\(^+\)CD38\(^-\)/loCD90\(^+\)CD45\(^+\)) were sorted and re-plated on OP9-M2 co-culture to assess their expansion (Fig. 2a). Both FL- and hESC-HSPC cultures maintained an immunophenotypic HSPC population 1 week later (Fig. 2b,c); however, at 3 weeks, hESC-HSPCs had disappeared (Fig. 2b,c). BrdU incorporation analysis did not reveal differences in cell cycle between FL- and hESC-HSPCs (Supplementary Fig. 2A), suggesting that loss of hESC-HSPCs was not due to inability to divide.

hESC-HSPCs also produced fewer clonogenic progenitors on OP9-M2 than FL-HSPCs (Fig. 2d), although both formed erythroid (BFU-E), granulocyte-macrophage (CFU-GM), macrophage (CFU-M) and mixed myelo-erythroid (CFU-mixed) colonies (Fig. 2e). CFU-C potential in OP9-M2 was higher in the CD34\(^+\)CD38\(^-\)/loCD90\(^+\)CD45\(^+\) fraction than total CD34\(^+\) cells (Supplementary Fig. 2B), which also showed minimal CFU-C expansion (Supplementary Fig. 2C). These data indicate that hESC-derived immunophenotypic HSPCs exhibit poor proliferative potential.

Previous studies documented impaired differentiation of hESCs to adult-type erythroid cells. Colonies from EB CD34\(^+\) cells generated erythroid cells that predominantly expressed embryonic \(\gamma\)-globin, whereas fetal \(\gamma\)-globin (HBG1) was low (16.3% of FL levels) and adult \(\beta\)-globin (HB\(B\)) nearly undetectable (1.8% of FL levels). However, colonies derived from EB CD34\(^+\) cells after 2 weeks on OP9-M2 induced some expression of \(\gamma\)- and \(\beta\)-globins (84.9% and 54.0% of FL-derived cells, respectively; Supplementary Fig. 2D). \(\beta\)-globin induction suggests that hESC-HSPCs can differentiate to adult-type erythroid cells if given adequate maturation time in culture.

T-cell generation is another hallmark of definitive haematopoiesis. Both EB- and FL-derived CD34\(^+\) cells generated CD4 and CD8 single- and double-positive T-lymphoid cells in OP9-DL1 culture (Supplementary Fig. 2E). These results imply that the main defect in hESC-derived haematopoietic cells is not differentiation, but self-renewal.

**Microarray uncovers molecular defects in hESC-HSPCs**

To understand the defective self-renewal of hESC-HSPCs, we assessed their relationship to primary human fetal HSPCs using microarray analysis. Gene expression differences were compared between immunophenotypic HSPCs (CD34\(^+\)CD38\(^-\)/loCD90\(^+\)CD45\(^+\)) derived from hESCs through EB or two-step differentiation (EB-OP9), and from FL. To assess culture-induced effects, FL-HSPCs were cultured on OP9-M2 for 2 or 5 weeks (FL-OP9). CD34\(^+\)CD38\(^-\)/loCD90\(^+\)CD45\(^+\) cells from early human placenta (PL) were used as a reference for immature human HSPCs (Supplementary Table 1). Spearman rank coefficient comparisons and hierarchical clustering of samples revealed that FL-HSPCs are more similar to EB-OP9-HSPCs than EB- or PL-HSPCs (Fig. 3a,b). EB-OP9-HSPCs were most similar to cultured FL-HSPCs, implying that hESC-HSPC transcriptome was influenced by prolonged culture (Fig. 3a). Many transcription factors governing the development of definitive HSPCs (SCL/TAL1, GATA2, RUNX1, MYB, ETV6, HOX4, GFI1B, BCL11A and so on) were expressed in both EB-OP9-HSPCs and FL-HSPCs (Fig. 3c). These data reveal that EB-OP9-HSPCs are remarkably similar to FL-HSPCs at the molecular level.

To identify co-regulated and differentially expressed programs between samples, K-means clustering and DAVID GO (gene ontology) analysis was performed for all differentially expressed genes (\(>2\)-fold, \(P\) value <0.05; Supplementary Tables 2 and 3). K-means clustering
Figure 1 Two-step culture of hESCs generates immunophenotypic HSPCs that engraft poorly. (a) Culture and isolation strategy for differentiating H1 hESCs to HSPCs. (b) Representative FACS plots from 11 experiments staining for CD34, CD90, CD38, CD45 and CD43 on hESC-derived CD34+ cells isolated from 2-week EBs (EB) and after 2-week maturation culture on OP9-M2 (EB-OP9), compared with cells from second-trimester fetal liver that were isolated directly (FL) or cultured on OP9-M2 (FL-OP9). (c) Representative FACS plots from nine experiments staining for CD38, CD34, CD90 and human fetal HSC self-renewal marker GPI-80 on hESC- and FL-derived cells. (d) Human engraftment in NSG mice with hESC-derived and FL-derived CD34+ cells, before and after OP9-M2 co-culture (individual values and mean are shown; n = 5 EB, n = 4 EB-OP9 and FL, and n = 3 FL-OP9 transplanted mice; statistical significance was calculated using the Wilcoxon rank sum test; see Supplementary Table 7 for statistics source data). (e) Representative FACS plots showing the human CD45+ fraction in the mouse bone marrow 12 weeks post-transplantation. Multi-lineage engraftment is assessed by CD19 and CD3 (B and T lymphoid), and CD66 and CD13 (myeloid) stainings.

showed that OP9-M2 co-culture of EB-derived cells induced many FL-HSPC-associated genes (Fig. 3d, clusters 2 and 9) that encoded DNA-repair factors and transcriptional regulators (Fig. 3e), whereas vascular genes were downregulated (Fig. 3d, clusters 5 and 6, Fig. 3f).

K-means clustering also identified genes that remained dysregulated in hESC-HSPCs. The most striking differences were observed in clusters 4 and 8 that contained genes highly expressed in FL-HSPCs but suppressed in hESC-HSPCs and early PL. The most highly enriched GO category was ‘Transcription’ (Fig. 3d), which included many HOXA genes. All HOXA genes except for HOXA13 (a regulator of placental vascular labyrinth specification) were severely suppressed in hESC-derived HSPCs (Fig. 3g). A similar
**Figure 2** hESC-derived haematopoietic cells have limited proliferative potential in vitro. (a) Strategy for comparing the expansion of hESC- and FL-HSPCs. (b) FACS staining for the HSPC surface markers CD34<sup>+</sup>, CD38<sup>lo</sup>, CD45<sup>+</sup> and CD90<sup>+</sup> at various time points in OP9-M2 coculture. (c) Expansion of FL- and hESC-derived haematopoietic cells sorted for HSPC phenotype after two-step differentiation and cultured for additional weeks on OP9-M2. Mean ± s.e.m. (upward bars) from n=4 experiments; statistical significance was assessed using the Wilcoxon rank sum test. (d) CFU-C expansions from 10,000 hESC-derived or FL-derived immunophenotypic HSPCs in methylcellulose following 0, 1 and 3 additional weeks on OP9-M2 co-culture. Mean ± s.e.m. (upward bars) from n=4 experiments; statistical significance was assessed using the Wilcoxon rank sum test. (e) The morphology of myelo-erythroid colonies generated from hESC- or FL-HSPCs on methylcellulose as assessed by light microscopy and May–Grünwald–Giemsa (MGG) staining. E, erythroid; M, macrophage; GM, granulocyte-macrophage; mixed, mixed myelo-erythroid. Statistics source data used to generate graphs in c and d can be found in Supplementary Table 7.

**HOXA pattern was observed with early PL, raising the hypothesis that medial HOXA gene induction in FL-HSPCs reflects developmental maturation and acquisition of HSC properties. Clusters 4 and 8 also included the HSC regulators HMGN, HLF, PRDM16 and MECOM/EVI1<sup>29,30</sup>, and the HSC surface markers PROM1/CD133, EMCN and ROBO4 (Fig. 3h). These analyses demonstrated that the two-step conditioning fails to induce transcriptional regulators highly expressed in FL-HSCs, including HOXA genes.**

**Medial HOXA genes govern human HSPC function**

We next asked whether HOXA gene silencing contributes to the poor self-renewal of hESC-HSPCs. Microarray analysis of human
FL-HSC subsets documented expression of several medial HOXA genes in GPI-80 HSCs and their immediate progeny (Fig. 4a), and downregulation on differentiation (Fig. 4b). As HOXA9 is a known regulator of mouse HSC proliferation, FL-HSPCs were transduced with pLKO.1 shRNA lentiviral vectors targeting HOXA5 or HOXA7 (Fig. 4c) to test whether other medial HOXA genes regulate human...
Figure 4 Medial HOXA genes govern the function and identity of human fetal HSPCs. (a, b) Microarray analysis of HOXA gene expression in CD34+CD38−/−CD90+GPI-80− cells and their progeny (mean values are shown; left, n = 3 samples, GEO database GSE54316; and right, n = 3 samples (CD34+CD38−CD90+) or 2 (CD34+CD38−CD90− and CD34−CD38−), GSE34974. (c) A schematic showing the strategy for lentiviral shRNA knockdown of HOXA5 or HOXA7 in FL-HSPCs. (d) Knockdown is confirmed using qRT-PCR 1 week post-infection (mean ± s.d. shown from n = 3 different FL samples). (e) Representative FACS plots 30 days after HOXA5 or HOXA7 knockdown. (f) Quantification of HSPC subsets in empty-vector (Ctrl)- and shRNA-infected cells (HOXA5 shRNA or HOXA7 shRNA) after 5, 14 and 30 days in culture (mean and s.e.m., n = 6 independent experiments per condition for day 14 and n = 3 for days 5 and 30). Statistical significance was assessed using the Wilcoxon signed rank test. (g) A schematic showing the transplantation strategy with HOXA5- or HOXA7-knockdown FL-HSPCs. (h) Representative FACS plots from mouse bone marrow 10 weeks post-transplantation assessing human CD45+ cells and multi-lineage engraftment (CD19 and CD3 for B and T lymphoid (middle column), and CD66 and CD33 for myeloid (right column)). (i) Quantification of human engraftment (n = 9 mice per condition from three independent experiments; individual values and mean are shown). Statistical significance was assessed using the Wilcoxon rank sum test. (j) RNA sequencing of HOXA7-knockdown FL-HSPCs at day 5 post-infection. Numbers of genes up- or downregulated in HOXA7 shRNA FL-HSPCs are shown. Genes dysregulated both in HOXA7-knockdown FL-HSPCs (RNA-Seq 1.8-fold change, n = 4 independent experiments, P value < 0.05) and in EB-OP9-HSCs compared with FL-HSPCs (microarray, twofold change, P < 0.05) are shown in blue pattern overlay. (k) Bar plots showing gene expression of HSC factors downregulated in HOXA7-knockdown FL-HSPCs (l) and differentiation-associated genes upregulated (l) in HOXA7-knockdown FL-HSPCs. Mean fragments per kilobase of exon per million fragments mapped (FPKM) from n = 4 independent specimens are shown; values used to generate graphs can be found in Supplementary Table 4 and GEO database GSE76685. See Supplementary Table 7 for statistics source data for d, f, i.
HSCs. Knockdown was confirmed using quantitative PCR with reverse transcription (qRT-PCR) 1 week post-transduction (Fig. 4d and Supplementary Table 6). By 2–4 weeks, HOXA5 and HOXA7 shRNA-treated cells were depleted of HSPCs (Fig. 4e,f). Cell cycle analysis 1 week post-transduction did not reveal significant differences in BrdU incorporation of HOXA5- or HOXA7-knockdown HSPCs (Supplementary Fig. 3A), implying that HOXA5- and HOXA7-deficient HSPCs can divide, but cannot self-renew. Transplantation of HOXA5 and HOXA7 shRNA-transduced FL-HSPCs into NSG mice (Fig. 4g) showed minimal engraftment (Fig. 4h,i). These results suggest that both HOXA5 and HOXA7 are necessary for human FL-HSPC expansion in vitro and reconstitution in vivo. HOXA7 showed stronger phenotypes in both assays.

To investigate how HOXA7 regulates human HSCs, pLKO1-control- and HOXA7-shRNA-transduced FL-HSPCs from four different FL tissues were sorted 5 days post-infection and subjected to RNA sequencing (Fig. 4e). Five hundred significantly differentially expressed genes (306 upregulated, 194 downregulated, 1.8-fold, \( P < 0.05 \)) between HOXA7-knockdown and pLKO1-control HSPCs were identified (Fig. 4j and Supplementary Table 4). Comparison with hESC-HSPC microarray data revealed that 30.1% of the genes upregulated on HOXA7 knockdown were also significantly upregulated in EB-OP9-HSPCs compared with FL-HSPCs, and 34.0% of the downregulated HOXA7-dependent genes showed low expression in EB-OP9 HSPCs. The shared downregulated genes included HSC regulators HOXA9-10, HLF, and HMGN, and HSC surface proteins PROM1/CD133 and MPL (Fig. 4k). HOXA7-knockdown HSPCs and EB-OP9-HSPCs also upregulated genes associated with megakaryocytic and erythroid differentiation (Fig. 4l). The shared upregulated genes also included cell cycle inhibitors CDKN1A (p21Cip1) and CDKN2D (p19Ink4d), whereas other cell cycle regulators and proliferation markers were unaffected at the messenger RNA level (Supplementary Fig. 3B,C). These results imply that HOXA7 activates factors regulating definitive HSC identity and suppresses programs associated with differentiation-primed embryonic progenitors.

### HOXA gene overexpression expands FL-HSPCs

We next assessed whether HOXA gene overexpression improves in vitro expansion of FL-HSPCs and hESC-HSPCs. Tetracycline-inducible PNL lentiviral vector (Supplementary Fig. 4A), which induced 30-fold overexpression of HOXA5 or HOXA7 in FL-HSPCs (Supplementary Fig. 4B), prolonged FL-HSPC maintenance in culture with both HOXA5 and HOXA7 (Supplementary Fig. 4C,D). Constitutively active FUGW vector\(^4\), which resulted in 4–5-fold transgene overexpression, prolonged FL-HSPC maintenance with HOXA7 (Fig. 5a–e). HSPCs overexpressing HOXA7 genes showed comparable differentiation ability in colony assays, implying that HOXA5 or HOXA7 expression does not prevent differentiation (Fig. 5f).

Although HOXA5 or HOXA7 overexpression promoted FL-HSPC expansion, their overexpression in EB-derived CD34\(^+\) cells did not improve hESC-HSPC expansion in vitro (Fig. 5g–i) and engraftment in vivo (Fig. 5j,k). Similar results were observed overexpressing HOXA5, HOXA7 and HOXA9 simultaneously. RNA-sequencing of HOXA7-overexpressing hESC-HSPCs showed that despite the confirmed HOXA7 overexpression (Fig. 5i), there was minimal change in putative HOXA7 target genes (Fig. 5m). Altogether, these studies showed that medial HOXA genes promote HSPC expansion when expressed in a correct cellular context.

### RA signalling induces HOXA genes in hESC-HSPCs

We next sought for upstream regulators that could specify the definitive HSPC fate and induce HOXA genes in a correct cellular context. Medial HOXA genes are developmentally regulated by RA signalling in other cell types.\(^3,7,38\) RA generated by ALDH1A2 (RALDH2) in haemogenic endothelium and its binding to RA receptor alpha (RARA) is necessary for generating HSCs in mouse AGM.\(^39\) Microarray analysis showed robust expression of RARA, and to a lesser extent RARB and RARG in HSPCs at different stages. Notably, RALDH2 was expressed in CD34\(^+\)CD38\(^{−}/lo\)CD90\(^+\)CD43\(^+\) cells in the early placenta, but not in EB. FL-HSPCs did not express RALDH2, consistent with its function in haemogenic endothelium (Fig. 6a). RALDH1 was expressed in FL-HSPCs but at low levels in both EB- and PL-derived cells. These results nominated defective RA signalling during hESC haematopoietic specification as a potential barrier for inducing HOXA genes.

We next asked whether RALDH2 function could be bypassed by administering all-trans retinoic acid (ATRA) or the RARA agonist AM580\(^4\) during hESC differentiation. CD34\(^+\) cells were isolated from EBs at 2 weeks and cultured for 6 days on OP9-M2 with or without ATRA (1.0 \(\mu M\)) or AM580 (0.2 \(\mu M\)) (Fig. 6b). qRT-PCR analysis showed that treatment of EB-derived CD34\(^+\) cells with AM580 or ATRA induced several medial HOXA genes (Fig. 6c). AM580 had a more robust effect on HOXA genes and was used for further studies.

Fluorescence-activated cell sorting (FACS) analysis at 6 days of AM580 treatment revealed strong induction of CD38, a known RA target\(^4\), in hESC- and FL-HSPCs (Fig. 6d). Persistent CD90 expression suggested that they had not differentiated. AM580-induced CD38 expression was reversible (Fig. 6d), and AM580-treated EB CD34\(^+\) cells typically retained a higher fraction of CD34\(^+\) cells with AM580 or ATRA agonist induced several medial HOXA genes (Fig. 6c). AM580 had a more robust effect on HOXA genes and was used for further studies.

Fluorescence-activated cell sorting (FACS) analysis at 6 days of AM580 treatment revealed strong induction of CD38, a known RA target\(^4\), in hESC- and FL-HSPCs (Fig. 6d). Persistent CD90 expression suggested that they had not differentiated. AM580-induced CD38 expression was reversible (Fig. 6d), and AM580-treated EB CD34\(^+\) cells typically retained a higher fraction of CD34\(^+\) cells with AM580 or ATRA agonist induced several medial HOXA genes (Fig. 6c). AM580 had a more robust effect on HOXA genes and was used for further studies.

### RA signalling promotes definitive haemogenic endothelium and HSC fate

To understand how RA signalling modulates hESC-derived haematopoietic cell development, RNA sequencing was performed to define AM580-induced genes in hESC-HSPCs (CD45\(^+\)CD34\(^+\)CD90\(^+\)). Six days after AM580 treatment, RARB and RARG, known RARA targets\(^4\), were upregulated (Fig. 7a). The HOXA cluster was also induced (Fig. 7b,c), whereas RUNX1, which is expressed in both progenitors and HSCs, showed no difference. Altogether, AM580 induced 408 genes and repressed 562 (twofold, \( P < 0.05 \)) (Supplementary Table 5). AM580-induced genes showed the highest enrichment in GO categories reflecting vasculature development, cell adhesion and migration (Fig. 7d). These include factors implicated in definitive haemogenic endothelium and HSC development.
Figure 5 Overexpression of medial HOXA genes enhances proliferative potential in FL-HSPCs but does not confer HSC properties to hESC-HSPCs. (a) A schematic showing the strategy for constitutive lentiviral overexpression of HOXA5 or HOXA7 in FUGW vectors in FL-HSPCs. (b) Representative FACS plots of FUGW empty vector, HOXA5- or HOXA7-overexpressing FL-HSPCs. (c,d) Expansion of total FL cells (c) or HSPCs (d) transduced with HOXA5- or HOXA7-overexpression vectors or empty control vector (Ctrl) (mean and s.e.m. values from \( n = 3 \) independent experiments; statistical significance was assessed using the paired Student’s \( t \)-test). (e) qRT-PCR confirming overexpression in transduced HSPCs sorted 1 week post-infection (\( n = 1 \) experiment with two pooled donors). (f) CFU-Cs from 2,000 HSPCs sorted after day 10 of infection with vectors overexpressing HOXA5 or HOXA7, or FUGW empty control vector (mean and s.d. values shown from \( n = 4 \) transductions from two independent experiments; \( P \) values shown correspond to Ctrl versus OE-HOXA7). (g) A schematic showing the strategy for lentiviral overexpression of HOXA5 and/or HOXA7 and/or HOXA9 in FUGW vectors in EB CD34\(^+\) cells. (h) Representative examples of FACS plots of EB CD34\(^+\) cells overexpressing HOXA5 or HOXA7, or a combination of HOXA5, HOXA7 and HOXA9. Un-transduced FL is shown as a control. (i) Quantification of CD34\(^+\)CD38\(^-\)CD45\(^-\) haematopoietic cells from \( h \); mean from \( n = 4 \) independent experiments for Ctrl and \( n = 3 \) for HOXA5/7/9, HOXA5 and HOXA7 at days 0 and 24, and \( n = 2 \) at all other time points. (j,k) Transplantation assay of EB-CD34\(^+\) cells into NSG mice. Representative FACS plots (j) and quantification (k) of human CD45 cells in the bone marrow of NSG mice 12 weeks post-transplantation. Multilineage engraftment is assessed by CD19 and CD3 (\( B \) and \( T \) lymphoid, right column) and CD66 and CD33 (myeloid, middle column) (mean from \( n = 5 \) mice per condition (except for FL, \( n = 4 \)) from two independent experiments). (l) qRT-PCR for HOXA7 from transduced EB-OP9-HSPCs 2 weeks post-infection from one representative experiment. (m) Graphs representing RNA-Seq of EB-OP9 cells overexpressing HOXA7 for genes regulated by HOXA7 in FL-HSPCs (Fig. 4k,l) (one representative experiment, GEO database GSE76685). See Supplementary Table 7 for statistics source data in d,i,k.
Figure 6 Retinoic acid signalling activates medial HOXA genes during human haematopoietic development. (a) Microarray analysis of gene expression of components of RA signalling pathway compared with FL-HSPCs (for n values see Fig. 4b; for mean values see Supplementary Table 1 and GEO database GSE64865). (b) A schematic showing 6-day treatment of CD34+EB and FL cells by all-trans retinoic acid (ATRA) and the RARA agonist AM580. Cells were reseeded on OP9-M2 stroma after 12 days and analysed after an additional 12 ± 1 days (day 24 ± 1). (c) qRT-PCR of HOXA3, HOXA5, HOXA6, HOXA7 and HOXA9 expression in EB or FL cells treated with RA and AM580 (mean ± s.e.m. from n = 4 independent experiments; see Supplementary Table 7 for statistics source data). (d) Representative FACS plots of CD45+ cells from AM580-treated EB and FL cells at 6, 12 and 24 days on OP9-M2 culture (n = 8 independent experiments). (e,f) Quantification of CD34+CD38−/− (e) and HSPC (f) fraction of EB- and FL-derived haematopoietic cells at day 24 ± 1 of OP9-M2 culture (mean ± s.e.m. from n = 8 independent experiments). Statistical significance was assessed using the Student’s paired t-test for c,e,f, one-tailed for c and two-tailed for e,f.

Analysis of hESC-HSPCs at day 12 of OP9-M2 culture (6 days after AM580 removal) evidenced expression of HOXA, albeit at lower levels, and many vascular factors and HSC regulators (Supplementary Fig. 6A,B). Interestingly, many genes upregulated in HOXA7-knockdown FL-HSPCs and hESC-HSPCs, including cell cycle inhibitors CDKN1A and CDKN2D and erythroid and...
RA signalling

**Figure 7** Retinoic acid signalling pulse in EB-derived haematopo-vascular cells induces transcriptional programs associated with definitive haemogenic endothelium and HSC fate. (a,b) RNA sequencing analysis showing FPKM quantification for the RARA targets RARB and RARG (a) and HOXA genes (b) in sorted AM580-treated and DMSO control hESC-HSPCs (mean from two independent experiments). (c) RNA-Seq genome browser screen shot for the HOXA cluster and RUNX1 in hESC- and FL-HSPCs after 6 days of AM580 treatment. (d) GO categories of biological processes significantly upregulated in hESC-HSPCs by AM580 treatment at day 6. (e) FPKM quantification values from representative genes from vasculature development and transcription GO categories from genes significantly upregulated in hESC-HSPCs by 6-day AM580 treatment (twofold or greater change, P < 0.05, mean from two independent experiments). (f) ATAC sequencing genome browser shot for the HOXA cluster and RUNX1 assessing change in accessibility of regulatory regions in hESC- and FL-HSPCs on AM580 treatment. (g) Peaks significantly induced by AM580 treatment grouped on the basis of the distance from the TSS. (h) GO categories enriched among genes showing significant difference in accessibility after AM580 treatment. (i) ATAC-seq signal proximal to the TSS of genes up- or downregulated by AM580 treatment. (ATAC-seq data show one representative data set from two independent experiments that showed comparable results.) See Supplementary Table 5 and GEO database GSE76685 for values used to generate graphs in b,e,g,i.

megakaryocytic genes, started to decline in AM580-treated hESC-HSPCs (Supplementary Fig. 6C). These data suggest that RARA signalling during endothelial-to-haematopoietic transition induces HOXA genes and other regulators that establish the definitive HSC fate, while suppressing embryonic progenitor programs.

To investigate whether RA signalling induces genes by modulating chromatin accessibility, ATAC sequencing was performed for CD45<sup>+</sup>CD34<sup>+</sup>CD90<sup>+</sup> hESC-HSPCs (DMSO- or AM580-treated for 6 days) and FL-HSPCs. AM580 stimulation increased chromatin accessibility throughout the HOXA cluster (Fig. 7f). Unbiased analysis...
of differentially accessible peaks (q value <0.05) between DMSO- and AM580-treated hESC-HSPCs showed that most ATAC sequencing signal was 50–500 kilobases from the transcriptional start site (TSS), which are likely to represent enhancers (Fig. 7g). GREAT analysis identified vasculature-related categories as top differentially accessible gene groups (Fig. 7h). The average ATAC sequencing signal around the TSS of AM580-induced genes increased twofold on AM580 treatment (Fig. 7i), whereas genes repressed on AM580 treatment did not show a difference. These data suggest that RA signalling facilitates chromatin opening in the regulatory regions of the genes it activates.

DISCUSSION

Inability to replicate the microenvironment where HSCs develop has prevented the generation of clinically valuable HSCs from PSCs. However, the molecular barriers underlying the dysfunction of PSC-HSPCs are unknown. Here we document the generation of hESC-HSPCs that possess the immunophenotype of human fetal HSCs and differentiate into adult-type erythroid cells and T-lymphoid cells, but lack self-renewal ability. We pinpointed defective medial HOXA gene activation as a major developmental barrier preventing the establishment of self-renewing HSCs from hESCs, and identified RA signalling during endothelial-to-haematopoietic transition as a key inducer of HOXA genes and HSC fate.

Cells obtained from EBs possessed a surface phenotype of haemogenic endothelium and immature haematopoietic precursors comparable to the early placenta (CD34+CD38−/−/CD90+CD43+CD45−GPI-80−)13,14. Following culture on HSC-supportive stroma, hESC-HSPCs acquired surface expression of CD45 and GPI-8014 and a closer molecular resemblance to FL-HSPCs. However, specific developmentally regulated genes, including the HOXA genes, remained silenced in hESC-HSPCs. Low expression of HOXA genes can also be found in published transcriptome analysis of HESC-derived CD34+ cells2,43–45, implying that HOXA gene silencing is independent of the differentiation protocol or PSC line used.

HOXA genes are dysregulated in leukaemias46,47 and HOXA9 has been previously implicated in mouse fetal HSC self-renewal52,48. Knockout mice for other HOXA genes have not revealed strong HSC phenotypes, presumably owing to compensation by other HOXA genes9,50. Deletion of the entire HOXa cluster in adult mice severely reduced HSC activity and even HOXa cluster haploinsufficiency compromised HSC function51,52. HOXA9 overexpression enhanced haematopoietic specification from hESCs, but did not confer HSC function44. Our finding that HOXA5 and HOXA7 are critical for in vitro expansion and engrafment of FL-HSPCs, and their over-expression promotes FL-HSPC maintenance, documents a broader requirement for medial HOXA genes in human HSC regulation.

As HOXA7 had the strongest loss- and gain-of-function phenotype, it was chosen for molecular analysis. Many genes dysregulated in HOXA7-deficient FL-HSPCs were also dysregulated in EB-OP9-HSPCs. Similar to hESC-HSPCs, HOXA7-deficient FL-HSPCs upregulated embryonic and fetal globins and other erythroid and megakaryocyte differentiation genes, implying that they acquire embryonic progenitor-like identity. Although the cell cycle was unaffected in HOXA7-deficient FL-HSPCs and hESC-HSPCs, they upregulated cell cycle inhibitors CDKN1A (p21Cip1) and CDKN2D (p19Ink4d). These inhibitors are low in human FL-HSPCs, but become induced after prolonged culture26. They are also highly expressed in FL CD34+CD38−/−CD90+ progenitors that proliferate, but cannot self-renew54. Thus, upregulation of these CDKNs in haematopoietic cells associates with poor proliferative potential and onset of differentiation.

Although the overexpression of HOXA7, and to a lesser extent HOXA9, improved FL-HSPC maintenance in culture, their overexpression, even together with HOXA9, was insufficient to rescue hESC-HSPC function. HOXA7 overexpression in hESC-HSPCs did not rescue the genes regulated by HOXA7 in FL-HSPCs. The context-dependent regulation of HOXA target genes may depend on their epigenetic state and/or complementary regulatory factors present only in properly specified HSPCs.

We identified RA signalling as an upstream regulator for HOXA genes and the definitive HSPC transcriptional program. RA production by RALDH2 regulates HSC development during endothelial-to-haematopoietic transition in mouse AGM59. EB-derived haemato-vascular cells showed low RALDH2 expression compared with developmentally matched PL-HSPCs, suggesting that HESC-derived haemato-vascular precursors cannot produce RA. Supplementation with RARA agonist to EB-derived haemato-endothelial cells for 6 days induced the expression of medial HOXA genes, and increased chromatin accessibility in the HOXA cluster and other RA-induced genes. Although many RA-regulated vascular genes are not highly expressed in FL-HSPCs, they promote definitive haemogenic endothelium17 and HSC fate. RA stimulation also induced the HSC regulators MECOM, HLF, GFI1 and GATA3 and the HSC surface markers ROBO4, EMCN and PROCR. Strikingly, RARA stimulation partially suppressed the HOXA7-controlled differentiation programs and cell cycle inhibitors in hESC-HSPCs. These results suggest that RARA-induced medial HOXA gene activation is critical for HSC fate.

The level of HOXA expression in hESC-HSPCs attenuated after RA stimulation ended, implying that other regulators contribute to maintaining high HOXA expression in HSCs. Moreover, prolonged RA signalling can be detrimental for HSC development and maintenance53,54 and RA effects vary with experimental conditions and developmental stage54–56. Therefore, the timing and dosage of RA stimulation must be optimized. Although other genetic programs are needed to impart full self-renewal capacity in HSCs, RA signalling and medial HOXA genes emerge as key regulators governing the patterning of haemogenic endothelium to definitive HSC fate.

METHODS

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

Note: Supplementary Information is available in the online version of the paper.

ACKNOWLEDGEMENTS

We thank BSCR FACS Core at UCLA, the UCLA Clinical Pathology Microarray Core, BSCR Sequencing Core, UCLA Tissue and Pathology Core and CEAR Gene and Cell Therapy Core (NIH grant AI028697-21) and Novogenix LLC. We thank H. Coller for discussions, T. Bolan for assistance with experiments, Y. Xing and Y.-T. Tseng for consultation on RNA sequencing analysis, and T. Stoyanova, D. Johnson and O. Witte for help with NSG mice. This work was supported by CIRM RN1-00557 and RT3-00783, NIH R01 DK100959, LLS Scholar award and Rose Hills Foundation Scholar Award to E.K.A.M.; Broad Stem Cell Research Center at UCLA and JCC Foundation; NIH P01 GM081621 to I.A.Z. and Z.G.; and NIH PO1 HL073104 and CIRM R83-05217 to G.M.C. D.R.D. was supported by the NSF GRFP.
16. Pick, M., Azzola, L., Mossman, A., Stanley, E. G. & Elefanty, A. G. Differentiation of human pluripotent stem cells towards hematopoietic cells: progress and pitfalls. *Curr. Opin. Hematol.* **15**, 312–318 (2008).

21. Shojaei, F. & Menendez, P. Molecular profiling of candidate human hematopoietic stem cells derived from human embryonic stem cells. *Exp. Hematol.* **36**, 1343–1448 (2008).

22. Martin, C. H., Woll, P. S., Ni, Z., Zuniga-Pflucker, J. C. & Kaufman, D. Differences in lymphocyte developmental potential between human embryonic stem cell and umbilical cord blood-derived hematopoietic progenitor cells. *Blood** **112**, 2730–2737 (2008).

23. Qiu, C., et al. Differentiation of human embryonic stem cells into hematopoietic cells by coculture with human fetal liver cells recapitulates the globin switch that occurs during erythropoiesis. *Exp. Hematol.* **33**, 1450–1458 (2005).

25. Wang, L., Cerdan, C., Menendez, P. & Bhatia, M. Derivation and characterization of hematopoietic cells from human embryonic stem cells. *Methods Mol. Biol.* **331**, 179–200 (2006).

26. Manessison, V. et al. Expansion on stromal cells preserves the undifferentiated state of human hematopoietic stem cells despite compromised reconstitution ability. *PLoS One** **8**, e59312 (2013).

27. Kennedy, M. et al. T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures. *Cell Rep.* **2**, 1722–1735 (2012).

28. Shaut, C. A., Keene, D. R., Sorensen, L. K., Li, D. Y. & Stadler, H. S. HOXA13 is essential for placental vascular patterning and labyrinth endothelial specification. *PLoS Genet.* **4**, e1000737 (2008).

29. Aguilo, F. et al. Pdmn16 is a physiologic regulator of hematopoietic stem cells. *Blood* **117**, 5057–5066 (2011).

30. Klimmcke, D. et al. Transcriptome-wide profiling and posttranscriptional analysis of hematopoietic stem/progenitor cell differentiation toward myeloid commitment. *Stem Cell Rep.* **3**, 858–875 (2014).

31. Thorsteinsson, T. et al. Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. *Blood* **99**, 121–129 (2002).

32. Lawrence, H. J. et al. Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. *Blood** **106**, 3988–3994 (2005).

33. Wang, Y., Shulute, B. A., LaRue, A. C., Ogawa, M. & Zhou, D. Total body irradiation selectively induces murine hematopoetic stem cell senescence. *Blood** **107**, 358–366 (2006).

34. Blackwell, M. et al. p19 INK4d controls hematopoietic stem cells in a cell-autonomous manner during genotoxic stress and through the microenvironment during aging. *Stem Cell Rep.* **3**, 1085–1102 (2014).

35. Pluta, K., Luce, M. J., Bao, L., Agha-Mohammadi, S. & Reiser, J. T. Tight control of transgene expression by lentivirus vectors containing second-generation tetracycline-responsive promoters. *J. Gene Med.* **7**, 803–817 (2005).

36. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germ line transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science** **295**, 868–872 (2002).

37. Marshall, H., Morrison, A., Studer, M., Poppert, H. & Krumlauf, R. Retinoic and Hox genes. *FASEB J.* **10**, 969–978 (1996).

38. Gavalas, A. & Krumlauf, R. Retinoid signalling and hindbrain patterning. *Curr. Opin. Dev. Biol.* **10**, 380–386 (2000).

39. Chanda, B., Dittadi, A., Issac, N. N. & Keller, G. Retinoic acid signaling is essential for activating embryonic hematopoietic stem cell development. *Cell** **155**, 215–227 (2013).

40. Delecouse, C. et al. Selective high affinity retinoic acid receptor alpha or beta-gamma ligands. *Mol. Pharmacol.* **50**, 556–562 (1991).

41. Kishimoto, H. et al. Molecular mechanism of human CD38 gene expression by retinoic acid. Identification of retinoic acid response element in the first intron. *J. Biol. Chem.* **273**, 15429–15434 (1998).

42. Balmer, J. E. & Blomhoff, R. Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773–1808 (2002).

43. Salvagiotto, G. et al. Molecular profiling reveals similarities and differences between primitive subsets of hematopoietic stem cells generated in vitro from human embryonic stem cells and in vivo during embryogenesis. *Exp. Hematol.* **36**, 1377–1389 (2008).

44. Ramos-Mejia, V. et al. HOXA9 promotes hematopoietic commitment of human embryonic stem cells. *Blood** **124**, 5056–5075 (2014).

45. Lai, L. et al. Generation of hematopoietic-repopulating cells from human embryonic stem cells independent of ectopic HOX4 expression. *Exp. Med.** **201**, 1603–1614 (2015).

46. Beachy, S. H. et al. Isolated Hoxa9 overexpression predisposes to the development of lymphoid but not myeloid leukemia. *Exp. Hematol.* **41**, 518–529 (2013).

47. Alharbi, R. A., Pettenegil, R., Pandha, H. S. & Morgan, R. The role of Hox genes in normal hematopoiesis and acute leukemia. *Leukemia** **27**, 1000–1008 (2013).

48. McKinney-Freeman, S. et al. The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell** **11**, 701–714 (2012).

49. Chen, F., Greer, J. & Capecchi, M. R. Analysis of Hoxa7/8Hoxb7 mutants suggests periodicity in the generation of the different sets of vertebrae. *Mech. Dev.* **77**, 49–57 (1998).

50. Boucherat, O. et al. Partial functional redundancy between Hoxa5 and Hoxb5 paralog genes during lung morphogenesis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **304**, L817–L830 (2013).

51. Lebert-Ghali, C. E. et al. Hoxa cluster is haploinsufficient for activity of hematopoietic stem and progenitor cells. *Exp. Hematol.* **38**, 1074–1086 (2010).

52. Lebert-Ghali, C. E. et al. Hoxa cluster genes determine the proliferative activity of adult mouse hematopoietic stem and progenitor cells. *Blood** **127**, 87–90 (2016).

53. Muramoto, G. G. et al. Inhibition of aldehyde dehydrogenase expands hematopoietic stem cells with radioprotective capacity. *Stem Cells** **28**, 523–534 (2010).

54. Szatmari, I., Iacovino, M. & Kaya, M. The retinoid signaling pathway inhibits hematopoiesis and uncouples from the Hox genes during hematopoietic development. *Stem Cells** **28**, 1518–1529 (2010).

55. Rönn, R. E. et al. Retinoic acid regulates hematopoietic development from human pluripotent stem cells. *Stem Cell Rep.* **4**, 269–281 (2015).

56. Cato, E., Ariza, L., Muñoz-González, R. & Carmona, R. Signaling by retinoic acid in embryonic and adult hematopoiesis. *J. Dev. Biol.* **2**, 18–33 (2014).
METHODS

Human ESC culture. The human ESC line H1, obtained from WiCell (WA01), was maintained on 50-Gy-irradiated (CF1) murine embryonic fibroblast feeders, in DMEM/F12 medium (Invitrogen) containing 20% knockout serum replacement (Invitrogen), 0.5% penicillin/streptomycin (50 U ml⁻¹), and 1 mM l-glutamine, 0.5% MEM non-essential amino acids (NEA, Invitrogen), 10 ng ml⁻¹ basic FGF (R&D Systems) and 0.11 mM BME (Invitrogen). hESCs were passaged weekly using 1 mg ml⁻¹ collagenase IV (Invitrogen) in DMEM/F12 for 10 min, detached by gentle pipetting or cell scraper to maintain cells in small clumps, washed twice and split to new 6-well plates containing irradiated CF1 mouse embryonic fibroblasts. Cells were mycoplasma-tested and authenticated by short-term repeat analysis. hESC work was approved by the UCLA Embryonic Stem Cell Research Oversight committee. WA01 is not listed in the ICLAC or NCBI Biosample as a commonly misidentified cell line.

Differentiation of embryoid bodies. hESCs were treated with 0.5 mg ml⁻¹ Dispase (Invitrogen) in DMEM/F12 for 5 min at 37 °C. Clumps were mechanically detached and transferred to low-attachment plates (Corning) in IMDM (Invitrogen), 15% FBS (Hyclone), 1% MEM-NEA (Invitrogen) and 1% peni-cillin/streptomycin (100 U ml⁻¹). hESCs were maintained on 50-Gy-irradiated (CF1) murine embryonic fibroblast feeders, The human ESC line H1, obtained from WiCell (WA01), was cultured. OP9-D cell cultures were maintained on 50-Gy-irradiated (CF1) murine embryonic fibroblast feeders. Cells were mycoplasma-tested and authenticated by short-term repeat analysis. hESC work was approved by the UCLA Embryonic Stem Cell Research Oversight committee. WA01 is not listed in the ICLAC or NCBI Biosample as a commonly misidentified cell line.

Human tissue collection. Placenta and fetal liver were de-identified, discarded material obtained from elective termination of first- and second-trimester pregnancies following informed consent. As these tissues are discarded material with no personal identifiers, this research does not constitute research on human subjects. This protocol was reviewed by the UCLA IRB committee, who determined such studies can be performed without further IRB review. Specimen age for this study is denoted as developmental age, 2 weeks less than gestational age, and was determined by ultrasound or estimated by the date of the last menstrual period. Tissues were collected into PBS with 5% FBS (Hyclone), ciprofloxacin HCl (10 ng ml⁻¹), Sigma, amphotericin B (2.50 µg ml⁻¹), (Invitrogen) and 1% penicillin/streptomycin, transported on ice and processed the same day.

Human tissue processing. Single-cell suspensions were prepared from FL at 14–17 weeks of developmental age. Tissues were mechanically dissociated using scalpels and syringes. Mononuclear cells were enriched on a Ficol layer according to the manufacturer's protocol (GE Healthcare Biosciences AB) and strained through a 70 µm mesh. Single cell suspensions were prepared from placenta at 3–5 weeks of developmental age. Tissues were mechanically dissociated and digested in 2.5 U Dispase (Gibco), 90 mg collagenase (Worthington) and 0.075 mg DNase I (Sigma) per gram of tissue for 90 min at 37 °C with agitation. Cells were then filtered through a 70 µm cell strainer. A step-by-step protocol detailing EB generation and AM580/ATRA treatment can be found at Nature Protocol Exchange².

Selection of CD34⁺ cells by magnetic beads. Single-cell suspensions obtained from 2-week EB differentiation cultures, human fetal livers or human placentae were magnetically isolated with anti-CD34 microbeads (Miltenyi Biotech) according to the manufacturer's protocol.

OP9-M2 stromal co-culture for HSPC maturation and expansion. OP9-M2 cells were irradiated (20 Gy) and pre-plated (50,000 cells cm⁻²) onto tissue-culture-treated wells 24 h before the start of co-culture in OP9 medium (γ-MEM (Invitrogen), 20% FBS (Hyclone), P/S/G). Haematopoietic cells derived from hESCs or haematopoietic tissues were plated on a stromal layer in OP9 medium supplemented with SCF (25 ng ml⁻¹), Peprotech, Flt3L (10 ng ml⁻¹), Peprotech and IL-7 (20 ng ml⁻¹, Peprotech). Cells were lifted and reseeded on OP9-DL1 for 12–15 weeks to allow FACS analysis.

Flow cytometry and cell sorting. FACS analysis was performed using single-cell suspensions prepared as described. Cells were stained with mouse anti-human monoclonal antibodies against human CD45–PE (cl. J.33, Beckman Coulter IM2078U; diluted 1:100), CD45–APC–H7 (cl. D21, Biolegend 368512 and 368516; 1:100) and CD45–BV711 (cl. H130, Biolegend 304050; 1:100) and mouse CD45–APC–H7 (cl. 30-F11, BD557659; 1:100), CD34–APC (cl. 581, BD 555824; 1:200), CD90–FITC (cl. SE10 BD 555955; 1:100), CD38–PE–Cy7 (cl. HIT2, BD 560677; 1:100), CD19–PE (cl. 1D3 or HIB19, eBiosciences 12-0193 and 12-0199; 1:50), CD43–PE (cl. MT1, SCB 51772; 1:25) or –FITC (cl. G10, BD 555475; 1:20), GPI-80–PE (cl. 3H9, MBL DO875-7; 1:50), CD3–PE–Cy7 (cl. SK7, BD 557851; 1:50), CD4–APC–CL3 (5.35, Invitrogen MHCDD0455; 1:50), CD8–PE (cl. HIT8A, BD 556350; 1:50), CD13–APC (cl. WM15, BD 557454; 1:500), CD66b–FITC (cl. G10F5, BD 555724; 1:500) and CD33–APC–PE (cl. WM53, BD 561816; 1:100).

Dead cells were excluded with 7-aminocytotoxin D (DAAD) (BD Biosciences, used at 1:50). Cells were assayed on a BD-LSR II flow cytometer and data were analysed with FlowJo software (Tree Star). Cell sorting was performed using a BD FACSAria II.

Methylesscell colony-forming assays. Myeloid-erythroid progenitor potential was assessed on methylesscell (MethoCult G4H4435, SCT) supplemented with TPO (10 ng ml⁻¹, Peprotech), 1% penicillin/streptomycin (100 U ml⁻¹, Invitrogen) and 1% amphotericin B (2.50 µg ml⁻¹, Invitrogen). Cultures were incubated at 37 °C and 5% CO₂ for 14–16 days and colonies were scored on the basis of morphological characteristics. Images were taken using an Olympus BX51 microscope with a DP72 camera.

Cytospins. Representative myeloid-erythroid colonies from methylesscell assays were picked and reseeded in PBS with 40% FBS (Hyclone). The cell suspension was cytospun on slides using a Shandon Cytospin 4 (Thermo Electron Corporation) spun at 10g under medium acceleration. Slides were air-dried overnight and stained using May–Grünewald–Giemsa (MGG) stain (Sigma–Aldrich).

Imaging. Bright-field images of individual colonies and images of MGG-stained colonies were taken using the Zeiss Axovert 40 CFL microscope under the x10 objective with an attached Canon PC TO1989 camera at x4 zoom for a total magnification of x40.

T-lymphoid differentiation. hESC or FL-derived CD34⁺ cells were plated on non-irradiated OP9-DL1 stroma (25,000 cells cm⁻²) in OP9 medium supplemented with SCF (25 ng ml⁻¹, Peprotech), FLT3L (10 ng ml⁻¹, Peprotech) and IL-7 (20 ng ml⁻¹, Peprotech). Cells were lifted and reseeded on OP9-DL1 every week until FACS analysis.

Cell cycle analysis. Cultured cells were pulse-labelled with 10 µM BrdU for 35 min in culture. Cells were sorted for the indicated surface phenotypes and processed according to the FITC–BrdU flow kit (BD) instructions.

RNA isolation, cDNA synthesis and quantitative reverse transcriptase PCR. RNA isolation was performed using the RNeasy Mini kit (Qiagen) with an additional DNase (Qiagen) step using the manufacturer's protocol. cDNAs were prepared using the Quantitect reverse transcription kit (Qiagen) and quantitative polymerase chain reaction (qPCR) for GAPDH, glycophorin A (GYPHA), haemoglobin subunit epsilon (HBE), haemoglobin subunit gamma (HBB) and beta globin (HHB) was performed with the LightCycler 480 SYBR Green I Master Mix (Roche) on the Lightcycler 480 (Roche); qPCR for GAPD and HOXA7 with the SYBR Select Master Mix (Life Technology, LT) using the ViiA 7 Real-Time PCR System (LT); SYBR Green–compatible primers were obtained from the PrimerBank database or literature²⁴,²⁵. Taqman primers were purchased from LT. Primers were tested against OP9-M2 cDNA to rule out amplification of murine genes, and gDNA or water as negative controls. Primers are presented in Supplementary Table 6.

Microarray analysis. RNA isolation was performed using RNeasy Mini kit (<50,000 cells) or RNeasy Micro kit (<50,000 cells) (Qiagen) with DNase digestion using the manufacturer's protocol. RNA was amplified by the NuGen amplification kit and hybridized on Affymetrix arrays (Human U133plus2.0 Array). Samples were quantile normalized.

K-means clustering of differentially expressed genes. Pairwise differential expression analysis was performed for the various populations by independently comparing immunophenotypic HSPCs derived from hESCs (EB and EB–OP9), placenta (PL), cultured fetal liver (FL–OP9 2 weeks and FL–OP9 5 weeks), with freshly isolated FL–HSPCs (FL). The following criteria were used to filter probes and identify differentially expressed genes: fold change value of 2 or higher; a P value less than 0.05; probes that are called 'Absent' in all replicates in all samples were excluded; and probes that have an absolute expression level of 50 or less were excluded. The union of all probes that met these criteria was chosen for further analysis, after which normalized expression values, which were replicate-averaged and standardized, were obtained. The standardized expression values were clustered using the K-means method. GO enrichment analysis was performed for genes in the various clusters. Differential expression assessment was performed using the
Induction of retinoic acid signalling. ATRA (all-\textit{trans} retinoic acid, Sigma-Aldrich) was dissolved in DMSO at 25 mM and applied at a final concentration of 1 \mu M; AM580 (Sigma-Aldrich) was dissolved in DMSO at 10 \mu M at a final concentration of 0.2 \mu M; DMSO was used at 1:2,500 final dilution. Treatments were performed on FL- or hESC-derived CD34+ cells in HSC medium. Cells were treated at day 6 for qRT–PCR and RNA sequencing analysis. For longer cultures, RA stimulation media were removed at day 6 from the wells avoiding disruption of the cell layer and replaced with HSC medium with no treatment. Cells were then collected at day 12 reseeded on OP9-M2 for another 12–13 days, and then assayed for HSPC expansion by FACS or colony-forming potential in methylcellulose. A step-by-step protocol detailing EB generation and AM580/ATRA treatment can be found at Nature Protocol Exchange\textsuperscript{76}.

ATAC sequencing. Cells (12,000 to 60,000) were sorted in PBS and processed according to the protocol indicated\textsuperscript{77}, with minor adjustments. Nuclei were purified by the addition of 250 \mu l of lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl\textsubscript{2}, 0.1% IGEPAL), pelleted and resuspended in the transposition reaction mix (Nextera DNA Library Prep Kit, Illumina) and incubated at 37 °C for 30 min. Transposed DNA was sequenced and used for library amplification with custom-made adaptor primers (see Supplementary Table 7) using NEBNext High-Fidelity 2x PCR Master Mix (NEB). The amplification was interrupted after five cycles and a SyBR green qPCR was performed with 1/10 of the sample to estimate for each sample the additional number of cycles to perform, before saturation was achieved. Total amplification was between 10 and 15 cycles. Purified libraries were sequenced using the HiSeq-2000 (Illumina) to obtain paired-end 50 bp-long reads. Demultiplexing of the reads and creation of the fastq files was performed using an in-house Unix shell script. Read mapping to the genome (hg19) was performed using Bowtie2 or v2.2.5\textsuperscript{87} with parameters --local -X 2000 -n 1 --no-mixed. The Bamcovergae tool from DeepTools was used to create the coverage .bw files for visualization\textsuperscript{88}. Samtools v1.2 was used to remove duplicates and reads aligned to chrM. MACS\textsuperscript{27} was used to call the differentially accessible peaks between the AM580-treated and -untreated EB-derived cells, using the parameters –broad –broad-cutoff 0.1.

Statistics and reproducibility. Graphs were generated with GraphPad PRISM software. Statistical significance was calculated in R version 3.2.3. Statistical significance was assessed using the Wilcoxon rank sum test for unpaired data sets and the Student's paired t-test for paired data sets. All tests are two-tailed with the exception of that in Fig. 6c, where a one-tailed distribution was used because it is an established fact that RA treatment enhances HOX gene expression. Data sets were considered paired when treatments (that is, DMSO versus AM580 or LKO versus shRNA) were performed in parallel on the same batch of cells (EB) or from the same donor (FL) for the same time periods. Data were considered unpaired when cells from different batches, donors or cell types were compared. The null hypothesis of the means being equal was rejected at \( \alpha = 0.05 \) and significant \( P \) values are shown in each graph. The comparisons to be made were decided a priori with the intention of limiting the overall number of comparisons and are indicated by lines in the figures and adjustments for multiple comparisons would be applied to \( k \leq 4 \) comparisons. To account for multiple comparisons in microarray analysis, \( P \) values were adjusted using the Benjamini–Hochberg correction method to control the false discovery rate (FDR). A 5% FDR was used as the cuto.

Accession numbers deposited in GEO database. Primary accession: GSE76685. Referenced accessions: GSE54316\textsuperscript{11} and GSE49074\textsuperscript{40}.

57. Dou, D. R., Calvanese, V., Saarikoski, P., Galic, Z. & Mikkola, H. K. A. Induction of HOXA genes in hESC-derived HSPC by two-step differentiation and RA signalling pathway. Protocol Exchange (2016). DOI: 10.1038/protex.2016.035 (2016).
58. Thoma, S. J., Lamping, C. P. & Ziegler, B. L. Phenotype analysis of hematopoietic CD34+ cell populations derived from human umbilical cord blood using flow cytometry and cDNA-polymerase chain reaction. Blood 83, 2103–2114 (1994).
59. Bauchwitz, R. & Costantini, F. Developmentally distinct effects on human epsilon-, gamma- and delta-globin levels caused by the absence of alpha-globin regulation of the human beta-globin gene in YAC transgenic mice. Hum. Mol. Genet. 9, 561–574 (2000).
60. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nat. Methods 43, e15 (2015).
61. Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 9, R80 (2004).

METHODS

DOI: 10.1038/ncb3354

© 2016 Macmillan Publishers Limited. All rights reserved
62. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. 
Bioinformatics 20, 1453–1454 (2004).
63. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. 
Bioinformatics 20, 3246–3248 (2004).
64. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with 
RNA-Seq. Bioinformatics 25, 1105–1111 (2009).
65. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing 
genomic features. Bioinformatics 26, 841–842 (2010).
66. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals 
unannotated transcripts and isoform switching during cell differentiation. Nat. 
Biotechnol. 28, 511–515 (2010).
67. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. 
Transposition of native chromatin for fast and sensitive epigenomic profiling of 
open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 
1213–1218 (2013).
68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. 
Methods 9, 357–359 (2012).
69. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a 
flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, 
W187–W191 (2014).
70. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, 
R137 (2008).
Supplementary Figure 1 FL derived, but not hESC derived haematopoietic cells can reconstitute human HSPC compartment in recipient BM. (A) Schematic of transplantation of CD34+ cells into irradiated NSG mice. (B) NSG mice were transplanted with CD34+ cells from hESCs (EB and EB-OP9) and fetal liver (FL or FL-OP9) and human engraftment in the BM assessed at 12 weeks for CD45+CD34+CD38-CD90+ immunophenotypic HSPCs. Results shown are from representative animals for each group of transplanted mice (5 mice transplanted with EB cells, 4 with EB-OP9 cells, 4 with FL cells, and 3 with FL-OP9 cells).
Supplementary Figure 2 hESC-derived haematopoietic cells can upregulate adult haemoglobin-beta (HBB) and differentiate into T-lymphoid cells (A) Representative FACS plots and quantification of BrdU incorporation and 7-AAD to determine cell cycle distribution in EB and FL CD90+ immunophenotypic HSPCs and CD90- cells is shown (mean +/- SEM from n=3 independent experiments). (B) Comparison between CD34+ haematopoietic cells and immunophenotypic (CD34+CD38-CD45+CD90+) HSPCs, all seeded at an initial density of 10,000 cells per sample, from FL and hESC-derived cells (mean +/- SEM of n=5 independent experiments). (C) CFU-C expansions from 10,000 hESC-derived or FL-derived CD34+ cells in methylcellulose following 0, 1, 2 and 3 additional weeks on OP9-M2 co-culture (mean +/- SEM of n=3 independent experiments). (D) Haemoglobin levels (expression measured from colonies derived from CD34+ cells) of embryonic epsilon (HBE), fetal gamma (HBG), and adult beta (HBB) measured through qRT-PCR and normalized to Glycophorin A levels (mean +/- SEM shown from n=3 independent experiments). (E) FACS staining of hESC- and FL-derived CD34+ haematopoietic cells grown on OP9-DL1 stroma for 4 weeks is shown. Cells were stained for CD45, the myeloid exclusion marker CD14, and T-cell markers CD4 and CD8 (mean +/- SEM from n=3 independent experiments). Statistics source data for graphs shown in A, B, C, and E can be found in Supplementary Table 7. Statistical significance was assessed using the Wilcoxon Rank Sum test for A, B, C, and E.
Supplementary Figure 3 Knockdown of HOXA5 or HOXA7 does not lead to changes in BRDU incorporation in FL immunophenotypic HSPCs. (A) Representative FACS plots and quantification of cell cycle analyses based on BrdU incorporation (mean from one experiment with 2 independent donors, statistics source data can be found in Supplementary Table 7) of control vector and HOXA5 and HOXA7 shRNA vector transduced FL-HSPCs. (B, C) Examples of cell cycle activators (B) and inhibitors (C) from RNA-seq analyses of FL immunophenotypic HSPCs with HOXA7 knockdown compared to empty vector controls (showing mean from 4 independent experiments, values used to generate graphs can be found in Supplementary Table 4 and GEO database GSE76685).
Supplementary Figure 4 Lentiviral overexpression of HOXA5, HOXA7 and HOXA9 in EB-derived CD34+ cells is not sufficient for rescuing HSC function. (A) Schematic showing the strategy for tet-inducible overexpression of HOXA5 or HOXA7 in FL-HSPCs using a PNL vector. (B) q-RT-PCR showing induction of HOXA5 or HOXA7 expression in FL-HSPCs overexpressing HOXA5 or HOXA7, compared to empty vector control 1 week post-transduction (plotting one representative experiment). (C, D) Representative FACS plots (C) and quantification (D) of FL-HSPCs overexpressing HOXA5 or HOXA7 (mean from 3 independent experiments, except for 2 independent experiments for 7-8 weeks timepoint). (E) Representative FACS plots assessing concurrent overexpression of HOXA5, HOXA7 and HOXA9 using PNL vector. EB and FL CD34+ cells transduced with empty-vector were used as controls (mean from 2 independent experiments, except for EB-control and HOXA5/7/9 at day 14, 1 independent experiment). Statistics source data for values used to generate graphs shown in B, D, and E can be found in Supplementary Table 7.
Supplementary Figure 5 AM580 treatment prolongs CFU-C potential in hESC-derived cells. (A) Quantification of CFU-Cs generated from 10,000 EB- or FL-derived haematopoietic cells at day 24 ± 1 of OP9-M2 culture (mean +/- SEM from n=4 independent experiments, statistics source data can be found in Supplementary Figure 5B). (B) Table showing CFU counts for the indicated samples. Counts were rounded to the closest integer value (DM=DMSO, AM=AM580).
Supplementary Figure 6 Analysis of gene expression changes in hESC-HSPCs upon AM580 treatment shows partial conversion to definitive HSC transcriptome. (A) RNA-seq genome browser screenshot of the HOXA cluster of day 12 EB and FL derived immunophenotypic HSPCS that were treated with AM580 for 6 days (6 days of treatment and 6 additional days in culture). (B) Representative genes upregulated by AM580 treatment at day 6, shown at day 12 as compared to FL-HSPCs (C) Representative genes upregulated by HOXA7 shRNA knockdown (see Figure 4L) shown in day 12 EB derived cells treated with AM580 (6 days of treatment and 6 additional days in culture) as compared to FL-HSPCs (showing mean from 2 independent experiments, values used to generate graphs in B and C can be found in Supplementary Table 5 and GEO database GSE76685).
Supplementary Table Legends

Supplementary Table 1 Average probe values for all genes in HSPCs isolated at different stages of human developmental hematopoiesis in vivo and in vitro. Average probe values for all genes are shown for each of the analyzed samples on Affymetrics Human U133plus Microarray. All populations were sorted for HSPC phenotype (CD34^+CD38^-CD90^+CD43^+CD45^-/+) (FL = second trimester fetal liver; EB = 2 week embryoid bodies; OP9 = sorted cells after culture on OP9-M2 stroma for the indicated number of weeks (2 or 5); and PL = first trimester placenta).

Supplementary Table 2 Gene Ontology categories of genes differentially expressed between immunophenotypic HSPCs at different stages of human developmental haematopoiesis in vitro and in vivo. Genes differentially expressed (>2 fold, p-value < 0.05) between any of the samples are included in the clustering (see Figure 3C).

Supplementary Table 3 Differentially expressed gene list from microarray analysis with raw expression values and K means cluster organization. Genes differentially expressed (>2 fold, p-value < 0.05) between any of the samples are included in the clustering (see Figure 3C).

Supplementary Table 4 RNA sequencing analysis of HOXA7 shRNA in FL HSPCs showing differentially expressed gene list. Genes that were significantly up- or down-regulated (>1.8 fold, p-value < 0.05; 4 independent fetal liver tissues) between HOXA7 shRNA and control HSPCs are shown.

Supplementary Table 5 RNA sequencing analysis of hESC-HSPCs after AM580 treatment showing differentially expressed genes. Genes up- or down-regulated (>2 fold, p-value < 0.05) after 6 days of AM580 treatment are listed. For comparison, the average FKPM values for the same treatment after 6 additional days in culture and uncultured FL are shown.

Supplementary Table 6 Primers used for qPCR Quantification and Cloning. (A) SYBR Green Primers used in qPCR quantification studies. (B) Taqman® Primers used in qPCR quantification studies. (C) Primers used in cloning.

Supplementary Table 7 Statistics source data file. In experiments where n < 5, individual values for each replicate and independent experiment used to generate graphs in the figures are shown. Data corresponding to each Figure and Supplementary Figure are shown in separate spreadsheets.