LETTERS
https://doi.org/10.1038/s41556-022-00898-9
nature cell biology

Check for updates

Identification of a retinoic acid-dependent haemogenic endothelial progenitor from human pluripotent stem cells

Stephanie A. Luff[1,2,3], J. Philip Creamer[3], Sara Valsoni[4], Carissa Dege[3], Rebecca Scarfò[4], Analisa Dacunto[1,2], Sara Cascione[4], Lauren N. Randolph[4], Eleonora Cavalca[4], Ivan Merelli[4,5], Samantha A. Morris[6,7], Andrea Ditadi[4] ✉ and Christopher M. Sturgeon[1,2,3] ✉

The generation of haematopoietic stem cells (HSCs) from human pluripotent stem cells (hPSCs) is a major goal for regenerative medicine. During embryonic development, HSCs derive from haemogenic endothelium (HE) in a NOTCH- and retinoic acid (RA)-dependent manner. Although a WNT-dependent (WNTd) patterning of nascent hPSC mesoderm specifies clonally multipotent intra-embryonic-like *HOXA*$^+$ definitive HE, this HE is functionally unresponsive to RA. Here we show that WNTd mesoderm, before HE specification, is actually composed of two distinct KDR$^+$ CD34^{neg} populations. CXCR4^{neg}*CYP26A1*$^+$ mesoderm gives rise to *HOXA*$^+$ multilineage definitive HE in an RA-independent manner, whereas CXCR4$^+$ *ALDH1A2*$^+$ mesoderm gives rise to *HOXA*$^+$ multilineage definitive HE in a stage-specific, RA-dependent manner. Furthermore, both RA-independent (RAi) and RA-dependent (RAd) HE harbour transcriptional similarity to distinct populations found in the early human embryo, including HSC-competent HE. This revised model of human haematopoietic development provides essential resolution to the regulation and origins of the multiple waves of haematopoiesis. These insights provide the basis for the generation of specific haematopoietic populations, including the de novo specification of HSCs.

Haematopoietic development during embryogenesis comprises multiple spatio-temporally orchestrated haematopoietic programs, each regulated by bone morphogenetic protein (BMP), WNT, NOTCH and retinoic acid (RA) signalling[1]. Of these, the specific role of RA in the specification of human pluripotent stem cell (hPSC)-derived haemogenic endothelium (HE) has not been extensively characterized. We thus sought to define the dependence of RA on known hPSC-derived HE populations[2]. Briefly, through a WNT-independent (WNTi) process, hPSCs differentiate rapidly towards either a CD43$^+$ population of primitive haematopoietic progenitor cells (HPCs) or a CD34$^+$ HE[2–4], consistent with extra-embryonic haematopoiesis. Conversely, in a WNT-dependent (WNTd) process, hPSCs give rise to *HOXA*$^+$ HE with definitive erythroid–myeloid–lymphoid potential, indicative of intra-embryonic-like definitive haematopoiesis (Extended Data Fig. 1a–c)[2,3,5–7].

RA is essential for definitive haematopoietic development during embryogenesis[1,8,9]. We thus sought to identify an RA-dependent hPSC-derived haematopoietic program. Both WNTi and WNTd hPSC-derived populations are obtained in an RA-independent manner, as these are chemically defined conditions with no exogenous sources of RA[2,3,5]. Similarly, manipulation of RA signalling on hPSC differentiation cultures have failed to yield functional improvements[10–12]. As precise mesodermal patterning is critical for specifying ontogenically distinct haematopoietic programs[2], we performed single-cell RNA sequencing (scRNA-seq) on day 3 of differentiation cultures to better understand the mesodermal population(s) obtained during early differentiation. Unsupervised clustering (Extended Data Fig. 2a–c) revealed 19 transcriptionally distinct clusters (Fig. 1a,b). *CDX1/2/4*, which correlates with the development of intra-embryonic-like HE[6,13], was uniquely expressed across multiple clusters. Conversely, cluster 6 was highly enriched for WNTi cells that expressed *GYPA*/*GYPB*, which is a unique subset of mesoderm that harbours extra-embryonic-like haemogenic potential[2] (Fig. 1b,c and Supplementary Table 1). This indicated that it is possible to identify functionally distinct haemogenic mesodermal populations. We next searched for cells expressing *ALDH1A2*, which governs the enzymatic conversion of retinol (ROH) to all-*trans* retinoic acid (ATRA) during embryogenesis[14], and is essential for intra-embryonic HE development[1]. *KDR*$^+$ *GYPA*$^+$ WNTi cells had virtually no *ALDH1A2* expression, but instead robustly expressed *CYP26A1*, an enzyme that degrades RA (Extended Data Fig. 2d). By contrast, there was a substantial population of *ALDH1A2*$^+$ cells within WNTd *KDR*$^+$ cells (Extended Data Fig. 2d).

As ~40% of the WNTd cells harboured some degree of *KDR* expression (Fig. 1c), we first sought to discriminate mesodermal cells from non-mesodermal cell types. Independent clustering of the WNTd population revealed *KDR* expression within most clusters (Extended Data Fig. 2e). However, there was separation of germ-layer-like populations within many of these clusters (Extended Data Fig. 2f(i),(ii)), indicating that there is significant heterogeneity within these cultures. The population identified as mesoderm exclusively harboured *KDR*-expressing cells, although only 58% of

[1]Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai School of Medicine, New York, NY, USA. [2]Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Medicine, Division of Hematology, Washington University School of Medicine, St Louis, MO, USA. [4]San Raffaele Telethon Institute for Gene Therapy, IRCCS San Raffaele Scientific Institute, Milan, Italy. [5]Institute for Biomedical Technologies, National Research Council, Milan, Italy. [6]Department of Developmental Biology, Washington University in Saint Louis, St Louis, MO, USA. [7]Department of Genetics, Washington University in Saint Louis, St Louis, MO, USA. ✉e-mail: ditadi.andrea@hsr.it; christopher.sturgeon@mssm.edu

them had detectable *KDR* transcripts (Extended Data Fig. 2f(iii)). As we have previously demonstrated that all WNTd haematopoiesis is derived from a KDR$^+$ population[2], we conservatively focused on the cells with detectable *KDR* expression, which consisted of nine transcriptionally distinct clusters (Fig. 1d(i)). Although multiple clusters harboured *ALDH1A2*$^+$ cells, no single cluster was uniquely enriched in *ALDH1A2* expression (Fig. 1d(i),(ii) and Supplementary Table 2a,b). However, when we performed differential gene expression analysis between all of the *KDR*$^+$ *ALDH1A2*$^+$ versus *KDR*$^+$ *ALDH1A2*^{neg} populations, we identified *CXCR4* as a candidate cell surface marker (Supplementary Table 2c).

Flow cytometry confirmed that, across the hESC and iPSC lines, CXCR4 was expressed within day 3 of differentiation of KDR$^+$ cells, and its expression is significantly upregulated under WNTd conditions in comparison to WNTi (Fig. 1e). We next performed whole-transcriptome analyses on fluorescence activated cell sorting (FACS)-isolated WNTd KDR$^+$ CXCR4$^{+/neg}$ populations, WNTi KDR$^+$ CD235a$^+$ cells[13] and WNTi and WNTd CD34$^+$CD43^{neg}CD73^{neg}CD184^{neg} HE cells[4]. In comparison to later-stage HE, each KDR$^+$ population had higher expression of mesodermal genes, but substantially lower expression of canonical haemato-endothelial markers (Extended Data Fig. 3a and Supplementary Table 3a). This was further reflected in flow cytometric analyses, which confirmed that WNTd KDR$^+$ cells did not express CD34, CD144/VE-Cadherin (*CDH5*) or TIE2 (*TEK*) on their surface (Extended Data Fig. 3b), establishing these cells as an early-stage mesodermal cell state that precedes haemato-endothelial specification.

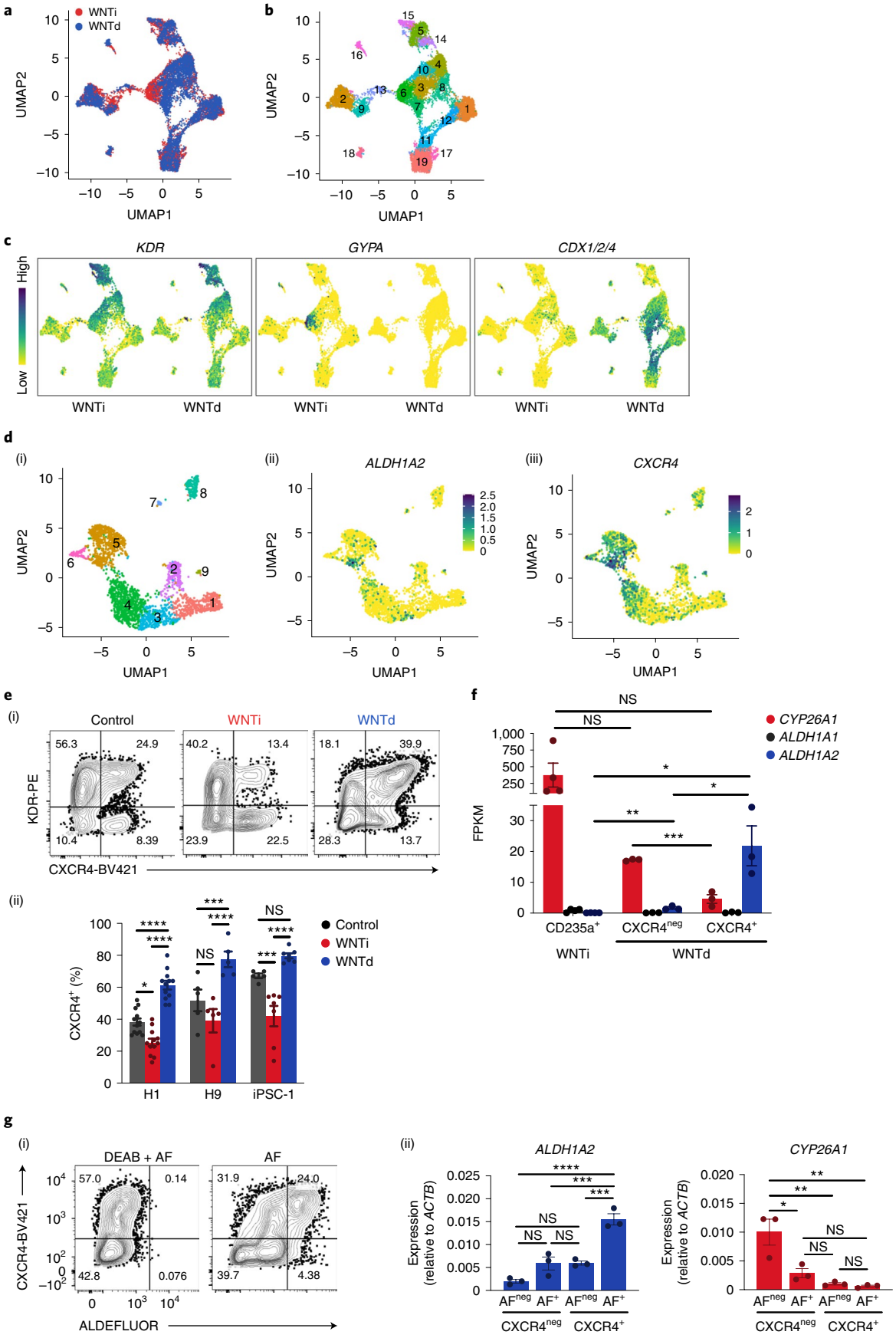
In comparison to WNTi CD235a$^+$ mesoderm, both WNTd KDR$^+$ populations exhibited substantially higher expression of primitive streak markers such as *T* (*TBXT*), *CDX1/2/4* and members of the *HOXA* cluster[6,13,15] (Extended Data Fig. 3a,c and Supplementary Table 3a), indicating that these populations comprise gastrulation and/or early mesodermal cell states. Hierarchical clustering of KDR$^+$ and HE populations revealed that the WNTd KDR$^+$ populations were more similar to each other than WNTi KDR$^+$ CD235a$^+$ cells (Extended Data Fig. 3a). Despite these similarities, each WNTd mesodermal population was transcriptionally distinct from the others (Extended Data Fig. 3d,e and Supplementary Table 3b). Critically, these populations harboured differences in the gene expression of two key enzymes regulating RA signalling (Fig. 1f). *CYP26A1* was highly expressed within both WNTi CD235a$^+$ and WNTd CXCR4^{neg} mesoderm, whereas *ALDH1A2* was highly expressed in CXCR4$^+$ mesoderm (Fig. 1f and Supplementary Table 3a,b). ALDEFLUOR ('AF') analysis confirmed that overall ALDH activity, which includes that of ALDH1A2, is enriched within this CXCR4$^+$ mesoderm (Fig. 1g(i)). Consistent with this, *ALDH1A2* expression was significantly enriched within KDR$^+$ CXCR4$^+$ AF$^+$ cells, whereas *CYP26A1* was enriched within KDR$^+$ CXCR4^{neg}AF^{neg} cells (Fig. 1g(ii)). Collectively, these observations reveal that at least two mesodermal populations exist following WNTd differentiation conditions, with a subset of CXCR4$^+$ cells being enriched in *ALDH1A2*, and are poised to respond to RA.

To assess which WNTd KDR$^+$ subset(s) could give rise to *HOXA*$^+$ definitive HE[6], we isolated each KDR$^+$ CXCR4$^{+/neg}$ population by FACS, and then cultured them for an additional five days to allow for HE specification (Fig. 2a(i))[2,5]. Both populations gave rise to a CD34$^+$CD43^{neg} population (Fig. 2a(ii)). However, only the CD34$^+$ cells derived from CXCR4^{neg} mesoderm harboured HE, as they formed a monolayer of endothelial cells that gave rise to non-adherent haematopoietic progenitors (Extended Data Fig. 4ai). This HE harboured multilineage definitive haematopoietic potential, as evidenced by its ability to give rise to definitive erythro-myeloid and T-lymphoid cells ('P1'; Fig. 2a(iii),b). By contrast, CD34$^+$ cells derived from the KDR$^+$ CXCR4$^+$ population ('P2') formed an endothelial monolayer (Extended Data Fig. 4a(i)), but did not efficiently give rise to multilineage lympho-myeloid potential (Fig. 2a(iii),b). This strongly suggested that definitive haematopoietic specification from hPSCs[2,5,6] originates from a KDR$^+$ CXCR4^{neg}CD34^{neg}*CDX4*$^+$ mesodermal population. Furthermore, consistent with its *CYP26A1* expression (Fig. 1f,g), HE can be specified from CXCR4^{neg} mesoderm in the presence of the pan-ALDH inhibitor N,N-diethylaminobenzaldehyde (DEAB, Extended Data Fig. 4b), indicating that this is an RA-independent (RAi) haematopoietic program.

Given that the non-haemogenic CXCR4$^+$ population was enriched in *ALDH1A2* expression, we hypothesized that this population may instead require RA signalling for HE specification. We therefore cultured the freshly isolated KDR$^+$ populations with ROH (Fig. 2a(i)). Critically, this treatment resulted in the specification of CD34$^+$ HE that harboured definitive erythroid, myeloid and lymphoid haematopoietic potential (P2; Fig. 2a(ii),(iii),b and Extended Data Fig. 4a(ii)). Interestingly, this RA-mediated response was temporally restricted, as only treatment of freshly isolated CXCR4$^+$ mesoderm on day 3 of differentiation, but not thereafter, resulted in the robust specification of HE (Fig. 2c). Therefore, a KDR$^+$ CD34^{neg}CXCR4$^+$ mesodermal population harbours stage-specific, RA-dependent (RAd) definitive haematopoietic potential.

ATRA has been identified as being required for definitive haematopoiesis, but repressive of extra-embryonic haematopoiesis[8,15].

Fig. 1 | scRNA-seq reveals distinct mesoderm populations. a,**b**, UMAP plots of sample origin (**a**) or transcriptionally distinct clusters (**b**) within WNTi or WNTd day 3 of differentiation cultures. **c**, Expression of *KDR*, *GYPA* or *CDX* genes within each differentiation culture. Colour bar: relative expression scaled for *KDR* and *GYPA*, and calculated module score for expression of *CDX1/2/4*. **d**, Clusters of WNTd *KDR*$^+$ cells. UMAP plots visualizing (i) clustering of *KDR*$^+$ WNTd cells, (ii) *ALDH1A2* and (iii) *CXCR4* expression. Colour bar: scaled expression within *KDR*$^+$ mesodermal cells. **e**, Mesodermal CXCR4 expression under WNTd or WNTi differentiation conditions. (i) Representative flow cytometric analysis of KDR and CXCR4 expression on day 3 of differentiation, following control, WNTi or WNTd differentiation conditions. (ii) Quantification of CXCR4$^+$ cells within each KDR$^+$ fraction, on day 3 of differentiation, across various hPSC lines. Two-way analysis of variance (ANOVA) with Tukey's test comparing all biological replicates: H1 ($n=12$; control versus WNTi, $P=0.0112$; control versus WNTd, WNTi versus WNTd, $P<0.0001$), H9 ($n=5$; control versus WNTi, $P=0.1388$; control versus WNTd, $P=0.0008$; WNTi versus WNTd, $P<0.0001$) and iPSC-1 (control $n=5$, WNTi/WNTd $n=7$; control versus WNTi, $P=0.0003$; control versus WNTd, $P=0.1109$; WNTi versus WNTd, $P<0.0001$). **f**, Expression of *CYP26A1*, *ALDH1A1* and *ALDH1A2* in day 3 KDR$^+$ cells, as in **e**. SEM, two-way ANOVA with Tukey's test comparing all biological replicates ($n=3$), CD235a$^+$ versus CXCR4$^+$ (*CYP26A1*, $P=0.152225$; *ALDH1A1*, $P=0.068067$; *ALDH1A2*, $P=0.003529$), CD235a$^+$ versus CXCR4^{neg} (*CYP26A1*, $P=0.140911$; *ALDH1A1*, $P=0.103219$; *ALDH1A2*, $P=0.010088$), CXCR4$^+$ versus CXCR4^{neg} (*CYP26A1*, $P=0.000833$; *ALDH1A1*, $P=0.429912$; *ALDH1A2*, $P=0.035653$). **g**, WNTd KDR$^+$ cells with ALDEFLUOR (AF) activity have enriched expression of *ALDH1A2*. (i) Representative AF flow cytometric analysis. (ii) *ALDH1A2* and *CYP26A1* expression within CXCR4$^{+/neg}$ALDF$^{+/neg}$ KDR$^+$ cells. One-way ANOVA with Tukey's test comparing all biological replicates ($n=3$). SEM, NS, not significant, *ALDH1A2* (CXCR4$^-$ALDF$^-$ versus CXCR4$^-$ALDF$^+$, $P=0.0904$; CXCR4$^-$ALDF$^-$ versus CXCR4$^+$ALDF$^-$, $P=0.0818$; CXCR4$^-$ALDF$^-$ versus CXCR4$^+$ALDF$^+$, $P<0.0001$; CXCR4$^-$ALDF$^+$ versus CXCR4$^+$ALDF$^-$, $P=0.9999$; CXCR4$^-$ALDF$^+$ versus CXCR4$^+$ALDF$^+$, $P=0.0006$; CXCR4$^+$ALDF$^-$ versus CXCR4$^+$ALDF$^+$, $P=0.0006$), *CYP26A1* (CXCR4$^-$ALDF$^-$ versus CXCR4$^-$ALDF$^+$, $P=0.0128$; CXCR4$^-$ALDF$^-$ versus CXCR4$^+$ALDF$^-$, $P=0.0033$; CXCR4$^-$ALDF$^-$ versus CXCR4$^+$ALDF$^+$, $P=0.0025$; CXCR4$^-$ALDF$^+$ versus CXCR4$^+$ALDF$^-$, $P=0.7111$; CXCR4$^-$ALDF$^+$ versus CXCR4$^+$ALDF$^+$, $P=0.5751$; CXCR4$^+$ALDF$^-$ versus CXCR4$^+$ALDF$^+$, $P=0.9945$).


a
WNTi
WNTd
UMAP2
UMAP1
b
c
KDR
GYPA
CDX1/2/4
High
Low
WNTi
WNTd
d
(i)
(ii)
ALDH1A2
(iii)
CXCR4
e
(i)
Control
WNTi
WNTd
KDR-PE
CXCR4-BV421
56.3
24.9
10.4
8.39
40.2
13.4
23.9
22.5
18.1
39.9
28.3
13.7
(ii)
CXCR4⁺ (%)
H1
H9
iPSC-1
Control
WNTi
WNTd
f
CYP26A1
ALDH1A1
ALDH1A2
FPKM
CD235a⁺
CXCR4neg
CXCR4⁺
WNTi
WNTd
NS
g
(i)
DEAB + AF
AF
CXCR4-BV421
ALDEFLUOR
57.0
0.14
42.8
0.076
31.9
24.0
39.7
4.38
(ii)
ALDH1A2
CYP26A1
Expression (relative to ACTB)
AFneg
AF⁺
CXCR4neg
CXCR4⁺

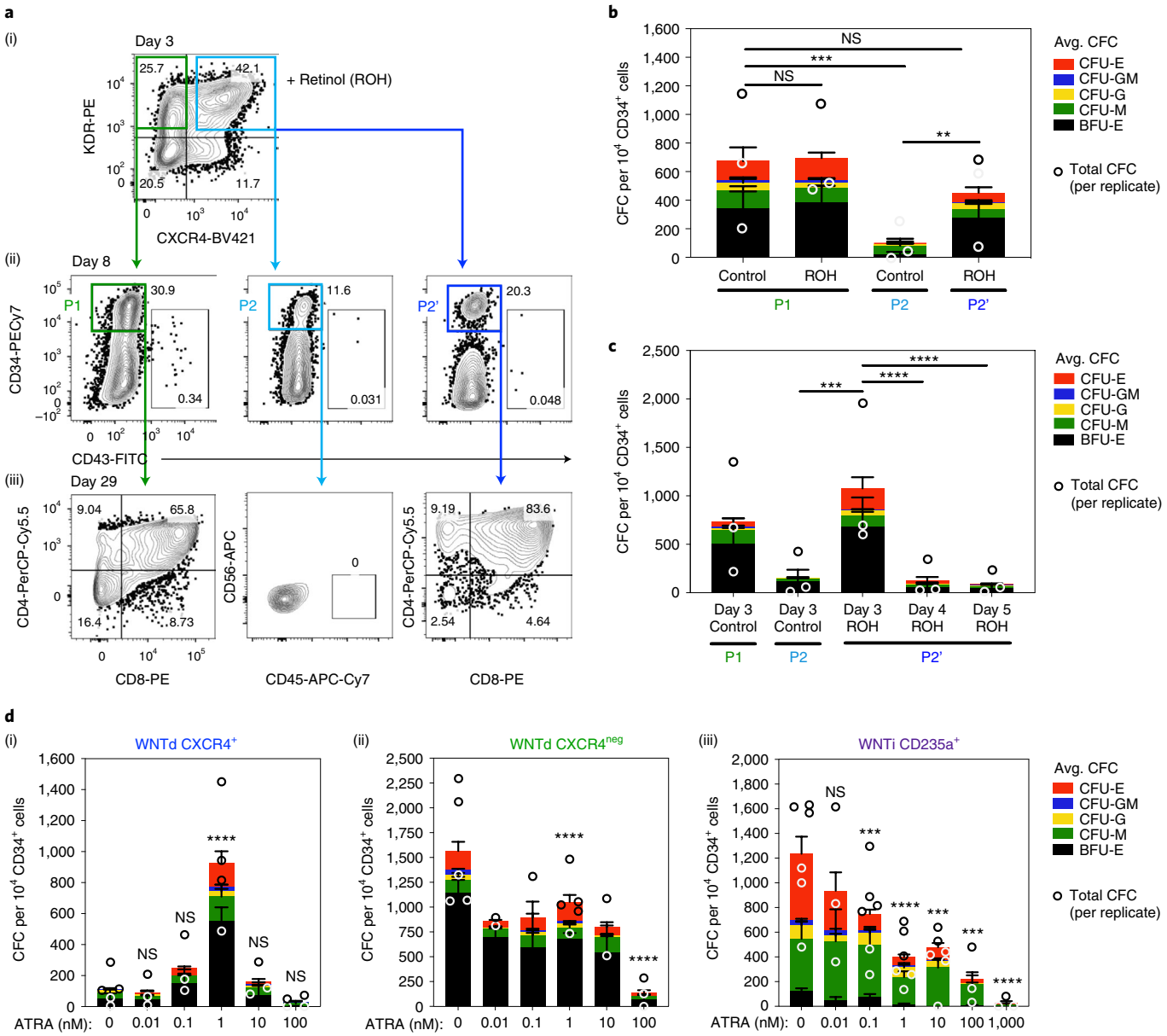


Fig. 2 | CXCR4^{neg} and CXCR4^{+} mesoderm gives rise to haemogenic endothelium in RA-independent and RA-dependent manners, respectively. **a**, Separation of mesodermal progenitors of HE, based on CXCR4 cell surface expression. (i) Representative FACS gating scheme of KDR^{+} mesoderm for CXCR4 expression, within day 3 WNTd differentiation cultures. (ii) Representative FACS gating scheme of CD34 and CD43 expression, following five days of culture after mesoderm isolation. (iii) Representative flow cytometric analyses of CD4^{+}CD8^{+} T-lymphoid potential of CD34^{+}CD43^{neg} populations. **b**, Quantification of erythro-myeloid CFC potential from different HE populations, as in **a**(ii). Two-way ANOVA with Tukey's test was used for all biological replicates ($n=3$), SEM; statistics are shown for BFU-E (CXCR4^{neg} control versus CXCR4^{+} control, $P=0.0008$; CXCR4^{+} control versus CXCR4^{+} ROH, $P=0.009$; NS, not significant); the remaining statistics are included in Source Data Fig. 2. Bars show the mean count by colony type across all biological replicates; open circles show the total CFCs for each biological replicate. **c**, Quantification of erythro-myeloid CFC potential of CD34^{+}CD43^{neg} populations, as in **a**(i),(ii), following ROH treatment initiating on day 3, 4 or 5. Two-way ANOVA with Tukey's test was used for all biological replicates ($n=3$), SEM; statistics are shown for BFU-E (CXCR4^{+} control versus D3 $P=0.0005$, D3 versus D4 and D3 versus D5 $P<0.0001$; NS, not significant); the remaining statistics are included in Source Data Fig. 2. Bars show the mean count by colony type across all biological replicates; open circles show the total CFCs for each biological replicate. **d**, Quantification of erythro-myeloid CFC potential of CD34^{+}CD43^{neg} cells, following ATRA treatment on day 3, from isolated (i) WNTd KDR^{+} CXCR4^{+}, (ii) WNTd KDR^{+} CXCR4^{neg} or (iii) WNTi KDR^{+} CD235a^{+} cells, as in **a**. Two-way ANOVA was used to compared to DMSO with Dunnett's test for all biological replicates; statistics are shown for RAi/RAd BFU-E and WNTi Ery-P; the remaining statistics are included in Source Data Fig. 2. Bars show the mean count by colony type across all biological replicates; open circles show total CFCs for each biological replicate. SEM, $^{**}P<0.01$, $^{***}P<0.001$, $^{****}P<0.0001$. WNTi CD235a^{+}: DMSO control, 0.1 nM, 1 nM ($n=6$); 0.01 nM, 100 nM, 1,000 nM ($n=3$); 10 nM ($n=5$); WNTd CXCR4^{neg}: DMSO control, 1 nM ($n=5$); 0.01 nM, 0.1 nM, 10 nM ($n=2$); 100 nM ($n=3$). WNTd CXCR4^{+}: DMSO control ($n=5$); 0.01 nM, 0.1 nM, 10 nM ($n=3$); 1 nM, 100 nM ($n=4$).

We next asked whether ATRA would also specify functional HE from WNTd CXCR4^{+} mesoderm. Titration of ATRA on isolated KDR^{+} CXCR4^{+} mesoderm revealed that a concentration of 1 nM was capable of eliciting robust specification of definitive HE, but concentrations lower than 1 nM and higher than 10 nM did not (Fig. 2d(i)). This indicated that a narrow range of RA signalling

is required to establish an RAd haematopoietic program. Indeed, 1 nM ATRA elicited an RA-dependent response within CXCR4$^+$ mesoderm across three hPSC lines (Extended Data Fig. 4c), indicating that this is a conserved, reproducible population that can be obtained from hPSCs. By contrast, 1–10 nM ATRA led to a reduction of overall definitive haematopoietic development from WNTd KDR$^+$ CXCR4^{neg} mesoderm, whereas ≥100 nM was repressive to HE specification (Fig. 2d(ii)). Similarly, ≥1 nM ATRA was repressive to extra-embryonic-like HE specification from WNTi CD235a$^+$ mesoderm (Fig. 2d(iii)), consistent with a repressive role of RA signalling on extra-embryonic haematopoiesis[8,15].

As these cells represent distinct haemogenic mesodermal populations, we next asked whether correlates exist for each within the early human embryo[16] (Fig. 3a). We first sought to identify mesodermal populations that either simultaneously expressed both *CXCR4* and *ALDH1A2*, or expressed *CDX* genes but lacked *CXCR4* expression, similar to our RNA-seq analyses (Fig. 1f and Extended Data Fig. 3e). Following Seurat label transfer (Fig. 3b), we observed that most embryonic cell types directly corresponded to similarly labelled hPSCs, and cells with matched labels shared many marker genes (Supplementary Table 4a), but emergent mesoderm and the yolk sac-derived lineages (yolk sac mesoderm, HE progenitors) had very few corresponding hPSC populations, while axial mesoderm and erythrocytes had very few cells in either dataset (Fig. 3b).

We next focused on identifying the mesodermal cell types present in both datasets that preceded haemato-endothelial specification (Extended Data Fig. 3a), but were not expressing markers characteristic of commitment to a non-haematopoietic fate (Supplementary Table 4b). This excluded epiblast, ectoderm, endoderm, erythrocytes and HE progenitors as candidate populations. Axial mesoderm was similarly excluded due to its enrichment of somitic genes, and both yolk sac mesoderm and advanced mesoderm were excluded for their enrichment of genes associated with cardiac development. Conversely, the poorly correlated emergent mesoderm-like cells were enriched in genes associated with kidney development. Therefore, if correlates to WNTd haemogenic mesoderm exist, they would be found within the remaining cell types of primitive streak and/or nascent mesoderm. Indeed, both these populations within the embryo had enrichment for *CDX1/2/4* expression, and nascent mesoderm was exclusively enriched in the expression of both *ALDH1A2* and *CXCR4* (Fig. 3c and Supplementary Table 4b).

Independent clustering of these two embryonic populations revealed seven transcriptionally distinct clusters (Fig. 3d). Of those lacking significant *CXCR4* expression (Fig. 3e and Supplementary Table 4c), only cluster 3 had enrichment of *CDX1/2* gene expression (Fig. 3f and Supplementary Table 4c). Gene set enrichment analyses (GSEA) using gene signatures derived from our whole-transcriptome analyses on hPSC-derived KDR$^+$ CXCR4$^{+/neg}$ mesodermal populations (Extended Data Fig. 3a and Supplementary Table 4d) revealed that clusters 1, 2 and 3 were enriched for an hPSC-derived KDR$^+$ CXCR4^{neg} transcriptional signature (Fig. 3g(i) and Supplementary Table 4e). Importantly, cluster 2 was also negatively correlated with the gene signature from KDR$^+$ CXCR4$^+$ cells, and had a negative association with an RA-related biological process Gene Ontology (GO) term (Fig. 3g(i),(ii)). Collectively, these comparisons suggest that the primitive streak cells within cluster 2 may be enriched for functionally equivalent RAi haemogenic mesoderm. However, in the absence of a positive-identifying or cell-autonomous marker for RAi haemogenic mesoderm, the discrete identification of its in vivo correlate within clusters 1, 2 and/or 3 remains unclear.

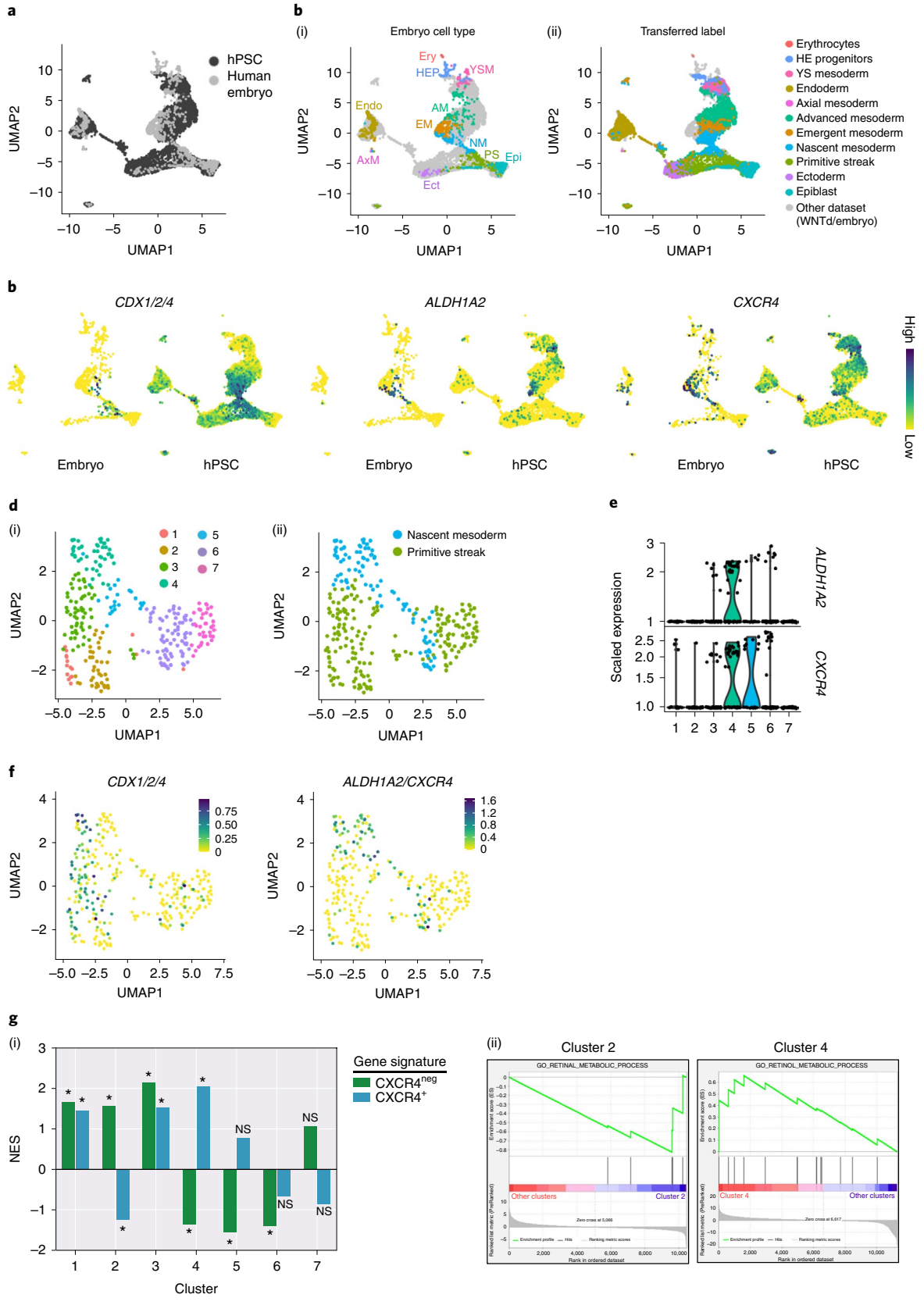
In sharp contrast, *ALDH1A2* and *CXCR4* were exclusively enriched in the nascent mesodermal cluster 4 (Fig. 3e and Supplementary Table 4c). Furthermore, GSEA revealed that only cluster 4 harboured a statistically significant KDR$^+$ CXCR4$^+$ gene signature and simultaneously negatively correlated with the KDR$^+$ CXCR4^{neg} gene signature (Fig. 3g(i) and Supplementary Table 4e). Finally, GO term analysis found that only cluster 4 exhibited a positive enrichment for the biological process relating to retinol metabolism (Fig. 3g(ii)). Collectively, these analyses strongly suggest that the cells within cluster 4 are the in vivo correlates to hPSC-derived RAd mesoderm. Thus, an *ALDH1A2*$^+$CXCR4$^+$ RA-responsive haemogenic mesoderm found in hPSC differentiation cultures is recapitulating the emergence of a similar population found within CS7 nascent mesoderm in the human embryo.

To better understand each hPSC-derived HE, we treated unfractionated WNTd differentiation cultures with either DEAB or ROH on day 3 of differentiation, to induce RAi or RAd definitive haematopoiesis, respectively (Extended Data Fig. 5a). Each culture gave rise to a CD34$^+$CD43^{neg} population, which could be subset by CD73 and CXCR4 expression (Extended Data Fig. 5b). Critically, the multilineage haematopoietic potential of both RAi and RAd HE was found within a NOTCH-dependent CD34$^+$CD43^{neg}CD73^{neg}CXCR4^{neg} population (Extended Data Fig. 5c(i),(ii)). Notably, compared to RAi HE, RAd HE gave rise to significantly more CD34$^+$CD45$^+$ progenitors after nine days of endothelial-to-haematopoietic transition (EHT) culture, had higher erythro-myeloid colony-forming cell (CFC) potential, and their resultant burst forming unit–erythroid (BFU-E) exhibited higher expression of both fetal (*HBG*) globin and *BCL11A* (Extended Data Fig. 5c(ii),(iii),d).

As a final functional assay, we isolated each CD34$^+$ population and assessed their ability to engraft in a murine neonatal recipient. As expected[17], RAi CD34$^+$ cells completely failed to persist in

Fig. 3 | In vivo correlates of hPSC-derived populations within the early human embryo. a,**b**, Cells from hPSC WNTd differentiation cultures and early gastrulating human embryos cluster together following integration of the datasets: UMAP visualizing the contribution of embryonic and hPSCs to the integrated dataset (one biological replicate each) (**a**); UMAP visualizing (i) cell types in the human embryo, as defined in ref. [16], and (ii) the cell type labels transferred from embryonic cells to hPSCs (**b**). **c**, UMAPs visualizing the calculated module score for the expression of *CDX1/2/4*, *CXCR4* or *ALDH1A2* within each dataset. The colour bar shows the expression scaled to each dataset and gene. **d**, UMAP of primitive streak and nascent mesoderm CS7 cells visualizing the (i) clusters and (ii) cell labels. **e**, Violin plot for the scaled expression of *ALDH1A2* and *CXCR4* across clusters, as in **d**. **f**, Simultaneous expression of *CDX1/2/4* or *ALDH1A2/CXCR4* across the subset dataset, as in **d**. The colour bars show the module score calculated for each gene combination. **g**, GSEA for enrichment of WNTd mesodermal gene signatures and RA-related processes within each subset cluster. (i) Normalized enrichment scores (NESs) for each cluster using genes upregulated in KDR$^+$ CXCR4$^+$ and KDR$^+$ CXCR4^{neg} cells, based on all cells within each cluster in **d**(i). Gene signatures were defined as in Supplementary Table 4d. *FDR < 0.25; NS, not significant. Clusters 1, 2 and 3 were enriched for the hPSC-derived KDR$^+$ CXCR4^{neg} transcriptional signature (cluster 1, NES = 1.689, FDR = 0; cluster 2, NES = 1.541, FDR = 0.001; cluster 3 CXCR4^{neg}, NES = 1.530, FDR = 0), and only cluster 4 harboured a statistically significant KDR$^+$ CXCR4$^+$ gene signature (NES = 1.972, FDR = 0). Additionally, cluster 2 was negatively correlated with the gene signature from KDR$^+$ CXCR4$^+$ cells (NES = −1.234, FDR = 0.051), and cluster 4 negatively correlated with the KDR$^+$ CXCR4^{neg} gene signature (NES = −1.350, FDR = 0.017). Other clusters include: cluster 1 CXCR4$^+$ (NES = 1.511, FDR = 0); cluster 3 CXCR4$^+$ (NES = 1.530, FDR = 0); cluster 5 CXCR4$^+$ (NES = 0.800, FDR = 0.880) and CXCR4^{neg} (NES = −1.675, FDR = 0); cluster 6 CXCR4$^+$ (NES = −0.716, FDR = 0.980) and CXCR4^{neg} (NES = −1.241, FDR = 0.123); cluster 7 CXCR4$^+$ (NES = 0.676, FDR = 0.983) and CXCR4^{neg} (NES = 1.081, FDR = 0.603). (ii) Enrichment plots for RA-related metabolic processing GO terms, with opposite trends in cluster 2 (NES = −1.60, P = 0.007, FDR = 0.2147) and cluster 4 (NES = 1.45, P = 0.047, FDR = 0.2498).

murine hosts (Extended Data Fig. 6a). However, the RAd CD34$^+$ cells gave rise to a transiently engrafting human CD45$^+$ population, detectable within both the marrow of the femurs, as well as peripheral blood (Extended Data Fig. 6b–e). The frequency of the observed human chimaerism remained low, becoming undetectable in the marrow after 8–10 weeks post-transplant while persisting in the blood until 10–12 weeks post-injection. Although we observed CD45$^+$CD19$^-$CD33$^-$ haematopoietic cells of unspecified lineage,



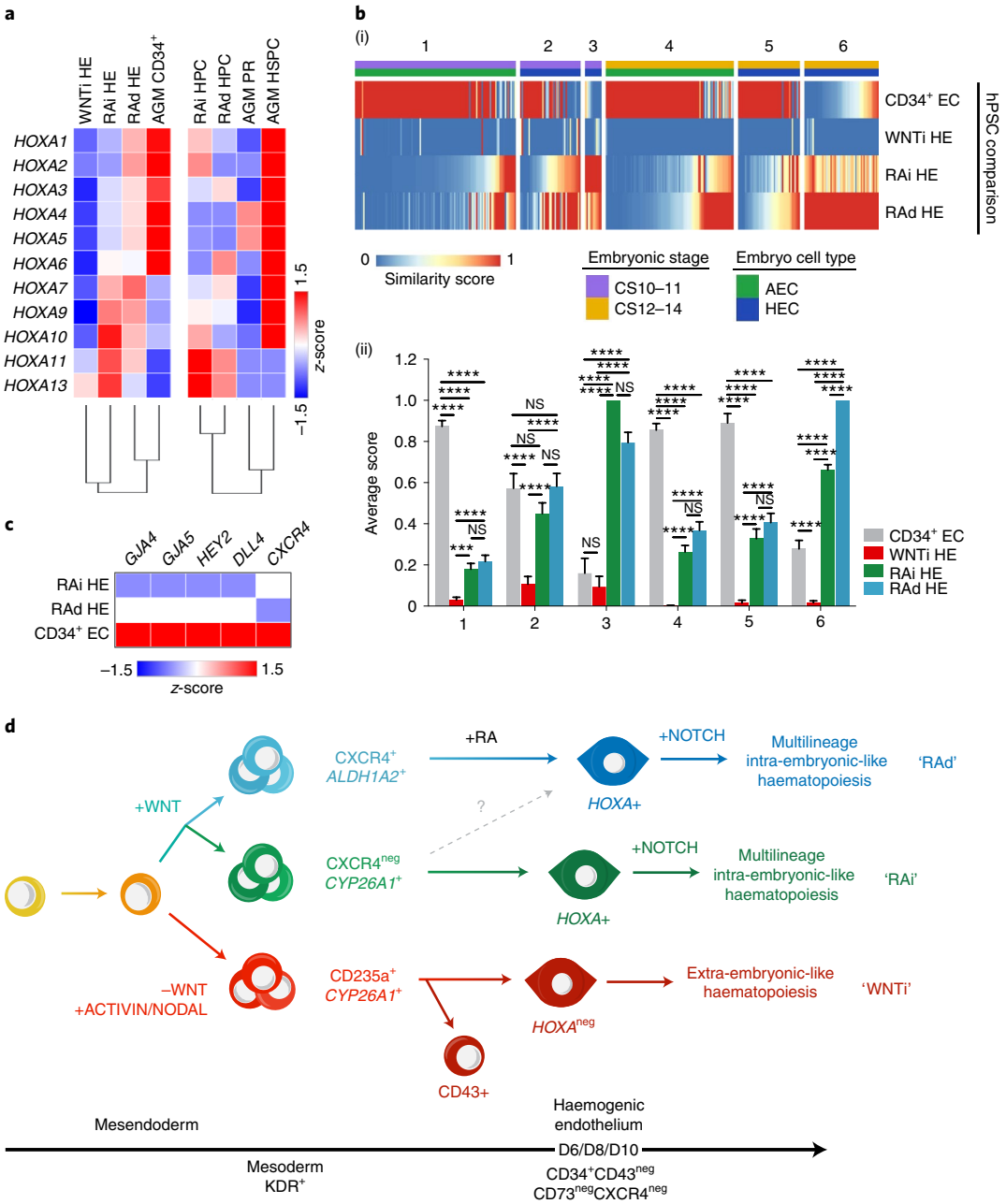


Fig. 4 | Distinct *HOXA*$^{+}$ intra-embryonic-like HE from different ontogenic origins can be specified from hPSCs. a, Heatmap visualizing the relative mean expression of *HOXA* genes within the indicated hPSC and embryonic populations, averaged across all biological replicates (WNTi HE, RAi HE, RAd HE, RAi HPC and RAd HPC: $n=4$; AGM CD34^{+}, AGM HSPC and AGM PR: $n=1$). Left: hPSC-derived CD34^{+} D43^{neg}CD73^{neg}CXCR4^{neg} HE cells and primary embryonic CD34^{+}CD90^{+}CD43^{-} cells from the fifth-week-of-gestation AGM (AGM CD34^{+}). Right: hPSC-derived CD34^{+}CD45^{+} cells (HPC) and primary embryonic CD34^{+}CD90^{+}CD43^{+} haematopoietic stem/progenitor (HSPC) or CD34^{+}CD90^{neg}CD43^{+} committed progenitor (AGM PR) from the fifth-week-of-gestation AGM. Colour bar: robust *z*-scores. **b**, RAi and RAd HE exhibit transcriptional similarity to distinct subsets of intra-embryonic HECs in the human embryo. (i) Heatmap visualizing the similarity of individual early (CS10–11) and late (CS12–14) human embryonic arterial endothelial cells (AEC, *CDH5*$^{+}$*CXCR4*$^{+}$*GJA5*$^{+}$*DLL4*$^{+}$*HEY2*$^{+}$*SPN*^{neg}*PTPRC*^{neg}) or HE cells (HEC, *CDH5*$^{+}$*RUNX1*$^{+}$*HOXA*$^{+}$*ITGA2B*^{neg}*SPN*^{neg}*PTPRC*^{neg}) compared to hPSC-derived CD34^{+}CXCR4^{+} arterial endothelium (CD34^{+} EC), WNTi, WNTd RAi or RAd HE. Colour bar: relative Spearman coefficients. Each column is representative of a single embryonic cell scored across each hPSC-derived population indicated by the row name. Biological replicates: CS10, CS11, CS12, CS14, CS15 ($n=1$), CS13 ($n=2$), all hPSC samples ($n=3$, mean expression). (ii) Average similarity scores for each group of human embryonic cells. Two-way ANOVA with Tukey's multiple comparison test was used to compare all single cells within each group illustrated in (i) (group 1, $n=117$; group 2, $n=44$; group 3, $n=12$; group 4, $n=93$; group 5, $n=45$; group 6, $n=54$), SEM, $^{**}P<0.01$, $^{***}P<0.001$, $^{****}P<0.0001$; NS, not significant. **c**, Heatmap of relative mean expression for arterial genes within hPSC-derived RAi HE, RAd HE and CD34^{+}CXCR4^{+} arterial endothelial cells, averaged across all biological replicates (RAi HE, RAd HE, CD34^{+} EC: $n=3$). Colour bar: robust *z*-score **d**, Schematic of hPSC-derived HE from different ontogenic origins. KDR^{+} CD235a^{+}*CYP26A1*$^{+}$ mesoderm (red) was obtained in a WNT-independent (WNTi) manner, and this population subsequently gives rise to *HOXA*$^{low/neg}$ extra-embryonic-like HE and HPCs. Conversely, two distinct KDR^{+} populations are obtained under WNTd differentiation conditions, each of which gives rise to *HOXA*$^{+}$ intra-embryonic-like HE. KDR^{+} CXCR4^{neg}*CYP26A1*$^{+}$ mesoderm (green) gives rise to multilineage, definitive HE in an RA-independent manner, and KDR^{+} CXCR4^{+}*ALDH1A2*$^{+}$ mesoderm (blue) gives rise to multilineage, definitive HE, in a stage-specific RA-dependent manner.

the presence of both myeloid CD33$^+$ and B-lymphoid CD19$^+$ lineages, the latter of which was not interrogated in vitro, confirmed the multilineage potential of RAd CD34$^+$ cells. Altogether, this indicates that RA signalling results in the generation of a unique population of CD34$^+$ cells that can transiently give rise to multilineage haematopoiesis in vivo.

We next performed whole-transcriptome analyses on each hPSC-derived population. Overall, although RAi and RAd HE were more similar to each other than to either WNTi HE or their derivative CD34$^+$CD45$^+$ HPCs, they did exhibit clear segregation from one another, with >300 differentially expressed genes (Extended Data Fig. 7a(i),(ii) and Supplementary Table 5a,b). Genes enriched in RAi HE were involved in a diverse array of transcriptional pathways, including endothelium development, cellular adhesion, the epithelial-to-mesenchymal transition and the response to mechanical stimuli (Supplementary Table 5c). In contrast, RAd HE was uniquely enriched in RA signalling, as expected, but also in several histone modification pathways, suggesting there may be unexplored regulation at the chromatin level within these cells. Collectively, the transcriptional differences between each population are subtle, highlighting the need for functional assays when contrasting these definitive haematopoietic populations.

We next asked whether either HE population had transcriptional similarity to that found in the developing human embryo. We first compared hPSC-derived RAi and RAd HE to fifth-week-of-gestation CD34$^+$CD90$^+$CD43^{neg} aorta-gonad-mesonephros (AGM) cells[6], where the first HSC-competent HE is detectable[18]. RAi and RAd HE both expressed haemato-endothelial genes, similar to primary embryonic tissue (Extended Data Fig. 7b(i)), but had vastly different enrichment of broad metabolic processes (Supplementary Table 5d–f), which could be reflective of differences between primary in vivo cells and their in vitro correlates. Both RAi and RAd HE exhibited enriched transcriptional signatures for haematopoietic stem and progenitor differentiation (Extended Data Fig. 7b(ii)), consistent with their multilineage in vitro potential. AGM cells, in contrast, harboured pathway enrichment for aorta and vascular development, and BMP and vascular endothelial growth factor (VEGF) signalling pathways (Extended Data Fig. 7b(ii) and Supplementary Table 5e,f), consistent with ongoing vasculogenesis at this developmental stage. Curiously, gene expression of the *HOXA* cluster varied between each population. RAi HE had higher expression of posterior *HOXA* genes, and RAd HE exhibited generally higher expression of anterior and medial *HOXA* genes, with overall higher similarity to that observed in AGM CD34$^+$ cells (Fig. 4a and Supplementary Table 5a). However, this *HOXA* expression was not retained in either population, as hPSC-derived HPCs had lower *HOXA* expression in comparison to the embryonic stem and progenitor population ('HSPC', CD34$^+$CD90$^+$CD43$^+$; Fig. 4a), consistent with their limited engraftment potential.

As a final comparison, we next contrasted these hPSC-derived CD34$^+$ populations with a human embryonic HE scRNA-seq dataset[19]. This comprised intra-embryonic populations of arterial endothelial cells (AEC) and HE cells (HEC)[6,19] (Extended Data Fig. 8a–c), segregated as CS10–11 and CS12–14 (Fig. 4b(i),(ii)), as these populations possess temporally distinct HE subsets, of which the latter is presumed to be HSC-competent[19]. As expected, there was relatively poor similarity between the intra-embryonic cell types with hPSC-derived WNTi HE (Fig. 4b), as the latter is extra-embryonic-like[4]. The embryonic AEC (groups 1 and 4) exhibited strong similarity to hPSC-derived CD34$^+$CXCR4$^+$ arterial-like endothelial cells, regardless of developmental stage (Fig. 4b). Curiously though, many transcriptionally defined embryonic HECs also harboured high similarity to hPSC-derived endothelial cells (groups 2 and 5; Fig. 4b). However, as *RUNX1* can also be expressed within non-haemogenic arterial endothelium[20,21], the strong similarity to hPSC-derived ECs suggests that groups 2 and 5 are probably composed of non-haemogenic endothelial cells. Critically, embryonic HECs in groups 3 and 6 exhibited stage-restricted correlation to either WNTd HE, with CS10–11 ('early') HECs harbouring high transcriptional similarity to RAi HE (group 3), and CS12–14 ('late') HECs exclusively having strong similarity to RAd HE (group 6; Fig. 4b). Interestingly, HECs in group 3 were characterized by only eight marker genes, whereas group 6 HECs were characterized by 194 marker genes, including several genes relating to HSC specification, such as *MYB* and *ANGPT1* (Supplementary Table 6a)[19].

The high similarity of group 6 HSC-competent HECs with RAd HE suggested that these cells may be a precursor to HSC-competent HE. HSC-competent HE harbours an arterial gene signature comprising *CXCR4*, *GJA4*, *GJA5*, *HEY2* and *DLL4*[19,22]. However, *CXCR4* expression is higher in AECs than in HECs[19]. Indeed, groups 3 and 6 exhibited the lowest *CXCR4* expression (Extended Data Fig. 8d and Supplementary Table 6b). This expression pattern was also recapitulated when comparing hPSC-derived CD34$^+$ populations, as RAd HE exhibited higher expression of *GJA4*, *GJA5*, *HEY2* and *DLL4* in comparison to RAi HE, but lower *CXCR4* expression in comparison to AECs (Fig. 4c). Collectively, these analyses have identified that both *HOXA*$^+$ HE populations derived from hPSCs have transcriptionally similar populations within the early human embryo. RAi HE is more similar to a non-arterial 'early' HE, and RAd HE exhibits high transcriptional similarity to embryonic HSC-competent HECs.

We sought to identify a unique signal pathway dependence that defines definitive HE development from hPSCs. Both NOTCH-dependency and *HOXA* expression have been reported as distinguishing characteristics of WNTd CD34$^+$ HE derived from hPSCs[5]. However, here we demonstrate the existence of two ontogenically distinct *HOXA*$^+$ NOTCH-dependent HE populations. Therefore, neither criterion can be solely used to distinguish between various hPSC-derived HE populations. In contrast, here we present evidence of a population that positively responds to RA in a stage-specific manner, giving rise to embryonic-like HE. As this population responds to a narrow concentration range of ATRA, but is inhibited at higher concentrations, our observations indicate that physiologically relevant concentrations are required for this process[23]. Thus, ATRA may not be inhibitory to extra-embryonic-like hPSC haematopoiesis, but is a positive regulator of intra-embryonic-like haematopoiesis.

Notably, these studies demonstrate that hPSC-derived haematopoietic potential is restricted to distinct KDR$^+$ CD34^{neg} mesodermal subpopulations, which are specified very rapidly within differentiation cultures and recapitulate populations found in the early embryo. This is reminiscent of a similar developmental trajectory for cardiomyocyte specification from hPSCs[24], suggesting that major cell fates are specified immediately following gastrulation. There are several lines of evidence that similarly suggest that haemogenic specification is a very early event in the murine conceptus[25,26]. Each hPSC-derived mesodermal population gives rise to an immunophenotypically similar HE population, but each respective population harbours subtle but significant transcriptional and functional differences.

Finally, these studies provide additional insight into the multiple, distinct haematopoietic progenitors that can be obtained from hPSCs (Fig. 4d). hPSC-derived WNTd HE expresses medial *HOXA* genes[6], which we also observe in RAi and RAd HE. However, RAi HE has posterior enrichment of *HOXA* expression, whereas RAd HE has more anterior and medial *HOXA* expression, giving it a higher similarity to primary HSC-competent HE. Our observations that hPSCs can give rise to functionally distinct *HOXA*$^+$ HE populations is inconsistent with the canonical three-wave model of embryonic haematopoietic development[27], wherein all intra-embryonic HE is assumed to be HSC-competent. Instead, these observations are consistent with the identification of heterogeneity in HE within the AGM[19,20,28,29].

The RAd HE derived under these conditions did not give rise to an HSC-like population. Determining whether RAd HE represent bona fide HSC-competent HE precursors or an additional HSC-independent wave of haematopoietic development is an ongoing challenge. Nevertheless, given its functional and transcriptional similarity to a primary intra-embryonic population that harbours HSC-competent HE, we anticipate that this methodology will provide a useful platform for developmental haematopoiesis and disease modelling studies, and in the pursuit of the in vitro generation of therapeutically relevant haematopoietic cells.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-022-00898-9.

Received: 17 September 2021; Accepted: 16 March 2022;
Published online: 28 April 2022

References

1. Chanda, B., Ditadi, A., Iscove, N. N. & Keller, G. Retinoic acid signaling is essential for embryonic hematopoietic stem cell development. *Cell* **155**, 215–227 (2013).
2. Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. & Keller, G. Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 554–561 (2014).
3. 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).
4. Dege, C. et al. Potently cytotoxic natural killer cells initially emerge from erythro-myeloid progenitors during mammalian development. *Dev. Cell* **53**, 229–239 (2020).
5. Ditadi, A. et al. Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).
6. Ng, E. S. et al. Differentiation of human embryonic stem cells to HOXA$^+$ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nat. Biotechnol.* **34**, 1168–1179 (2016).
7. Gao, L. et al. RUNX1 and the endothelial origin of blood. *Exp. Hematol.* **68**, 2–9 (2018).
8. Goldie, L. C., Lucitti, J. L., Dickinson, M. E. & Hirschi, K. K. Cell signaling directing the formation and function of hemogenic endothelium during murine embryogenesis. *Blood* **112**, 3194–3204 (2008).
9. Marcelo, K. L., Goldie, L. C. & Hirschi, K. K. Regulation of endothelial cell differentiation and specification. *Circ. Res.* **112**, 1272–1287 (2013).
10. Dou, D. R. et al. Medial HOXA genes demarcate haematopoietic stem cell fate during human development. *Nat. Cell Biol.* **18**, 595–606 (2016).
11. Yu, C. et al. Retinoic acid enhances the generation of hematopoietic progenitors from human embryonic stem cell-derived hemato-vascular precursors. *Blood* **116**, 4786–4794 (2010).
12. Ronn, R. E. et al. Retinoic acid regulates hematopoietic development from human pluripotent stem cells. *Stem Cell Rep.* **4**, 269–281 (2015).
13. Creamer, J. P. et al. Human definitive hematopoietic specification from pluripotent stem cells is regulated by mesodermal expression of CDX4. *Blood* **129**, 2988–2992 (2017).
14. Kumar, S., Sandell, L. L., Trainor, P. A., Koentgen, F. & Duester, G. Alcohol and aldehyde dehydrogenases: retinoid metabolic effects in mouse knockout models. *Biochim. Biophys. Acta* **1821**, 198–205 (2012).
15. de Jong, J. L. et al. Interaction of retinoic acid and scl controls primitive blood development. *Blood* **116**, 201–209 (2010).
16. Tyser, R. C. V. et al. Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).
17. Sugimura, R. et al. Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432–438 (2017).
18. Ivanovs, A., Rybtsov, S., Anderson, R. A., Turner, M. L. & Medvinsky, A. Identification of the niche and phenotype of the first human hematopoietic stem cells. *Stem Cell Rep.* **2**, 449–456 (2014).
19. Zeng, Y. et al. Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res* **29**, 881–894 (2019).
20. Zhu, Q. et al. Developmental trajectory of pre-hematopoietic stem cell formation from endothelium. *Blood* **136**, 845–856 (2020).
21. North, T. E. et al. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661–672 (2002).
22. Hou, S. et al. Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses. *Cell Res.* **30**, 376–392 (2020).
23. Hernandez, R. E., Putzke, A. P., Myers, J. P., Margaretha, L. & Moens, C. B. Cyp26 enzymes generate the retinoic acid response pattern necessary for hindbrain development. *Development* **134**, 177–187 (2007).
24. Lee, J. H., Protze, S. I., Laksman, Z., Backx, P. H. & Keller, G. M. Human pluripotent stem cell-derived atrial and ventricular cardiomyocytes develop from distinct mesoderm populations. *Cell Stem Cell* **21**, 179–194 (2017).
25. Tanaka, Y. et al. Early ontogenic origin of the hematopoietic stem cell lineage. *Proc. Natl Acad. Sci. USA* **109**, 4515–4520 (2012).
26. Tanaka, Y. et al. Circulation-independent differentiation pathway from extraembryonic mesoderm toward hematopoietic stem cells via hemogenic angioblasts. *Cell Rep.* **8**, 31–39 (2014).
27. Dzierzak, E. & Bigas, A. Blood development: hematopoietic stem cell dependence and independence. *Cell Stem Cell* **22**, 639–651 (2018).
28. Chen, M. J. et al. Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* **9**, 541–552 (2011).
29. Dignum, T. et al. Multipotent progenitors and hematopoietic stem cells arise independently from hemogenic endothelium in the mouse embryo. *Cell Rep.* **36**, 109675 (2021).

Methods

Maintenance and differentiation of human ES and iPS cells. The previously established hESC lines H1[30] and H9[6], and iPSC-1[31], were maintained on irradiated mouse embryonic fibroblasts in hESC media, as described previously[2,32]. For differentiation, hPSC were cultured on Matrigel-coated plasticware (Corning Life Sciences) for 24 h, followed by embryoid body (EB) generation, as described previously[3,33,34]. Briefly, hPSCs were dissociated with brief trypsin-EDTA (0.05%) treatment, followed by scraping. EB aggregates were resuspended in serum-free differentiation medium[35] supplemented with L-glutamine (2 mM), ascorbic acid (1 mM), monothioglycerol (MTG, 4×10^{-4} M; Sigma), transferrin (150 µg ml^{-1}) and BMP4 (10 ng ml^{-1}). After 24 h, basic fibroblast growth factor (bFGF; 5 ng ml^{-1}) was added. On the second day of differentiation, activin A (1 ng ml^{-1}), SB-431542 (6 µM), CHIR99021 (3 µM) and/or IWP2 (3 µM) were added, as indicated. On the third day, EBs were changed to StemPro-34 medium supplemented with L-glutamine, ascorbic acid, MTG and transferrin, as above, with additional bFGF (5 ng ml^{-1}) and VEGF (15 ng ml^{-1}). On day 6, interleukin 6 (IL-6; 10 ng ml^{-1}), insulin-like growth factor 1 (IGF-1; 25 ng ml^{-1}), IL-11 (5 ng ml^{-1}), stem cell factor (SCF; 50 ng ml^{-1}) and erythropoietin (EPO; 2 U ml^{-1} final) were added. All differentiation cultures were maintained at 37 °C. All EBs and mesodermal aggregates were cultured in a 5% CO_2/5% O_2/90% N_2 environment. All recombinant factors are human and were obtained from Biotechne. Analysis of haematopoietic colony potential via Methocult (H4034; Stem Cell Technologies) was performed as described previously[3,5].

For RA manipulation within bulk EB differentiation cultures. Cells were treated exactly as above, except that on day 3 of differentiation, cells were treated with either 10 µM pan-ALDH inhibitor DEAB (4-diethylaminobenzaldehyde, Sigma cat. no. D86256; 'RA-independent') or 5 µM retinol (ROH, Sigma cat. no. R7632; 'RA-dependent'). DEAB or ROH were added again on day 6 of differentiation. HE was FACS-isolated for terminal assays on day 8 (DEAB) or day 10 (ROH).

Stem cell studies were approved by the Washington University Embryonic Stem Cell Research Oversight Committee (ESCRO #14-001), the Ospedale San Raffaele Ethical Committee (TIGET-HPCT) and the Icahn School of Medicine Embryonic Stem Cell Research Oversight Committee (ESCRO #20-06).

Flow cytometry and cell sorting. Cultures were dissociated to single cells, as previously described[2]. All cell sorting was performed in the absence of fetal bovine serum (FBS). Cells were washed, labelled, sorted and collected in StemPro-34 medium. The antibodies used include those previously described[2,3,5]. For hPSC studies, the antibodies included KDR-biotin (clone 89106, Biotechne cat. no. MAB3572, 15:100), streptavidin-PE (BD cat. no. 554061, 1:200), CD4-PerCP-Cy5.5 (clone RPA-T4, BD cat. no. 560650, 1:100), CD8-PE (clone RPA-T8, BD cat. no. 561950, 5:100), CD34-APC (clone 8G12, BD cat. no. 340441, 1:100), CD34-PE-Cy7 (clone 8G12, BD cat. no. 348791, 1:100), CD43-FITC (clone 1G10, BD cat. no. 555475, 10:100), CD45-APC-Cy7 (clone 2D1, BD cat. no. 557833, 3:100), CD45-BV421 (clone 2D1, BD cat. no. 642275, 3:100), CD56-APC (clone B159, BD cat. no. 555518, 4:100), CD73-PE (clone AD2, BD cat. no. 550257, 2:100), CXCR4-APC (clone 12G5, BD cat. no. 555976, 2:100), CXCR4-BV421 (clone 12G5, BD cat. no. 562448, 2:100) and CD235a-APC (clone HIR-2, BD cat. no. 551336, 1:100). For murine xenograft analyses, the antibodies used were CD19-PerCP-Cy5.5 (clone HIB19, BD cat. no. 561295, 1:100), CD33-BV605 (clone HIM3-4, BD cat. no. 744352, 1:100), mCD45-Alexa Fluor 488 (clone 30-F11, BioLegend cat. no. 103122, 1:100) and hCD45-APC (clone 2D1, BD cat. no. 557833, 1:100). Cells were sorted with a FACSAria II (BD) cell sorter and analysed on an LSRFortessa (BD) cytometer. For murine xenograft analyses, the antibodies used were CD19-PerCP-Cy5.5 (clone HIB19) or -PE-Cy7 (clone SJ25C1), CD33-BV605 or -FITC (clone HIM3-4), mCD45-FITC (clone 30-F11, BioLegend) and hCD45-APC or -APC-Cy7 (clone 2D1). Representative gating strategies are demonstrated in Extended Data Fig. 9.

Mesoderm isolation. For isolation of mesodermal populations, day-3-of-differentiation WNTd KDR$^+$ CD235a^{neg}CXCR4$^{+/neg}$ were FACS-isolated and reaggregated at 250,000 cells ml^{-1} in day 3 medium, as above. Cultures were plated in 250-µl volumes in a 24-well low-adherence culture plate, and grown overnight in a 37 °C incubator, with a 5% CO_2/5% O_2/90% N_2 environment. As specified, RA was manipulated with either 5 µM ROH or ATRA (Sigma cat. no. R2625) or 10 µM DEAB. On day 4, an additional 1 ml of RA-supplemented day 3 medium was added to reaggregates. On day 6 of differentiation, CD34$^+$ and CD43$^+$ cells from WNTi cultures were FACS-isolated for terminal assays. WNTd cultures were fed as normal, but without additional RA manipulation. CD34$^+$ cells were sorted from all WNTd populations on day 8 of differentiation.

Haemato-endothelial growth conditions of hPSC-derived haemogenic endothelium. CD34$^+$CD43$^-$ haemogenic endothelium was isolated by FACS and allowed to undergo the EHT as described previously[5,34]. Briefly, cells (CD34$^+$CD43^{neg} or CD34$^+$CD43^{neg}CD73^{neg}CXCR4^{neg} cells) were aggregated overnight at a density of 2×10^5 cells ml^{-1} in StemPro-34 media supplemented with L-glutamine (2 mM), ascorbic acid (1 mM), MTG (4×10^{-4} M; Sigma-Aldrich), holo-transferrin (150 µg ml^{-1}), TPO (30 ng ml^{-1}), IL-3 (30 ng ml^{-1}), SCF (100 ng ml^{-1}), IL-6 (10 ng ml^{-1}), IL-11 (5 ng ml^{-1}), IGF-1 (25 ng ml^{-1}), EPO (2 U ml^{-1}), VEGF (5 ng ml^{-1}), bFGF (5 ng ml^{-1}), BMP4 (10 ng ml^{-1}), FLT3L (10 ng ml^{-1}) and SHH (20 ng ml^{-1}). Aggregates were spotted onto Matrigel-coated plasticware and were cultured for an additional three or nine days for WNTi and WNTd cultures, respectively. Cultures were maintained in a 37 °C incubator, in a 5% CO_2/5% O_2/90% N_2 environment. All resultant cells within the haemato-endothelial cultures were subsequently collected by trypsinization, and assessed for haematopoietic potential by Methocult in a 37 °C incubator, in a 5% CO_2/air environment. In some cases, NOTCH-dependency was assessed by culturing cells, for the entire duration, in the presence or absence of the NOTCH inhibitor gamma secretase inhibitor ('gSI', L-685,458 Sigma cat. no. L1790).

OP9-DL4 co-culture for T-lineage differentiation. OP9 cells expressing Delta-like 4 (OP9-DL4) were a gift from J.-C. Zúñiga-Pflücker, and were generated and described previously[36,37]. A total of $1–10\times10^4$ isolated CD34$^+$CD43^{neg} cells were added to individual wells of a six-well plate containing OP9-DL4 cells, and cultured with rhFlt-3L (5 ng ml^{-1}) and rhIL-7 (5 ng ml^{-1}). rhSCF (30 ng ml^{-1}) was added for the first five days. Cultures were maintained at 37 °C, in a 5% CO_2/air environment. Every five days, co-cultures were transferred onto fresh OP9-DL4 cells by vigorous pipetting and passaging through a 40-µm cell strainer. Following 21–28 days of co-culture, cells were analysed using a LSRFortessa flow cytometer (BD).

Gene expression analyses. Total RNA was prepared for whole-transcriptome sequencing using the Clontech SMARTer kit and was sequenced using an Illumina HiSeq 2500 system with 1×50 single reads. Reads were aligned to GENCODE GRCh38 version 23 using STAR (version 2.7.1a), gene counts were obtained using Subread (version 1.6.5), and batch correction was achieved using Limma. Differential gene expression analysis, fragments per kilobase of exon per million mapped fragments (FPKM) calculations and principal component analysis (PCA) were performed using DESeq2 (version 1.24.0), as described at https://github.com/sturgeonlab/Luff-etal-2022/tree/master/Bulk-analyses. Expression values were filtered to only include protein coding genes (GRCh38) using Ensembl BioMart[38]. Genes were considered significantly enriched with an adjusted P value of <0.05. Euclidean distance was calculated using the generic R function for generating distance matrices, and the accompanying heatmap with hierarchical clustering (Extended Data Fig. 8a(ii)) was plotted using the R package heatmap. Preranked GSEA (Broad Institute GSEA, version 4.1.0) using C5 biological processes as the gene set database was performed, including gene sets containing between 5 and 500 genes, with default parameters. All genes were ranked using the product of the signed $\log_2$(fold change) and inverse log of the P value[39]. GO terms were considered significantly enriched with a $P<0.05$ and a false discovery rate (FDR) of <0.25. Morpheus (https://software.broadinstitute.org/morpheus) was used to create all gene heatmaps of FPKM values, perform the accompanying hierarchical clustering with gene heatmaps (one minus the Pearson correlation with average linkage) and transform data to z-scores. A bulk RNA-seq comparison to scRNA-seq was performed using the SingleR package (version 1.0.1)[40] implemented in R (version 3.6.2), as described at https://github.com/sturgeonlab/Luff-etal-2022/tree/master/SingleR. Embryonic populations from ref. [19] were selected based on the expression of known markers for arterial endothelial cells (AEC, *CDH5*$^+$*CXCR4*$^+$*GJA5*$^+$ *DLL4*$^+$*HEY2*$^+$*SPN*^{neg}*PTPRC*^{neg}) and HE cells (HEC, *CDH5*$^+$*RUNX1*$^+$*HOXA*$^+$ *ITGA2B*^{neg}*SPN*^{neg}*PTPRC*^{neg})[6]. Genes driving the similarities scores are presented in Supplementary Table 6c.

Real-time quantitative polymerase chain reaction with reverse transcription (RT-PCR) was performed as previously described[2]. Briefly, total RNA was isolated with the RNAqueous RNA Isolation Kit (Ambion), followed by reverse transcription using random hexamers and Oligo (dT) with Superscript III Reverse Transcriptase (Invitrogen). Real-time quantitative PCR was performed on a StepOnePlus thermocycle (Applied Biosystems), using the Power Green SYBR mix (Invitrogen). Primers used include *ALDH1A2* (5′-TTGCATTCACA GGGTCTACTG-3′ and 5′-GCCTCCAAGTTCCAGAGTTAC-3′), *CYP26A1* (5′-CTGGACATGCAGGCACTAAA-3′ and 5′-TCTGGAGAACAT GTGGGTAGA-3′) and *BCL11A* (5′-GCCAGAGGATGACGATTGTTTA-3′ and 5′-CCCTCCAGTGCAGAAGTTTATC-3′). Gene expression was evaluated as DeltaCt relative to control (*ACTB*). For globin analysis, the following TaqMan assays (Applied Biosystems) were used: *HBB* (Hs00747223_g1), *HBE1* (Hs00362215_g1), *HBG1/2* (Hs00361131_g1) and *GAPDH* (Hs02786624_g1).

scRNA-seq analyses. Cells from each day-3 differentiation culture condition were methanol-fixed as previously described[41]. Libraries were prepared following the manufacturer's instructions using the 10X Genomics Chromium Single Cell 3′ library and Gel Bead kit v2 (PN-120237), Chromium Single Cell 3′ Chip kit v2 (PN-120236) and Chromium i7 Multiplex kit (PN-120262). A total of 17,000 cells were loaded per lane of the chip, capturing >6,000 cells per transcriptome. cDNA libraries were sequenced on an Illumina HiSeq 3000 system. Sequencing reads were processed and aligned to GRCh38 using the Cell Ranger software pipeline (version 3.1.0). A detailed workflow for these scRNA-seq analyses is available at https://github.com/sturgeonlab/Luff-etal-2022/tree/master/Seurat. Briefly, using

Seurat (version 3.2.2) implemented in R (version 3.6.2), the dataset was filtered by removing genes expressed in fewer than three cells and cells where more than 10% of the unique molecular identifiers (UMIs) are mitochondrial genes, and retaining cells with unique gene counts between 200 and 6,000. The remaining UMI counts were log-normalized with a scale factor of 10,000 and variable features were calculated using the vst selection method. The data were then scaled by regressing out the mitochondrial gene percentage and UMI count. PCA utilizing 50 dimensions was performed and a JackStraw plot was used to determine the dimensionality of the dataset. To finalize the initial processing, uniform manifold approximation and project (UMAP) and shared nearest neighbours were calculated utilizing 50 dimensions, with optimal Louvain clustering resolution determined by maximizing the cluster (SC3) stability, as calculated using the clustree package (version 0.4.3). All gene thresholds were determined using density plots. WNTi and WNTd cells were integrated to account for different sequencing runs using the maximum number of dimensions (50) at all applicable steps. For integration of WNTd mesodermal cultures with the human embryo gastrulation dataset[16], the standard workflow for integration between two datasets was performed in a similar manner to the previous dataset, as described in detail in the online repository (https://github.com/sturgeonlab/Luff-etal-2022/tree/master/Seurat).

Murine xenografts. A total of 0.5–4 × 10^5 $CD34^+$ cells from either anonymous cord blood or DEAB- or ROH-treated hPSC bulk differentiation cultures were isolated by $CD34^+$ MACS (Milenyi) and immediately transplanted into neonatal NOD.Cg-$Kit^{W\text{-}41J}$ Tyr^+ $Prkdc^{scid}$ $Il2rg^{tm1Wjl}$/ThomJ (NBSGW) or NOD.Cg-$Prkdc^{scid}$ $Il2rg^{tm1Wjl}$/SzJ (NSG) mice of both sexes via intrahepatic injection. To assess xenograft persistence, peripheral blood and bone marrow aspirates were obtained at two- to four-week intervals. Peripheral blood was obtained from tail veins, followed by successive washes with red blood cell lysis buffer. Flow cytometry was performed using a BD Fortessa or BD Canto cytometer. Bone marrow aspirates were obtained from right femurs under deep anaesthesia, by inserting a 29-G needle into the distal femur shaft and aspirating the content into a syringe containing 100 µl of phosphate buffered saline, followed by red blood cell lysis. Alternatively, following euthanisia, bone marrow were harvested from femurs by aspirating each bone with Iscove's modified Dulbecco's medium supplemented with 5% FBS, followed by red blood cell lysis. The animals were housed in a standard BSL2 level animal facility, in a dedicated room for immunocompromised animals, according to institutional guidelines.

This study is compliant with all relevant ethical regulations regarding animal research and has been approved by the Washington University Institutional Animal Care and Use Committee, IACUC #19-0959, the Animal Care and Use Committee of Ospedale San Raffaele (IACUC no. 841), and has been authorized by the Italian Ministry of Health and the Icahn School of Medicine Institutional Animal Care and Use Committee (IACUC #PROTO202000171).

Statistics and reproducibility. For all multivariate statistical analyses, analyses of variance (ANOVA) were performed with the appropriate corrections for multiple comparisons. Two-way ANOVA with Tukey's multiple comparison test was used when comparing ≥2 metrics with more than two sorted populations, and one-way ANOVA with Tukey's multiple comparison test was chosen for single metrics with more than two populations. For comparing ≥2 metrics with exactly two populations, two-way ANOVA with Bonferroni's multiple comparison test was used. For Fig. 2d, two-way ANOVA with Dunnett's multiple comparison test was used to analyse the difference in BFU-E between each concentration as compared to the vehicle control. The data distribution was not formally tested, but was assumed to be normally distributed with equal variance. Sample size and replication were determined by historical controls[2,5]. In general, biological replicates were only excluded if internal controls failed and technical replicates (for example, methylcellulose colony-forming assays) were not excluded. Specific to the CXCR4 quantification in the H9 cell line, experiments with poor mesodermal specification (<50% KDR^+) were excluded. Experimental conditions were not randomized, but covariates were controlled by an equal distribution of sorted cells across controls and experimental conditions. Blinding of experimental conditions was not relevant as our studies do not require grading of the results.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

All gene expression analysis datasets are available in the Gene Expression Omnibus (GEO) under accession no. GSE139853 (scRNA-seq in Figs. 1 and 3 and Extended Data Fig. 2; $CXCR4^{+/-}$mesoderm RNA-seq in Fig. 1 and Extended Data Fig. 3; RAi/RAd HE RNA-seq in Fig. 4 and Extended Data Figs. 1, 3 and 7; RAi/RAd HPC RNA-seq in Fig. 4 and Extended Data Fig. 7) or BioProjects under PRJNA352442 (WNTi mesoderm RNA-seq in Fig. 1, Extended Data Figs. 1 and 3; WNTd mesoderm RNA-seq in Extended Data Fig. 1) and PRJNA525404 (WNTi HE RNA-seq in Fig. 4 and Extended Data Figs. 1, 3 and 8). The publicly available datasets used include GEO SuperSeries GSE81102, 'AGM $CD34^+$' (GSM2142333), 'AGM PR' (GSM2142334), 'AGM HSPC' (GSM2142332) for human 5 week AGM (Fig. 4 and Extended Data Fig. 7); GEO GSE135202 for human CS10-CS15 embryo (Fig. 4 and Extended Data Fig. 8); ArrayExpress E-MTAB-9388 for human CS7 embryo (Fig. 3). Genome alignments were performed with GENCODE GRCh38.p3 version 23. All source data have been deposited within the San Raffaele Open Research Data Repository and are available at https://doi.org/10.17632/pbbn55mhy9.1. Source data are provided with this paper.

Code availability

All code used to analyse and visualize datasets has been shared on a repository and can be found at https://github.com/sturgeonlab/Luff-etal-2022.

References

30. Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
31. Park, I. H. et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
32. Kennedy, M., D'Souza, S. L., Lynch-Kattman, M., Schwantz, S. & Keller, G. Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures. *Blood* **109**, 2679–2687 (2007).
33. Dege, C. & Sturgeon, C. M. Directed differentiation of primitive and definitive hematopoietic progenitors from human pluripotent stem cells. *J. Vis. Exp.* **2017**, 55196 (2017).
34. Ditadi, A. & Sturgeon, C. M. Directed differentiation of definitive hemogenic endothelium and hematopoietic progenitors from human pluripotent stem cells. *Methods* **101**, 65–72 (2016).
35. Sturgeon, C. M. et al. Primitive erythropoiesis is regulated by miR-126 via nonhematopoietic $Vcam\text{-}1^+$ cells. *Dev. Cell* **23**, 45–57 (2012).
36. La Motte-Mohs, R. N., Herer, E. & Zuniga-Pflucker, J. C. Induction of T-cell development from human cord blood hematopoietic stem cells by Delta-like 1 in vitro. *Blood* **105**, 1431–1439 (2005).
37. Schmitt, T. M. et al. Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro. *Nat. Immunol.* **5**, 410–417 (2004).
38. Kinsella, R. J. et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxf.)* **2011**, bar030 (2011).
39. Kaspi, A. & Ziemann, M. mitch: multi-contrast pathway enrichment for multi-omics and single-cell profiling data. *BMC Genomics* **21**, 447 (2020).
40. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
41. Alles, J. et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 44 (2017).

Acknowledgements

S.A.L., C.D. and J.P.C. received support from an NHLBI T32 Training Grant (HL007088-41). S.M. is supported by a Vallee Scholar Award and an Allen Distinguished Investigator Award. A.D. is supported by the Telethon Foundation (TIGET grants nos. C4 and G3b) and San Raffaele Hospital (Seed Grant). C.M.S. is supported by an American Society of Hematology Scholar Award, an American Society of Hematology Bridge Grant, a Washington University Center of Regenerative Medicine Pilot Grant, the Bill & Melinda Gates Foundation INV-002414, and NIH R01HL145290 and R01HL151777. This publication was made possible, in part, by grant no. UL1 RR024992 from the NIH National Center for Research Resources (NCRR). R.S. conducted this study as partial fulfilment of an international PhD in Molecular Medicine, Vita-Salute San Raffaele University.

Author contributions

A. Ditadi and C.M.S. formulated the initial concept. S.A.L. and C.M.S. designed the experiments and analysed the data. S.A.L., J.P.C., C.D., R.S., A. Dacunto, S.C., L.N.R., E.C., S.M., A. Ditadi and C.M.S. performed the experiments. S.A.L., S.V. and I.M. performed bioinformatics analyses. S.A.L., A. Ditadi and C.M.S. wrote the manuscript.

Competing interests

The methodology described in this publication is subject to patent no. PCT/US2020/014626 (inventors: A. Ditadi and C.M.S.). The remaining authors declare no competing interests.

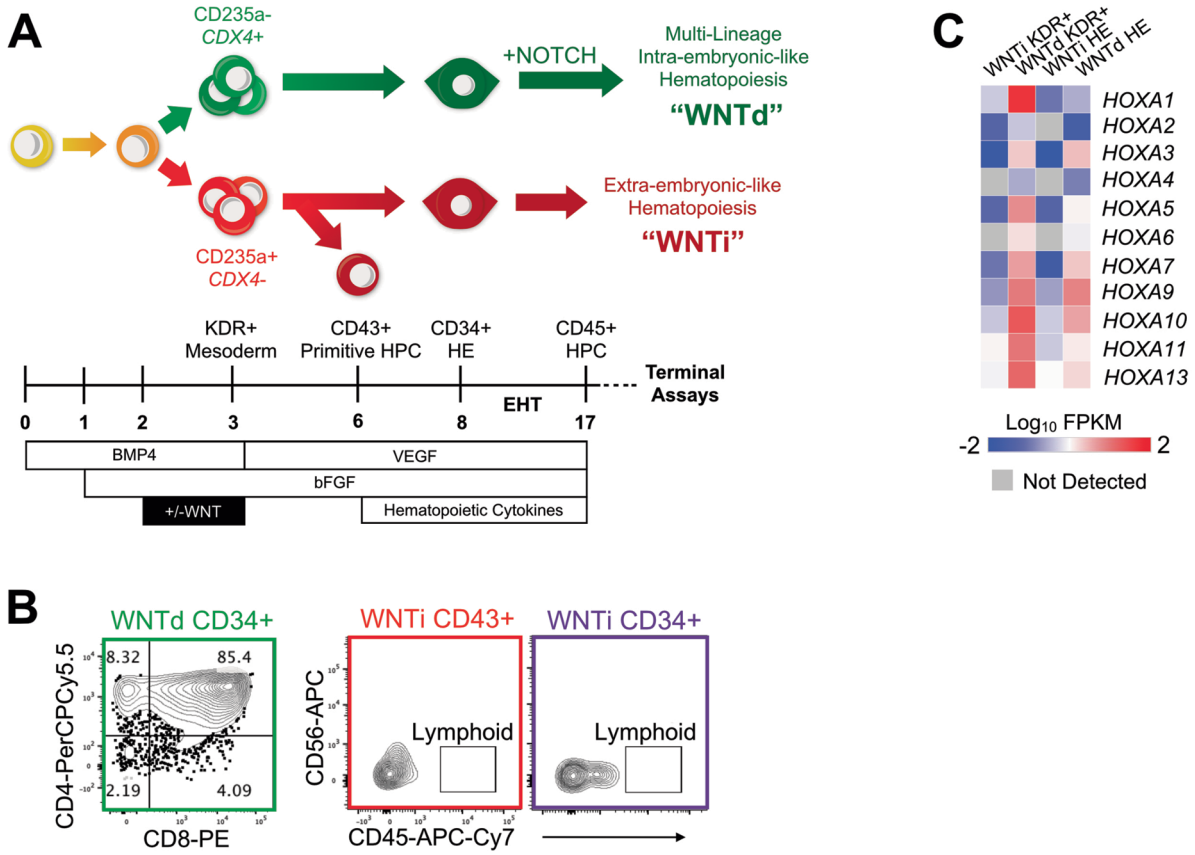
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-022-00898-9.

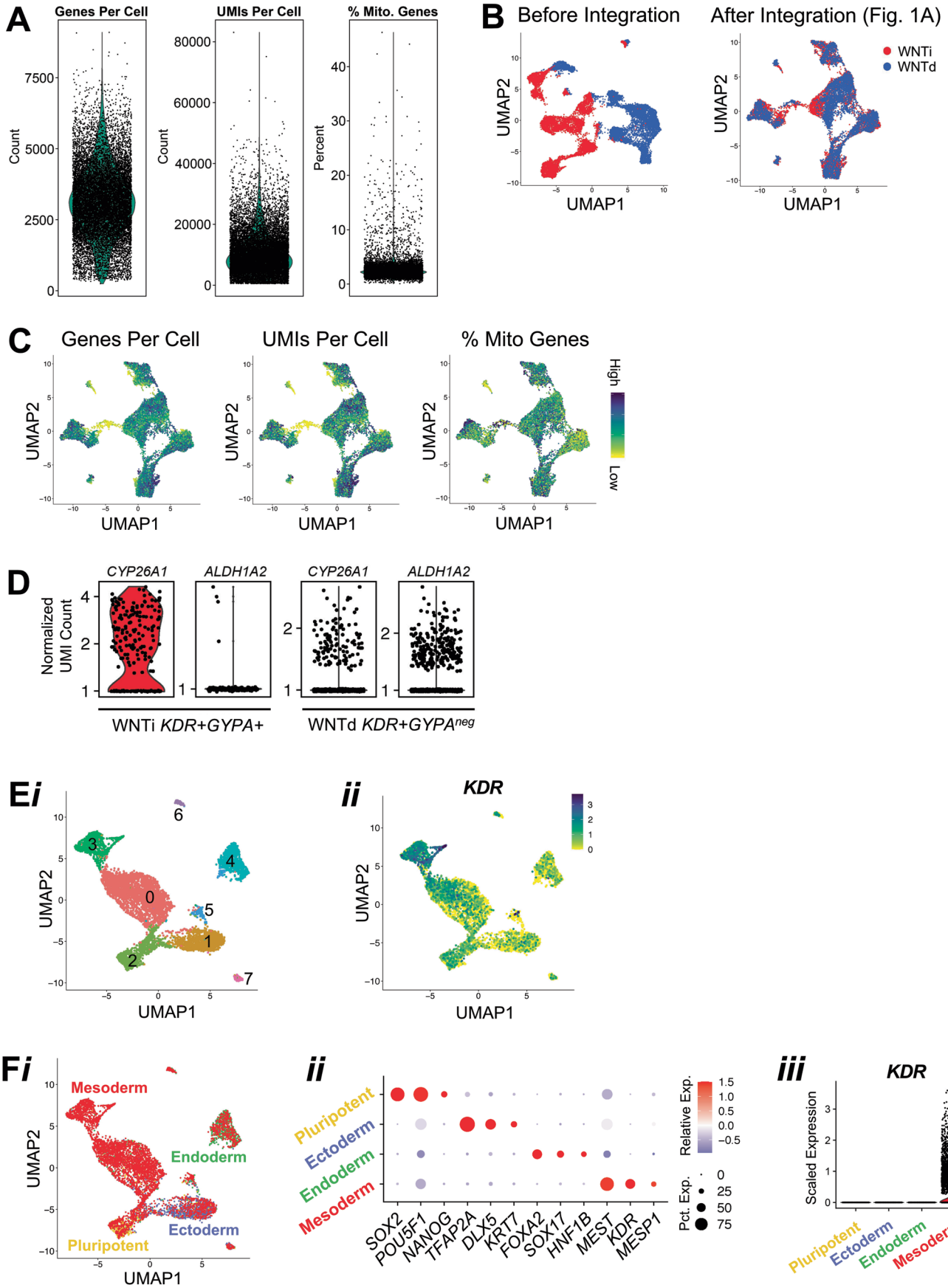
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-022-00898-9.

Correspondence and requests for materials should be addressed to Andrea Ditadi or Christopher M. Sturgeon.

Peer review information *Nature Cell Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

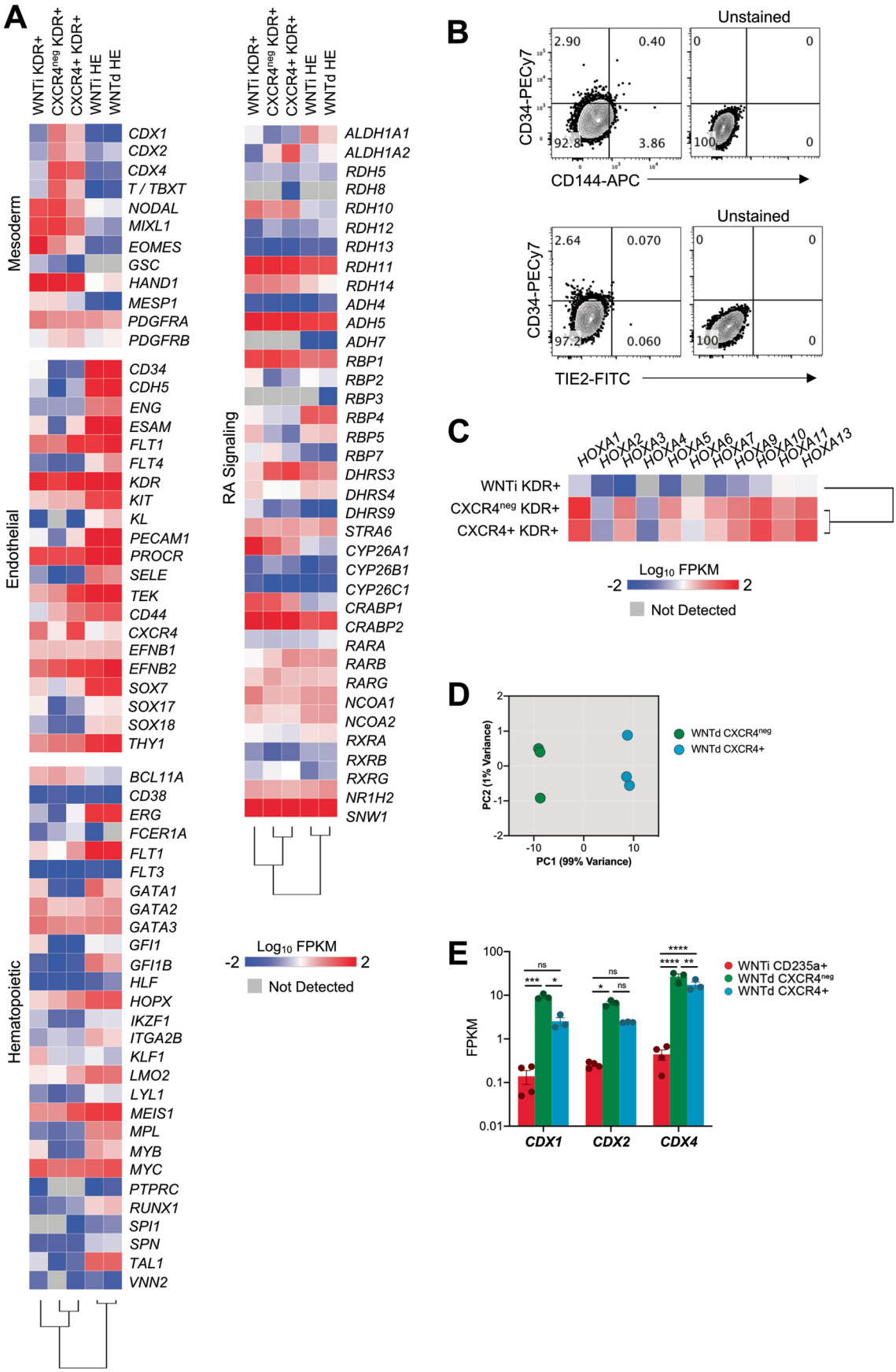


Extended Data Fig. 1 | Specification of HOXA⁺ HE from hPSCs in a WNT-dependent manner. A, Schematic of hPSC directed differentiation towards haemogenic endothelium, as described in Sturgeon et. al[2]. Definitive intra-embryonic-like hematopoietic potential is specified in a WNT-dependent ('WNTd') manner, while extra-embryonic-like hematopoietic potential is WNT-independent ('WNTi'). **B**, Representative flow cytometric analyses of the T-lymphoid potential of WNTd CD34⁺ cells and WNTi CD34⁺CD43^{neg} and CD43+ populations. T cell potential is positively identified by the presence of a CD4⁺CD8⁺ population following 21+ days of OP9-DL4 coculture, while an absence of potential is identified by an absence of CD45+ lymphocytes[3,30]. **C**, Heatmaps visualizing the mean expression of *HOXA* genes across all biological replicates in mesoderm and haemogenic endothelium populations, as in (A). Grey indicates undetected gene. Scale bar: $\log_{10}$ FPKM. Biological replicates: WNTi/WNTd KDR+ (n=4); WNTi/WNTd HE (n=3). The expression of *HOXA* genes within WNTd-derived populations is suggestive of an intra-embryonic-like population[6], while a lack of *HOXA* expression in WNTi-derived populations is suggestive of an extra-embryonic-like population[7].



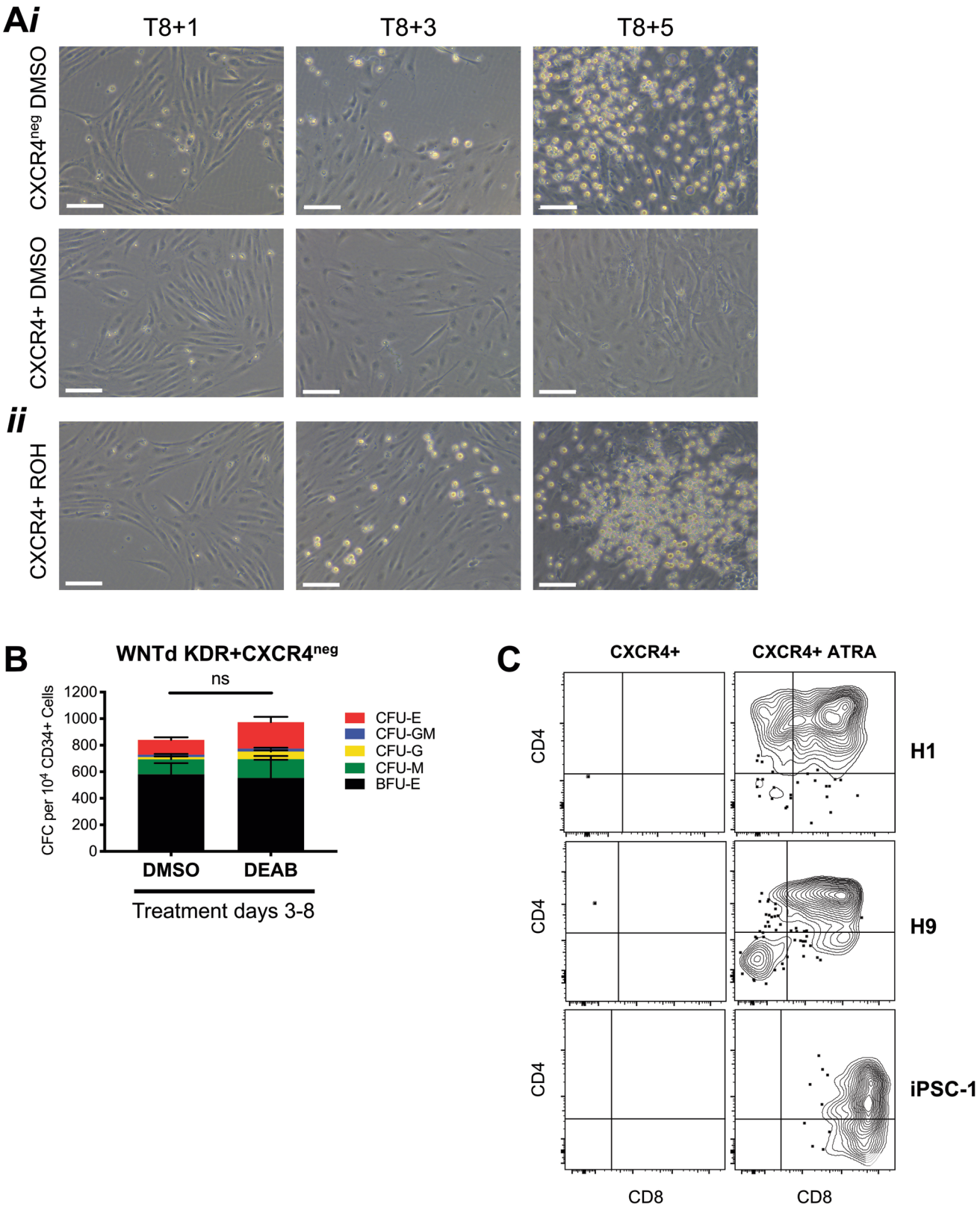
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | scRNA-seq analyses of day 3 WNTi and WNTd differentiation cultures. A, Violin plots visualizing the number of genes per cell (left), the number of unique molecular identifiers ('UMIs', middle), and the percent of expressed genes that are mitochondrial (right) following filtering of low quality cells from both WNTi and WNTd datasets combined (1 biological replicate each). **B**, UMAP visualizing before and after integration of WNTi (red) and WNTd (blue) datasets to account for batch effects between sequencing runs. **C**, UMAP visualizing quality control metrics, as in A, for the dataset following integration. Scale bar: values range as indicated in (A). **D**, Violin plots for *CYP26A1* and *ALDH1A2* expression within WNTi *KDR+GYPA+* and WNTd *KDR+GYPA*^{neg} cells, as indicated. **E,F**, Day 3 WNTd cultures are comprised of all germ layers. **E**, UMAP visualizing (*i*) clustering and (*ii*) *KDR* expression within WNTd differentiation cultures. Scale bar: gene expression scaled to WNTd subset. **F**, (*i*) UMAP plot with the projection of predicted germ layer type, where each label includes cells expressing the following genes: Pluripotent (*SOX2, NANOS3, DND1, POU5F1*, or *TBXT*), Ectoderm (*TFAP2A, DLX5*, or *GATA3*), Endoderm (*FOXA2, APOA1*, or *APOA2*), and Mesoderm (*KDR, MEST, MESP1, TEK*, or *FLT1*). 44 (0.64%) remaining cells were labeled based on clustering. (*ii*) Dot plot visualizing expression of germ layer-specific genes within each identified cell type, as in (*i*). Scale bar: scaled expression (*iii*) Violin plot visualizing the expression of *KDR* within all labeled populations, as in (*i*).

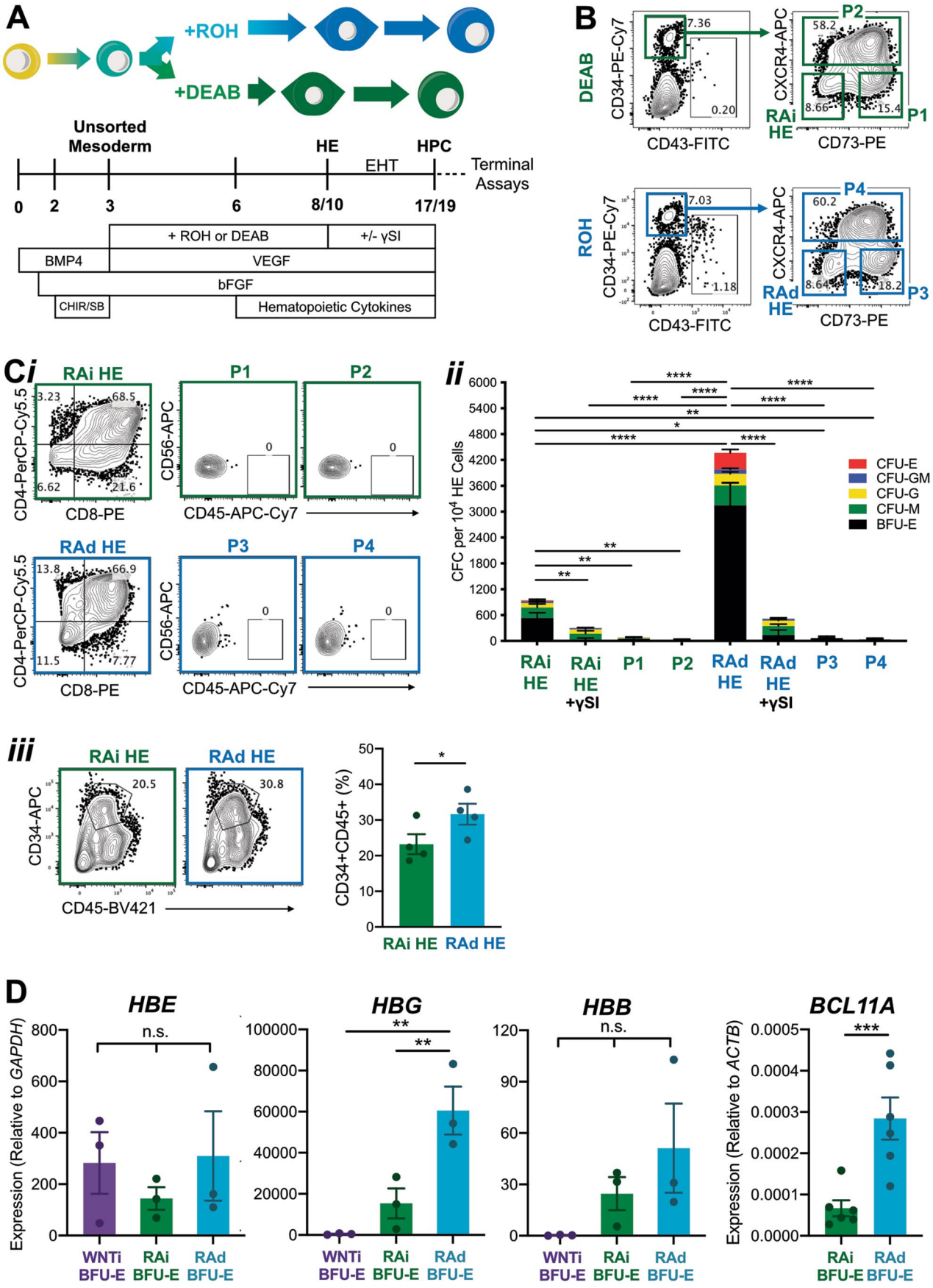


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Day 3 of differentiation WNTd KDR^{+}CXCR4^{neg} and KDR^{+}CXCR4^{+} cells are transcriptionally distinct mesodermal subsets. **A**, Heatmaps visualizing the mean expression across all biological replicates (log_{10} FPKM) of mesodermal, endothelial, hematopoietic, and RA-related genes within hPSC-derived day 3 WNTi KDR^{+}CD235a^{+} cells and WNTd KDR^{+}CXCR4+/^{neg} cells, day 6 WNTi HE, and day 8 WNTd HE. Grey indicates undetected gene. Hierarchical clustering based on the expression of genes shown. Scale bar: log_{10} FPKM. Biological replicates: WNTi KDR^{+} (n=4); CXCR4+/^{neg} KDR^{+}, WNTi HE, WNTd HE (n=3). **B**, Representative flow cytometric analysis for endothelial markers CD34, CD144 (VE-Cadherin/*CDH5*), and TIE2 (*TEK*) within WNTd KDR^{+} cells. **C**, Heatmaps visualizing the mean expression of *HOXA* genes across all biological replicates in day 3 WNTi KDR^{+}CD235a^{+} cells, or WNTd KDR^{+}CXCR4^{+} or KDR^{+}CXCR4^{neg} cells, as in A. Scale bar: log_{10} FPKM. Biological replicates: WNTi KDR^{+} (n=4); CXCR4+/^{neg} KDR+ (n=3). **D**, PCA plot for batch-corrected KDR+CXCR4+ and KDR+CXCR4^{neg} replicates, as in A. **E**, Expression of *CDX* genes within WNTi and WNTd mesodermal populations, as in A, Two-way ANOVA with Tukey's multiple comparison test comparing all biological replicates: WNTd CXCR4+/^{neg} (n=3), WNTi CD235a^{+} (n=4). SEM, **p<0.01, ***p<0.001, ****p<0.0001, ns=not significant.

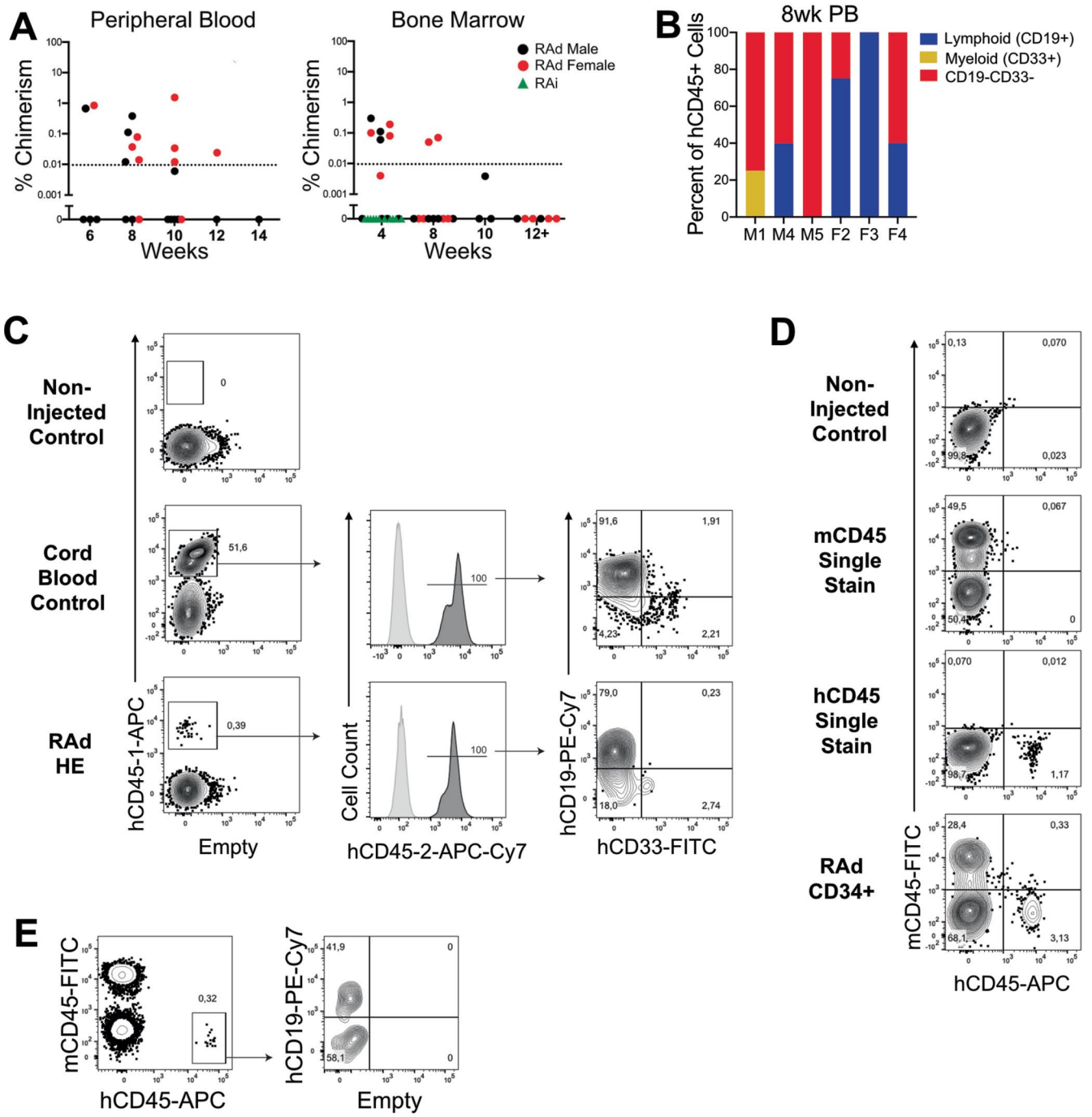


Extended Data Fig. 4 | CXCR4neg mesoderm gives rise to HE in an RA-independent manner, while specification of HE from CXCR4$^+$ mesoderm is RA-dependent. A, Representative phase contrast microscopy of CD34$^+$ cells from DMSO-treated CXCR4+/^{neg} mesoderm (*i*) and retinol-treated CXCR4+ mesoderm (*ii*) following 1, 3, and 5 days after FACS isolation. 100X magnification, scale bar: 50um. **B**, Erythro-myeloid colony forming potential of HE specified from CXCR4^{neg} mesoderm treated with DMSO or DEAB from days 3-8 (HE specification window). Two-way ANOVA with Bonferroni's multiple comparison test comparing all biological replicates, *n*=13, SEM, ns=not significant, p>0.9999 for all colony types. **C**, Representative flow cytometric analyses of T-lymphoid RA-dependent definitive hematopoietic potential of CXCR4$^+$ mesoderm in H1 hESC, H9 hESC, and iPSC-1 hPSC lines. CXCR4$^+$ mesoderm was isolated and cultured as in Fig. 2a. 1 nM ATRA was applied to CXCR4$^+$ cells immediately after FACS isolation. $n \geq 2$.

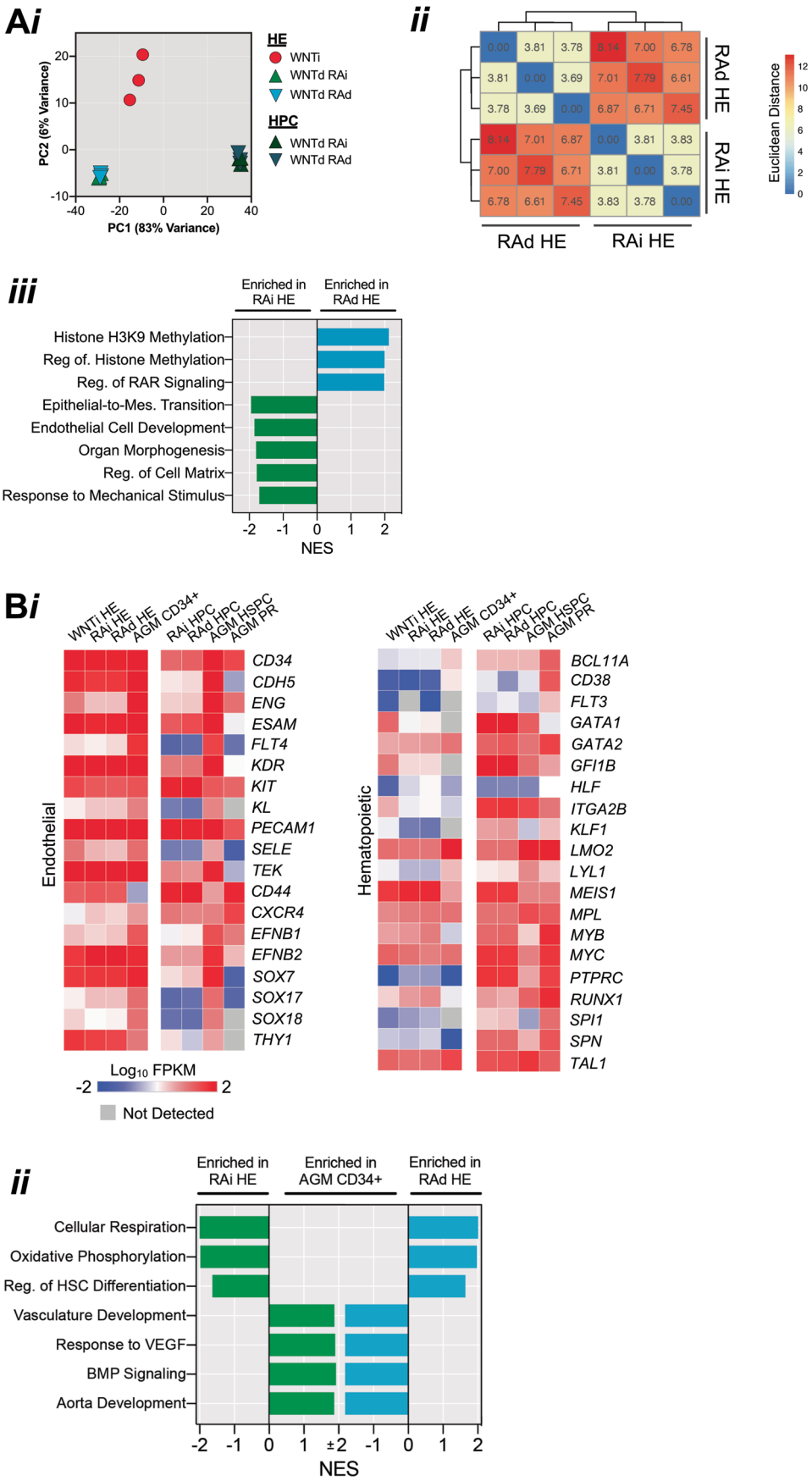


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | RA-dependent HE, similar to RAi HE, is a CD34^{+}CD43^{neg}CD73^{neg}CXCR4^{neg} population and undergoes the endothelial-to-hematopoietic transition (EHT) in a NOTCH-dependent manner. A, Schematic for the differentiation of WNTd cultures towards either RAi and RAd CD34^{+} cells and their respective hematopoietic progenitor cells (HPCs) in the presence or absence of NOTCH inhibitor L-685458. **B**, Representative flow cytometric analyses and FACS isolation strategies within DEAB-treated (RAi) or retinol ('ROH')-treated (RAd) cultures. Isolated populations were then assessed for hematopoietic potential. **C**(*i*), Representative flow cytometric analyses of T-lymphoid potential of populations, as in (B). (*ii*) Quantification of the erythro-myeloid CFC potential from the populations in (B), averaged across all biological replicates. Two-way ANOVA with Tukey's multiple comparison test comparing all biological replicates: RAd HE plus NOTCH inhibitor (n=3), remaining samples (n=4), SEM, statistics shown for BFU-E (RAi HE vs. RAi HE+gSI (p=0.0059), RAi HE vs. RAi CXCR4+ (p=0.008), RAi HE vs. RAi CD73+ (p=0.0044), RAi HE vs. RAd CXCR4^{+} (p=0.0128), RAi HE vs. RAd CD73^{+} (p=0.0033), p<0.0001 for RAi HE vs. RAd HE, RAi HE+gSI vs. RAd HE, RAi CXCR4^{+} vs. RAd HE, RAi CD73^{+} vs. RAd HE, RAd HE vs. RAd HE+gSI, RAd HE vs. RAd CXCR4^{+}, RAd HE vs. RAd CD73^{+}), all comparisons not shown are not significant. All colony counts and statistical analyses are included in Source Extended Data Fig. 5. (*iii*) Representative flow cytometric analysis of CD34 and CD45 at 9 days after HE FACS isolation and percentage of CD34^{+}CD45^{+} cells from each culture, averaged across all biological replicates (*n*=4). Two-tailed paired *t*-test, p=0.0458, SEM. **D**, Average expression across all biological replicates of embryonic (*HBE*), fetal (*HBG*), and adult (*HBB*) globin genes and *BCL11A* in BFU-E derived from WNTi CD34+ cells (purple), WNTd RAi HE (green), and RAd HE (blue). Ordinary one-way ANOVA comparing all biological replicates: *HBE* (*n*=3; WNTi vs. RAi, p=0.4636, WNTi vs. RAd, p=0.8811; RAi vs. RAd, p=0.3841), *HBG* (n=3; WNTi vs. RAi, p=0.234; WNTi vs. RAd, p=0.0018; RAi vs. RAd, p=0.007), *HBB* (*n*=3, WNTi vs. RAi, p=0.3253; WNTi vs. RAd, p=0.0662; RAi vs. RAd, p=0.286), *BCL11A* (*n*=6, p=0.002552), SEM, ns=not significant.

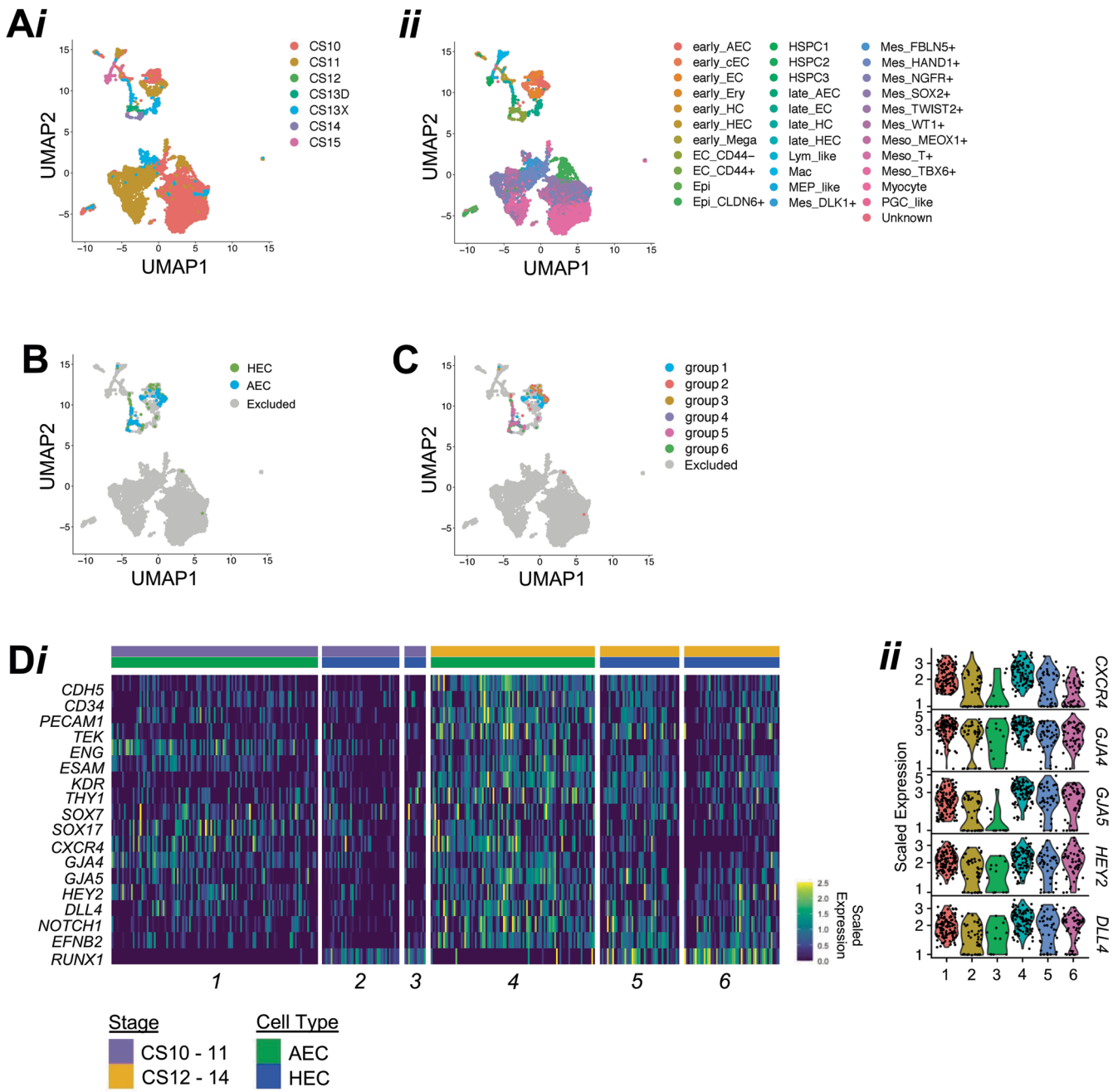


Extended Data Fig. 6 | Xenograft analyses of hPSC-derived HE populations. A, Transient xenograft persistence of RAd hematopoietic progenitors following injection in neonatal mice. Percent chimerism observed of hCD45 cells present in either the peripheral blood and bone marrow following injection with hPSC-derived CD34+ cells. **B**, Lineage distribution of hCD45+ cells present in the peripheral blood (PB) 8 weeks post-transplant. **C-E**, Detailed analysis of human engraftment in the bone marrow or peripheral blood of mice transplanted with RAd HE cells. **C**, Two independent CD45 antibodies ('CD45-1' and 'CD45-2') were used to label human cells. Representative flow cytometric 4-week analysis of the bone marrow of a non-injected control mouse (top), a recipient of 10^5 CD34+ cord blood cells (middle) and a recipient of $5x10^4$ RAd CD34+ cells. **D-E**, Alternative strategy for human chimerism analysis based on mCD45 and hCD45 exclusive staining. **D**, Representative flow cytometric analysis for single stains of both mCD45 and hCD45, from the bone marrow of an non-injected control mouse or a recipient of $2x10^5$ RAd CD34+ cells. **E**, Representative flow cytometric strategy for the detection of human chimerism in the peripheral blood, from a recipient of 10^5 RAd CD34+ cells.

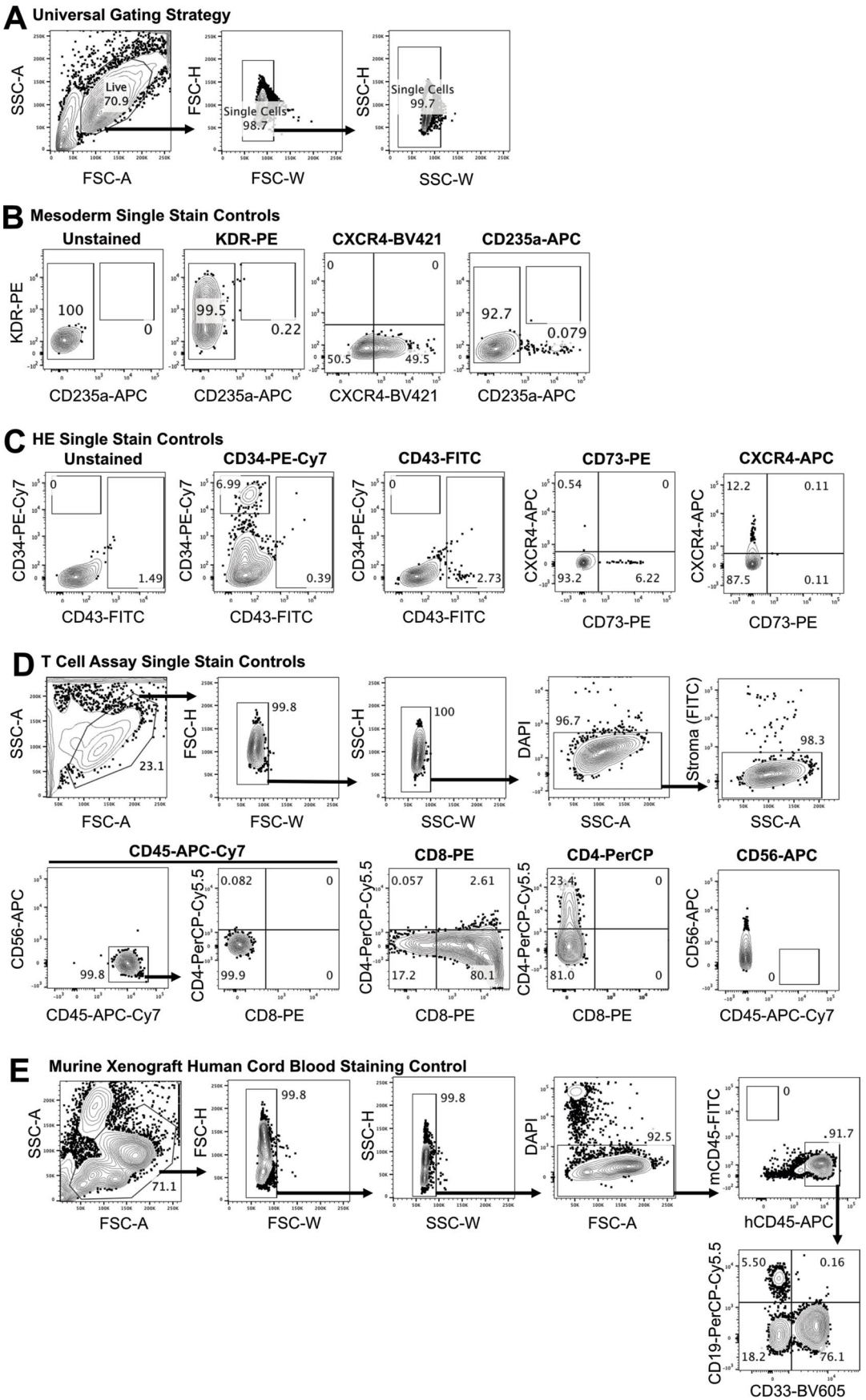


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Whole-transcriptome analysis of hPSC and human embryonic CD34$^+$ populations. A, Comparison of whole transcriptomes of hPSC-derived HE and HPCs (3 biological replicates each). (*i*) PCA plot demonstrating the similarity between batch-corrected biological replicates of WNTi, RAi, and RAd HE, along with CD34$^+$CD45$^+$ HPCs derived from RAi or RAd HE. (*ii*) Heatmap and hierarchical clustering showing the Euclidean distance between all biological replicates of WNTd RAi and RAd HE, as in (i). (*iii*) Selection of significant normalized enrichment scores (NES) from pre-ranked GSEA between RAi HE (green) and RAd HE (blue), as in Supplementary Table 5,c. RAi was enriched in embryonic development (NES=−1.80, FDR=0.23), endothelium development (NES=−1.85, FDR=0.23), cellular adhesion (NES=−1.78, FDR=0.24), the epithelial-to-mesenchymal transition (NES=−1.95, FDR 0.14), and response to mechanical stimuli (NES=−1.71, FDR=0.23). In contrast, RAd HE was enriched for RA signaling (NES=1.99, FDR=0.22) and several histone modification pathways (NES=2.12, FDR=0.14). **B**, (*i*) Mean expression (log_{10} FPKM) of hemato-endothelial genes for all biological replicates within hPSC-derived WNTi HE, RAi HE, RAd HE, RAi HPC, and RAd HPC, as in (A), with CD34$^+$CD90$^+$CD43^{neg} cells ('AGM CD34$^+$'), CD34$^+$CD90$^+$CD43$^+$ hematopoietic stem/progenitors ('AGM HSPC'), and CD34$^+$CD90^{neg}CD43$^+$ hematopoietic progenitors ('AGM PR') isolated from the aorta-gonad mesonephros (AGM) region of 5th-week of gestation human embryos[6]. Grey indicates undetected gene. Scale bar: log_{10} FPKM. Biological replicates: WNTi HE, RAi HE, RAd HE, RAi HPC, RAd HPC (*n*=4); AGM CD34+, AGM HSPC, AGM PR (*n*=1, GEO SuperSeries GSE81102). (*ii*) Selection of significant normalized enrichment scores (NES) from preranked GSEA comparing RAi HE to AGM CD34$^+$ cells (green) and RAd HE to AGM CD34$^+$ cells (blue), as in Supplementary Table 5,e,f. RAi and RAd HE was enriched for hematopoietic stem and progenitor differentiation (NES≥1.63, FDR≤0.015), while AGM CD34$^+$ cells were enriched for aorta and vascular development (NES≥1.82, FDR≤0.01), and BMP and VEGF signaling pathways (NES≥1.89, FDR≤0.004).



Extended Data Fig. 8 | Establishment of human embryonic dataset for comparative analysis with hPSC-derived haemogenic and arterial endothelial populations. A, UMAPs visualizing (*i*) Carnegie Stages and (*ii*) cell type labels from the complete dataset of Zeng, et al.[19]. Biological replicates: CS10, CS11, CS12, CS14, CS15 (*n*=1); CS13 (*n*=2, 1 each for CS13X (10X genomics) and CS13D (Modified STRT-Seq). **B**, UMAP visualizing the cells categorized as arterial endothelial cells ('AEC', defined as *CDH5+CXCR4+GJA5+DLL4+HEY2+SPN*^{neg}*PTPRC*^{neg}) and haemogenic endothelial cells ('HEC', defined as *CDH5+RUNX1+HOXA+ITGA2B*^{neg}*SPN*^{neg}*PTPRC*^{neg}), as in Fig. 3b. **C**, UMAP visualizing the numbered groups for embryonic cells, as in Fig. 3b, within the categorized AEC and HEC. **D***i*, Heatmap visualizing the expression of select broadly and arterial endothelial genes and *RUNX1* in human CS10-14 AEC and HEC. Clusters of AEC and HEC are segregated by their relative similarity to hPSC-derived RAi or RAd HE, as in Fig. 3b. Scale bar: gene expression scaled to subset. (*ii*) Violin plot for scaled expression of select arterial endothelial genes across 6 AEC/HEC clusters, as in (*i*).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Gating strategy and controls for flow cytometric analyses. A, Universal gating strategy for all hPSC-derived flow cytometric analyses. **B**, Single stain controls for markers assessed at the mesodermal stage (day 3 of differentiation). **C**, Single stain controls for markers assessed at the HE stage (day 8 of differentiation). **D**, Gating strategy and single stain controls for T cell assay (day 21 of OP9-DL4 co-culture). **E**, Gating strategy for assessment of xenograft persistence established using human cord blood CD34+ cells.

nature research

Corresponding author(s): Christopher Sturgeon, Ph.D.

Last updated by author(s): February 15, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

For whole transcriptome RNA-sequencing, total RNA was prepared using the Clontech SMARTer kit and was sequenced using an Illumina HiSeq 2500 with 1x50 single reads. Reads were aligned to GENCODE GRCh38 version 23 using STAR (version 2.7.1a).

For single cell RNA-sequencing, cells from each day 3 differentiation culture condition were methanol-fixed as previously described (doi:10.1186/s12915-017-0383-5). Libraries were prepared following the manufacturer's instruction using the 10X Genomics Chromium Single Cell 3' Library and Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3' Chip kit v2 (PN-120236), and Chromium i7 Multiplex Kit (PN-120262). 17,000 cells were loaded per lane of the chip, capturing >6000 cells per transcriptome. cDNA libraries were sequenced on an Illumina HiSeq 3000. Sequencing reads were processed and aligned to GRCh38 using the Cell Ranger software pipeline (version 3.1.0).

All flow cytometric data was collected using BD FACSDiva (version 8).

Data analysis

For whole transcriptome RNA-sequencing, gene counts were obtained using Subread (version 1.6.5), and batch correction was achieved using Limma (version 3.40.6). Differential gene expression analysis, FPKM calculation, and principal component analysis was performed using DESeq2 (version 1.24.0), as described on https://github.com/sturgeonlab/Luff-etal-2022/tree/master/Bulk-analyses. Expression values were filtered to only include protein coding genes using Ensembl BioMart (GRCh38 version 103). Genes were considered significantly enriched with an adjusted p value < 0.05. Euclidean distance was calculated using the generic R function for generating distance matrices, and the accompanying heatmap with hierarchical clustering (Extended Data Figure 7Aii) was plotted using the R package pheatmap (version 1.0.12). Preranked Gene Set Enrichment Analysis (Broad Institute GSEA, version 4.1.0) using C5 biological processes as the gene set database was performed including gene sets containing between 5 and 500 genes, with default parameters. All genes were ranked using the product of the signed log2 fold change and inverse log of the p-value. GO terms were considered significantly enriched with a p value < 0.05 and FDR < 0.25. Morpheus (https://software.broadinstitute.org/morpheus, no version control) was used to create all gene heatmaps of FPKM values, perform the accompanying hierarchical clustering with gene heatmaps (one minus the Pearson correlation with average linkage), and transform data to z-scores.

For single cell RNA-sequencing, a detailed workflow for these scRNA-seq analyses is available at https://github.com/sturgeonlab/Luff-

etal-2022/tree/master/Seurat. Briefly, using Seurat (version 3.2.2) implemented in R (version 3.6.2), the dataset was filtered by removing genes expressed in fewer than 3 cells and cells where greater than 10% of the UMIs are mitochondrial genes, and to retain cells with unique gene counts between 200 and 6000. The remaining UMI counts were log-normalized with a scale factor of 10,000 and variable features were calculated using the vst selection method. The data were then scaled by regressing out mitochondrial gene percentage and UMI count. Principal component analysis utilizing 50 dimensions was performed and a JackStraw plot was used to determine the dimensionality of the dataset. To finalize the initial processing, uniform manifold approximation and project (UMAP) and shared nearest neighbors were calculated utilizing 50 dimensions, with optimal Louvain clustering resolution determined by maximizing cluster (SC3) stability, as calculated using the clustree package (version 0.4.3). All gene thresholds were determined using density plots. WNTi and WNTd cells were integrated to account for different sequencing runs using the maximum number of dimensions (50) at all applicable steps. For integration of WNTd mesodermal cultures with the human embryo gastrulation dataset, the standard workflow for integration between two datasets was performed in a similar manner to the previous dataset, described in detail in the online repository (https://github.com/sturgeonlab/Luff-etal-2022/tree/master/Seurat).

Comparison of the human fetal single cell RNA-sequencing dataset with the whole transcriptomic hPSC datasets was performed using the SingleR package (version 1.0.1) implemented in R (version 3.6.2), as described at https://github.com/sturgeonlab/Luff-etal-2022/tree/master/SingleR. Embryonic populations from Zeng, et al. were selected based on the expression of known markers for arterial endothelial cells ("AEC", CDH5+CXCR4+GJA5+DLL4+HEY2+SPN-PTPRC-) and HE cells ("HEC", CDH5+RUNX1+HOXA+ITGA2B-SPN-PTPRC-). To create the reference dataset for SingleR implementation, FPKMs were log2-transformed. Samples were organized as follows: WNTi HE (type: WNTi HE, main type: HE), WNTd RAi HE (type: RAi HE, main type: HE), WNTd RAd HE (type: RAd HE, main type: HE), CD34+ EC (type: EC, main type: EC).

All flow cytometric data was anaylzed using FlowJo (version 10).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All gene expression analysis datasets are available in the Gene Expression Omnibus (GEO) under the accession numbers GSE139853 (scRNA-seq in Fig. 1, Fig. 3, and Extended Data Fig. 2; CXCR4+/- mesoderm RNA-seq in Fig. 1 and Extended Data Fig. 3; RAi/RAd HE RNA-seq in Fig. 4, Extended Data Fig. 1, Extended Data Fig. 3, and Extended Data Fig. 7; RAi/RAd HPC RNA-seq in Fig. 4, Extended Data Fig. 7), or BioProjects #PRJNA352442 (WNTi mesoderm RNA-seq in Fig. 1, Extended Data Fig. 1, and Extended Data Fig. 3; WNTd mesoderm RNA-seq in Extended Data Fig. 1) and #PRJNA525404 (WNTi HE RNA-seq in Fig. 4, Extended Data Fig. 1, Extended Data Fig. 3, and Extended Data Fig. 7). Publicly-available datasets used include: GEO SuperSeries GSE81102, "AGM CD34+" (GSM2142333), "AGM PR" (GSM2142334), "AGM HSPC" (GSM2142332) for Human 5 week AGM (Fig. 4 and Extended Data Fig. 7); GEO GSE135202 for Human CS10-CS15 embryo (Fig. 4 and Extended Data Fig. 8); ArrayExpress E-MTAB-9388 for Human CS7 embryo (Fig. 3). Genome alignments were performed with GENCODE GRCh38.p3 version 23.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Differentiation experiments were performed in triplicate or quadruplicate. Bulk RNA-sequencing was performed in triplicate, as prior experience instructs is sufficient for statistical power. Single cell RNA-sequencing was performed as one biological replicate, as is standard within the field. No statistical method was used to pre-determine sample size but replication was consistent with our prior publications.
Data exclusions	Differentiation results were excluded when internal control failed to produce hematopoietic progenitors. For CXCR4 quantification in the H9 cell line, experiments with poor mesodermal specification (<50% KDR+) were excluded.
Replication	Differentiation experiments were performed at minimum in triplicate. Number of replicates are indicated in all figure legends. Two additional, separate hPSC lines were employed in Figure 1Eii.
Randomization	Experimental conditions were not randomized but covariates were controlled by equal distribution of sorted cells across controls and experimental conditions.
Blinding	Blinding of experimental conditions was not relevant as our studies do not require grading of the results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	For hPSC studies, antibodies include KDR-Biotin (clone 89106, Biotechne #MAB3572, 15:100), Streptavidin-PE (BD #554061, 1:200), CD4-PerCP-Cy5.5 (clone RPA-T4, BD #560650, 1:100), CD8-PE (clone RPA-T8, BD #561950, 5:100), CD34-APC (clone 8G12, BD #340441, 1:100), CD34-PE-Cy7 (clone 8G12, BD #348791, 1:100), CD43-FITC (clone 1G10, BD #555475, 10:100), CD45-APC-Cy7 (clone 2D1, BD #557833, 3:100), CD45-BV421 (clone 2D1, BD #642275, 3:100), CD56-APC (clone B159, BD #555518, 4:100), CD73-PE (clone AD2, BD #550257, 2:100), CXCR4-APC (clone 12G5, BD #555976, 2:100), CXCR4-BV421 (clone 12G5, BD #562448, 2:100), and CD235a-APC (clone HIR-2, BD #551336, 1:100). For murine xenograft analyses, antibodies used were CD19-PerCP-Cy5.5 (clone HIB19, BD #561295, 1:100), CD33-BV605 (clone HIM3-4, BD #744352, 1:100), mCD45-Alexa Fluor 488 (clone 30-F11, BioLegend #103122, 1:100), and hCD45-APC (clone 2D1, BD #557833, 1:100). All antibodies were commercially validated.
Validation	All antibodies are commercially-validated. Each lot of an antibody is tested for conformance with characteristics of a standard reagent and representative flow cytometric data is included in data sheets to demonstrate specificity and/or sensitivity to a relevant cell population.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	WA01 (H1) and WA09 (H9) were obtained from WiCell Stemcell bank. MSC-iPSC is from an anonymous donor and were generated as in Park et al., 2008. OP9-DL4 were generated as in La Motte-Mohs et al., 2005.
Authentication	No new lines were generated, no lines were authenticated
Mycoplasma contamination	H1, H9, and MSC-iPSC line tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines in the ICLAC registry are used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	NSG and NBSGW mice were used in this study. All recipients were neonates and sex of each recipient is identified in the reported data. The animals were housed in standard BSL2 level animal facility, in dedicated room for immunocompromised animals, according to institutional guidelines.
Wild animals	The current study does not involve wild animals.
Field-collected samples	The current study does not involve field-collected samples.
Ethics oversight	Animal studies have been approved by the Washington University Institutional Animal Care and Use Committee, IACUC #19-0959, the Animal Care and Use Committee of Ospedale San Raffaele (IACUC no. 841) and authorized by the Italian Ministry of Health, and the Icahn School of Medicine Institutional Animal Care and Use Committee, IACUC #PROTO202000171.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Day 3 cells were trypsinized for 5 minutes and washed 5 times in IMDM, 10% FBS, and 10 µg/mL. Day 8 cells were trypsinized for 8 minutes and washed 2 times in IMDM, 10% FBS, and 10 µg/mL. They were further dissociated with Collagenase I for 30 minutes and washed an additional 2 times in Stem Pro 34. All samples were stained in Stem Pro 34 media.
Instrument	Flow cytometry was performed on BD LSR Fortessa and the BD Aria II was used for FACS isolation.
Software	BD FACS Diva was used for data acquisition and FlowJo was used for analysis.
Cell population abundance	500,000 - 1,000,000 of KDR+CXCR4+/- cells and 10,000 - 40,000 CD34+ cells were isolated for each WNTd experiment. 1,000,000 - 1,500,000 KDR+CD235a+ cells and 50,000 - 100,000 CD34+/CD43+ cells were isolated for each WNTi experiment.
Gating strategy	Viable cells were gated using FSC-A/SSC-A and doublets were removed using FSC-H/FSC-W and SSC-H/SSC-W. Autofluorescent cells are removed using the PerCP channel. Representative flow plots and gating strategy demonstrated in Extended Data Fig. 9 Fig. 1E: All plots are gated on CD235a- cells (>98%) Fig. 1G, Extended Data Fig. 3B: All plots are gated on KDR+ cells (60-80%) Fig. 2Ai: Plot is gated on CD235a- cells (>98%) Fig. 2Aiii, Extended Data Fig. 1B (WNTd), Extended Data Fig. 5Ci: CD4/CD8 were gated on CD56-CD45+ cells.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.