A novel human pluripotent stem cell-based gene activation system identifies IGFBP2 as a mediator in the production of hematopoietic progenitors in vitro.

Paolo Petazzi[1], Telma Ventura[6], Francesca Paola Luongo[6], Heather McClafferty[7], Alisha May[6], Helen Alice Taylor[6], Micheal J Shipston[7,8], Nicola Romanò[7,8], Lesley M. Forrester[6], Pablo Menéndez[1-5], Antonella Fidanza[6,9*].

1. Josep Carreras Leukemia Research Institute, Barcelona, Spain.
2. Red Española de Terapias Avanzadas (TERAV)-Instituto de Salud Carlos III (ISCIII), Madrid, Spain.
3. CIBER-ONC, ISCIII, Barcelona, Spain.
4. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.
5. Department of Biomedicine. School of Medicine, University of Barcelona, Barcelona, Spain.
6. Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, UK
7. Centre for Discovery Brain Sciences, Edinburgh Medical School, Biomedical Sciences, University of Edinburgh, Edinburgh, UK
8. Zhejiang University-University of Edinburgh Joint Institute, Zhejiang University School of Medicine, Zhejiang University, Haining 310058, China
9. Edinburgh Medical School, Biomedical Sciences, University of Edinburgh, Edinburgh, UK

* Correspondence to afidanza@ed.ac.uk

Abstract

A major challenge in the stem cell biology field is the ability to produce fully functional cells from induced pluripotent stem cells (iPSCs) that are a valuable resource for cell therapy, drug screening and disease modelling. Here we developed a novel inducible CRISPR-mediated activation strategy (iCRISPRa) to drive the expression of multiple endogenous transcription factors important for *in vitro* cell fate and differentiation of iPSCs to haematopoietic progenitor cells. This work has identified a key role for IGFBP2 in the development of hematopoietic progenitors. We first identified nine candidate transcription factors that we predicted to be involved in blood cell emergence during development, then generated tagged gRNAs directed to the transcriptional start site of these transcription factors that could also be detected during scRNAseq. iCRISPRa activation of these endogenous transcription factors resulted in a significant expansion of arterial-fated endothelial cells expressing high levels of IGFBP2 and our analysis indicated that IGFBP2 is involved in the remodeling of metabolic activity during *in vitro* endothelial to hematopoietic transition. As well as providing fundamental new insights into the mechanisms of

haematopoietic cell fate and differentiation, the broader applicability of iCRISPRa provides a valuable tool for studying dynamic processes controlling developmental events and for recapitulating abnormal phenotypes characterised by ectopic activation of specific endogenous gene expression in a wide range of systems.

Introduction

The haematopoietic system develops early during gestation through so called “waves” of hematopoietic progenitors that arise in different anatomical regions and result in the production of various progenitor and stem cells [1–5]. The precise signalling pathways leading to the production of hematopoietic stem and progenitor cells (HSPCs) during embryonic development are yet to be completely defined, posing a limitation on how to recapitulate the process *in vitro* from pluripotent stem cells.

During development, HSPCs are generated by a subset of endothelial cells, known as hemogenic endothelium [6–9], via the endothelial to hematopoietic transition (EHT) [10]. During the EHT, endothelial cells undergo profound transcriptional remodelling whereby the expression of endothelial genes is gradually downregulated, and the transcription of the hematopoietic program is initiated [11] while the cells round up and eventually detach to enter the circulation [12,13].

To explore the molecular control on the development of the hematopoietic system and to address the differences with the *in vitro* differentiation of induced pluripotent stem cells (iPSCs), we compared our single cell transcriptomics analysis of *in vitro* derived hemogenic endothelium and early progenitors [14] to that of the *in vivo* HSC-primed human hemogenic endothelium [15]. We then developed a novel DOX-inducible CRISPR gene activation system to assess the role of the genes that were expressed at a lower level within *in vitro*-derived cells compared to their *in vivo* counterparts. We employed single-cell RNA sequencing to track the presence of guide RNAs and we monitored the phenotypic effects of gene activation. With this experimental pipeline, we identified and functionally validated a novel role for IGFBP2, IGF Binding protein 2, in the development of *in vitro* progenitors and showed that IGBP2 remodels the metabolic activity during *in vitro* EHT.

Results

Comparison of iPSC-derived endothelial cells with AGM AGM endothelial cells dataset identifies 9 differentially expressed transcription factors.

We, and others, have shown that *in vitro* differentiation of human iPSCs is a powerful tool to model intraembryonic hematopoiesis [14,16–18]. To understand the molecular basis underlying the challenges associated with the *in vitro* production of blood stem and progenitor cells *in vitro* from differentiating iPSCs, we compared our scRNAseq dataset from differentiating iPSCs[14] to that of cells derived *in vivo* from the human aorta-gonad-mesonephros (AGM) region [15]. We integrated the transcriptomic data of *in vitro* derived endothelial (IVD_Endo) and hematopoietic cells (IVD_HPC) with that of arterial endothelial cells (aEC), arterial hemogenic cells (aHEC) and venous endothelial cells (vEC) derived from human embryos collected between Carnegie stage 12 and 14 (Figure 1A-B). To identify possible target genes that could be manipulated *in vitro* to improve iPSCs differentiation, we first determined the transcription factors marking the aHEC *in vivo* (Table 1). Then, we filtered the transcription factor genes that were expressed at lower levels in the IVD_Endo and IVD_HPC. This strategy identified 9 target transcription factors *RUNX1T1*, *NR4A1*, *GATA2*, *SMAD7*, *ZNF124*, *SOX6*, *ZNF33A*, *NFAT5* and *TFDP2* (Figure 1C), whose expression was detected across other AGM datasets in the hemogenic endothelium[17] (Supplementary Figure 1).

Development of a DOX-inducible dCAS9-SAM activation system in human iPSCs

We previously developed an all-in-one Synergistic Activator Mediator system, UniSAM, that mediates the transcriptional activation of endogenous gene expression[19]. To activate the nine target genes identified in this study, we developed a novel doxycycline-inducible SAM (iSAM) cassette targeted into the *AAVS1* locus of human iPSCs (Figure 2A). We demonstrated this strategy could activate of *RUNX1C* expression in HeLa cells and that this was correlated with the doxycycline (DOX) concentration in a linear manner (Figure S2A-B). To verify gene activation in human PSCs at single-cell resolution, we employed a RUNX1C-GFP human embryonic stem cell (hESC) reporter cell line (Figure S2C-G). As predicted, the level of expression of the mCherry tag, the fluorescent mCherry reporter tag within the iSAM cassette, was proportional to the concentration of DOX (Figure S2D-E) and to the number of cells in which RUNX1C-GFP was activated (Figure S2 D-F). Furthermore, the RUNX1C expression level, measured by the mean fluorescence intensity (MFI) of the RUNX1C-GFP reporter, also correlated with the concentration of DOX (Figure S2G). We then tested the iSAM cassette in the iPSC line (SFCi55) (Figure 2B-D). Only when iPSCs were transfected with both iSAM and the gRNA directed to *RUNX1C* and treated with DOX, was the expression of the *RUNX1C* gene and RUNX1 protein detected (Figure 2B-C).

110 We then targeted the iSAM cassette into the *AAVS1* locus using a Zinc Finger Nuclease
111 (ZFN) strategy [14,20,21]. iPSC clones that had specifically integrated the iSAM cassette into the
112 *AAVS1* locus were validated by genomic PCR screening (Figure S2A-B) and sequencing.
113 The *AAVS1* locus has been reported to be a “safe harbour” site that is resistant to epigenetic
114 silencing and indeed, we had previously demonstrated that transgenes inserted into the
115 *AAVS1* locus under the control of the constitutively active CAG promoter were efficiently
116 expressed both in undifferentiated and in differentiated iPSCs [20,22–24]. However, after the
117 iSAM line had been established and cultured under self-renewal conditions, we noted a
118 dramatic reduction in the number of mCherry+ cells in undifferentiated iSAM iPSCs upon
119 DOX induction. We predicted this to be due to transgene-silencing of the rTTA DOX-
120 inducible cassette (Figure S3C). To overcome this problem, we treated the iSAM iPSC line
121 with an inhibitor of histone deacetylases (HDACs), sodium butyrate (SB), reported to have
122 no adverse effect on iPSC maintenance [25,26] and we also confirmed that the treatment with
123 SB had no effect on viability and cell proliferation in our culture conditions (Figure S3F-G). A
124 short 48-hour treatment significantly increased the number of mCherry+ cells upon DOX
125 induction, proportional to the SB concentration (Figure S3D-E). We, therefore, maintained
126 the iSAM iPSCs in the presence of SB and this fully restored the inducibility of the transgene
127 with virtually all cells expressing mCherry in the presence of DOX (Figure S3G). We did not
128 treat the cells with SB during the differentiation since we detected robust expression of the
129 iSAM cassette during the differentiation (Figure S4F). Furthermore, since the HDACs are
130 also involved in the chromatin remodelling during the EHT process[27]we predicted that SB
131 treatment would negatively affect the differentiation.
132 To test the effect of activating the 9 target genes on the transcriptomes of differentiating
133 iPSC cells, we engineered the gRNAs to allow their detection within the single cell RNA
134 sequencing pipeline [28]. We inserted a capture sequence prior to the termination signal to
135 avoid any alteration in the secondary structure of the loops thus preserving the binding of the
136 synergistic activators of the SAM system to the stem loops of the gRNAs. Of the two capture
137 sequences available[28], we decided to use the one that was predicted to result in fewer
138 secondary structure alterations and this new gRNA was named 2.1 (Figure 2E). We
139 compared the activation level achieved with the new 2.1 gRNA to that of the original 2.0
140 backbone using various gRNAs targeting *RUNX1C* (Figure 2F). These results convincingly
141 demonstrated that the addition of the capture sequence in the gRNA 2.1 does not alter the
142 level of endogenous gene activation that could be achieved (Figure 2G). Altogether, these
143 results show that the iSAM system is able to induce gene expression that predictably
144 translates into an increased protein expression and thus provides a platform to steer
145 phenotypical changes in cell identity.

146 To activate our target genes, we designed 5-7 gRNAs in the 200bp upstream of the
147 transcriptional start sites of each of the 9 target genes. We subcloned a total of 49 gRNAs
148 (Table S1) into the gRNA 2.1 backbone and packaged them into lentiviral particles, (herein
149 referred to as the AGM library) as well as a non-targeting (NT) gRNA that was used as
150 control. The iSAM iPSC line was transduced with the targeting gRNAs or the control non-
151 targeting gRNA to generate the iSAM_AGM and iSAM_NT iPSCs line, respectively (Figure
152 S4A). After puromycin selection, their integration in the genome was also confirmed by PCR
153 and sequencing (Figure S3H-I).
154
155 **Single Cell RNA sequencing in combination with CRISPR activation identifies arterial**
156 **cell type expansion in association with higher hematopoietic progenitor potential**
157
158 To assess the transcriptional changes in response to the activation of the target genes, we
159 differentiated the iSAM iPSCs, induced with DOX and subjected them to single-cell RNA
160 sequencing using the 10X pipeline. After 10 days of differentiation, in the presence or
161 absence of DOX, we FAC-sorted live CD34+ cells from iSAM_AGM and the iSAM_NT
162 iPSCs (Figure S4B). Following data filtering, we selected cells in which the gRNAs
163 expression was detected and showed that our approach activated all the target genes,
164 except for ZFN124. A higher level of expression of *RUNX1T1*, *NR4A1*, *GATA2*, *SMAD7*,
165 *SOX6*, *ZNF33A*, *NFAT5*, *TFDP2* was observed following DOX treatment of cells of
166 iSAM_AGM cells compared to the iSAM_NT cells (Figures 3A, S3G). To study the effect of
167 gene activation on transcriptional and cellular phenotype, we performed clustering analysis
168 and detected a total of 7 clusters (Figure 3B-E). A high level of *GJA4* and *DLL4* expression
169 was used to annotate the arterial-like cell cluster, while high levels of hemogenic markers,
170 such as *RUNX1* and *CD44*, were used to annotate the hemogenic clusters, EHT_1 and
171 EHT_2 (Figure 3B, C-D). All the clusters expressed pan-endothelial markers such as
172 *PECAM1* and *CDH5*, coding for CD31 and VECAD, respectively, which get progressively
173 downregulated in EHT_1 and EHT_2, as expected for cells undergoing the EHT process.
174 To assess the effect of the activation on cell identity, we analysed the proportion of the cell
175 clusters between the libraries and detected a significant expansion of the arterial cluster in
176 the DOX-induced iSAM-AGM compared to the iSAM_NT cells (Figures 3E, S4D, S4I). To
177 validate the effect of the activation on the expansion of arterial cell population, we analysed
178 their prevalence by flow cytometry. Although DOX treatment resulted in an average 1.63 ±
179 0.13 fold increase of $CD34^{+}DLL4^{+}$ cells in the control iSAM_NT sample (Figure S3A), the
180 increase observed in the iSAM-AGM cells was significantly greater, with an expansion of
181 3.33 ± 0.73 fold increase of $CD34^{+}DLL4^{+}$ immunophenotypic arterial cells identified by flow
182 cytometry (Figure 3F). To assess the effect of the different cell composition in the CD34+

183 compartment upon activation on the emergence of colony forming progenitors, we isolated
184 $CD34^+$ cells using magnetic beads and cocultured 20,000 cells on OP9 supportive stromal
185 cells for 7 days in the presence of hematopoietic differentiation cytokines. After one week,
186 the progenitor cells were assessed by colony-forming unit (CFU) assays and scored 14 days
187 later. We detected an increased number of CFU-E (colony-forming unit erythroid) and CFU-
188 GM (colony-forming unit granulocyte/macrophage) and a reduction of CFU-M (colony-
189 forming unit macrophage) in iSAM-AGM samples cultured in the presence of DOX compared
190 to the absence of DOX but no significant effect of DOX in SAM_NT samples (Figure 3G).
191 These data indicate that activation of target transcription factors using our novel CRISPR
192 strategy results in transcriptional remodelling and a steer in cell identity that we detected as
193 a functional difference in the hematopoietic progenitor profile.
194
195 **The addition of IGFBP2 to the *in vitro* differentiation leads to a higher number of**
196 **functional hematopoietic progenitor cells.**
197
198 To better understand the molecular mechanism behind the increased progenitor
199 development, we compared the expression profile of the arterial cells between the different
200 activation libraries, and we obtained a list of genes upregulated upon activation of the
201 targets (Table 1). The most upregulated gene, *IGFBP2,* was expressed at significantly
202 higher levels in the iSAM_AGM library in the presence of DOX compared to the others
203 (Figure 4A, Table 1). Although the arterial cells expressed *IGFBP2* at the highest level
204 compared to other cell types, the activation of the gene was not cell-type specific and it was
205 detected across the various clusters (Figure S4H). We then compared the gRNA distribution
206 in these arterial cells from the iSAM_AGM treated with DOX to that of arterial cells without
207 activation. We observed a significant enrichment of the *RUNX1T1-specific* gRNAs (Table1),
208 indicating that the increased *IGFBP2* expression could be downstream of *RUNX1T1*
209 activation. IGF Binding Protein 2 is a member of the family of IGF binding proteins and is
210 thought to be secreted from cells where it then binds IGF1, IGF2, and other extracellular
211 matrix proteins, modulating their function. IGF1 and IGF2 are commonly used in
212 differentiation protocols[29–31], including ours[32,33], due to their direct role in blood development.
213 To test if the increased frequency of functional hematopoietic progenitors was due to
214 IGFBP2 signalling, presumably derived from the arterial cells, we supplemented the media
215 with IGFBP2 at 100ng/ml after the induction of endothelial cell differentiation. To explore the
216 role of IGFBP2 we employed the parental iPSCs line, SFCi55 from which the iSAM line was
217 derived. We isolated CD34+ cells at day 8 and co-cultured them on OP9 cells in presence of
218 IGFBP2 for one week, then tested for their hematopoietic clonogenic potential using CFU
219 assays (Figure 4B). Cells treated with IGFBP2 showed a significant increase in the total

220 number of haematopoietic CFU colonies compared to cells in control cultures. To assess
221 whether IGFBP2 also affected the production of arterial cells themselves, we analysed the
222 proportion of DLL4+ cells, but no difference was detected in the presence of IGFBP2 (Figure
223 4C). This implies that the mechanism of action of IGFBP2 is different from that mediated by
224 the gene activation that led to an expansion of the arterial population. To further understand
225 the relationship between colonies' potential with the arterial identity, we sortedCD34+ into
226 $DLL4^+$ and $DLL4^-$ and cultured them on OP9 co-culture for a week prior to methylcellulose
227 assay. $CD34^+DLL4^+$ and $CD34^+DLL4^-$ plated on OP9 showed different capacity to generate
228 suspension cells (Figure S5D), in line with the colonies' formation results showing that the
229 $DLL4^-$ contains the largest progenitor activity. This aligns with the observation that the EHT
230 process coincides with the downregulation of arterial markers such as *DLL4* and the
231 upregulation of hemogenic markers such as *RUNX1* (Figure 3C).
232 This observation supports our hypothesis that the change in hematopoietic progenitor
233 production following activation in the iSAM_AGM line is associated with both differential
234 gene expression within the arterial cells rather and their expansion. We then focused on the
235 characterization of the cells derived from the CD34+ cells after coculture with the OP9 in the
236 presence of IGFBP2. Our results showed a significant expansion of the CD34+ and CD43+
237 cell populations, further supporting our hypothesis (Figure 4D). To address the potential role
238 of *IGFBP2 in vivo*, we analysed single cell sequencing data from the human AGM at
239 Carnegie stages 14 and 15. *IGFBP2* is highly expressed within the AGM niche by stromal
240 and epithelial cells and, most importantly, highest in endothelial cells (Figure 4E). Together
241 these data show that IGFBP2 addition results in increased hematopoietic blood production *in*
242 *vitro* and that the endothelial compartment is the most likely source of IGFBP2 both *in vitro*
243 and *in vivo*.
244
245 **IGFBP2 enhances metabolic dependency on oxidative phosphorylation of**
246 **differentiating endothelial cells.**
247
248 Following the observation that IGFBP2 supports haematopoietic progenitor differentiation *in*
249 *vitro* from human iPSCs we performed a time course single-cell RNA experiment of SFCi55
250 iPSCs differentiated in its presence. We isolated $CD34^+$ cells and plated them in EHT culture
251 on laminin[34], rather than on OP9 support, to assess exclusively the specific effects of
252 IGFBP2- on differentiating iPSCs. FAC-sorted single/live adherent cells from day 10 and 13
253 in the presence and absence of IGFBP2 were subjected to single-cell RNAseq (Figure S4C).
254 Our time course transcriptomic analyses showed that IGFBP2 induced a change in the
255 transcriptome, specifically on day 13 (Figure 5A). A cluster of endothelial cells enriched for
256 genes associated with the KEGG pathway of growth factor binding, was detected almost

257 exclusively at day 13 (Figure S5F); these cells displayed a different transcriptional signature
258 (as indicated by a shift in their position in the UMAP embedding) in the presence of IGFBP2,
259 and were enriched upon treatment (Figure S5F). Interestingly, this cluster showed
260 expression of *GJA4* across the cells, indicating their broad arterial identity, with a specific
261 reduction of *DLL4* expression level concomitantly to an upregulation of *RUNX1* in the region
262 of the cluster induced by IGFBP2 (Figure 5B). This increase in hemogenic identity within the
263 endothelial cell compartment is consistent with the observation of increased functional
264 progenitor production. This suggests that IGFBP2 supports arterial cells in the acquisition of
265 hemogenic capacity.
266 We compared the transcriptome of cells at day 13 in the presence and absence of IGFBP2
267 and, using KEGG enrichment analysis, we observed that the genes that were upregulated
268 by IGFBP2 were highly enriched in the oxidative phosphorylation term (Figure 5C-D). Since
269 the metabolic switch between glycolytic to mitochondrial metabolism has been previously
270 reported in definitive hematopoiesis[35,36], we tested whether the addition of IGFBP2 could
271 result in a different ATP production profile. We analysed the ATP production at day 13, when
272 RUNX1 expression was induced by IGFBP2 addition, by using specific inhibitors of the
273 complex I and II and complex V to quantify the intracellular ATP and mitochondrial
274 synthesis, respectively. We detected a reduction in glycolytic-derived ATP (Figure 5E-F),
275 which translates to a higher contribution of mitochondria metabolism for IGFBP2-treated
276 cells (Figure 5G). We then looked at the expression levels of genes encoding the enzymes
277 of the glycolytic pathway and its checkpoints, glucose and lactate transporters,
278 monocarboxylate transporters and the enzymes associated with hexokinase and
279 phosphofructose reactions. We observed a general downregulation of the glycolytic
280 enzymes and its checkpoints, with the exception of *PFKM* and *SLC16A1*, (Figures S5A-B),
281 which were expressed at lower levels in the IGFBP2-treated cells at day 13 compared to the
282 control on the same day. These differences were not observed on day 10 cells, which is
283 consistent with our previous observations that the IGFBP2-induced remodelling of the
284 endothelial cells transcriptome happens exclusively on day 13.
285 To assess whether the higher clonogenic potential and effect on the bioenergetic profile of
286 the cells treated with IGFBP2 was the result of an increase in proliferation, we analysed the
287 cell cycle profile of suspension hematopoietic cells obtained from the OP9 cocultured in the
288 presence of IGFBP2. No differences in the cell cycle distribution of hematopoietic
289 progenitors (Figure S3C) was observed, indicating that the detected increase in
290 hematopoietic progenitors is not likely to be a consequence of their increased cycling. We,
291 therefore, concluded the effect of IGBP2 occurs prior to, or during their emergence via the
292 EHT process.

293 In summary, our results show that the activation of the 9 target transcription factors,
294 expressed at higher levels in vivo within the arterial cells in the AGM region, leads to the
295 expansion of the arterial cells and, consequently, to an increase in progenitor activity (Figure
296 6). IGFBP2 was identified as the most upregulated gene in the arterial cell upon gene
297 activation. The addition of IGFBP2 to the differentiation culture induced the upregulation of
298 *RUNX1* and OxPhos genes and accelerated the metabolic dependency on OXPHOS in
299 association with improved progenitor activity (Figure 6).
300
301 **Discussion**
302
303 The complexity and dynamism of developmental hematopoiesis *in vivo* have imposed
304 challenges in accurately reproducing the process *in vitro*. We hypothesized that this could, in
305 part, be due to the inability to recapitulate the appropriate transcriptional programme in
306 cultured cells. Here, we developed a novel CRISPR-activation system to induce the
307 expression of genes that are expressed at low levels in cells that are generated *in vitro*, to
308 explore the downstream consequences of their activation and to assess the effects on
309 emerging hematopoietic progenitor cells.
310 When we compared endothelial cells derived *in vitro* from hiPSCs to those in the AGM
311 region of the human embryo, at the time point when early hematopoietic commitment takes
312 place, we identified nine transcription factors that were expressed at a lower level in the
313 cultured cells. Some of these had been associated with blood cell development, including
314 *GATA2*[37–39], *SMAD7* [40], *NR4A1*[40], *SOX6*[41] and *RUNX1T1*[15,17]. Other genes such as,
315 ZNF124, ZNF33A, NFAT5 and TFDP2 had not been previously associated with
316 hematopoiesis and could provide early evidence of their possible role in hematopoiesis that
317 will require further studies. We were particularly interested in *RUNX1T1* (also known as
318 ETO) which is associated with the leukemic fusion protein, AML1/ETO resulting from the
319 t(8;21) chromosomal translocation[42]. Furthermore, *RUNX1T1* expression has been recently
320 detected in transcriptomic analyses of the human AGM region[15,17], but its precise role during
321 the ontogeny of the blood system has not been elucidated. The addition of a capture
322 sequence to the gRNA backbone enabled their detection coincidentally with the single-cell
323 transcriptome, and this allowed us to demonstrate that RUNX1T1 gRNAs were significantly
324 enriched in cells within an expanded arterial cluster and their presence was associated with
325 the highest expression of *IGFBP2* across clusters. *IGFBP2* KO mice show increased
326 expression of cell-cycle inhibitors and HSC apoptosis, implicating IGFBP2 as a modulator of
327 HSCs cell cycle and survival [43]. More recently *IGFBP2* was reported as being highly
328 expressed in the human AGM region at CS14 when HSCs are emerging [17], supporting a

329 possible role during developmental hematopoiesis but the precise molecular process was
330 unclear. In this study, we show that the addition of IGFBP2 recombinant protein in our *in*
331 *vitro* model of EHT results in the emergence of an increased number of functional
332 hematopoietic progenitors, and we provide an early observation to suggest that RUNX1T1
333 could be involved in the regulation its expression. Because RUNX1T1 lacks a DNA binding
334 domain, its direct involvement in the regulation of IGFBP2 expression, or in the development
335 of the hematopoietic system in general, must require association with other cofactors that
336 are yet to be identified.
337 Our data, together with *in vivo* data[17], show that IGFBP2 is expressed predominantly by
338 endothelial cells. The supportive role of the endothelial niche in the development of HSCs
339 has been studied *in vivo* [44–46] and exploited *in vitro* to support HSCs emergence [45,47].
340 We describe here a novel role for IGFBP2 in the remodelling of the metabolism of iPSC-
341 derived endothelial cells and that this is associated with the induction of RUNX1, a hallmark
342 of hemogenic fate. The addition of IGFBP2 in the culture induces upregulation of the genes
343 of the oxidative phosphorylation pathway in association with an increased relative
344 mitochondrial contribution to cellular ATP production. This is due to reduced glycolysis in
345 association with the downregulation of genes coding for glycolytic enzymes and their
346 checkpoints. The switch between glycolytic to mitochondrial metabolism has been shown to
347 be essential for the EHT in definitive hematopoiesis[36,48]. This switch is induced *in vivo* by
348 mechanical cues downstream of the establishment of circulation, and it is required for
349 functional HSCs development[35], while *in vitro*, the switch is driven by pyruvate mitochondrial
350 catabolism, leading to definitive EHT as opposed to primitive[36]. Further *in vivo* studies are
351 required to characterize the role of IGFBP2 during AGM hematopoiesis to overcome the
352 limitation of using an *in vitro* model such as iPSCs differentiation.
353 In conclusion, we detected transcriptional differences between *in vivo* and *in vitro*
354 developmental hematopoiesis and developed a novel inducible gene activation system to
355 identify novel molecular players during the endothelial-to-hematopoietic transition. Our
356 multidisciplinary approach identified IGFBP2 as novel signalling molecules that support
357 human blood progenitor development *in vitro*, inducing a metabolic switch from cytoplasmic
358 glycolysis to mitochondrial respiration. This study demonstrates that combining CRISPR-
359 mediated activation of target genes with single-cell transcriptomic analysis in differentiating
360 hPSCs can be a powerful approach to alter cell fate, providing a tool for gene function
361 studies during human development. The fine epigenetic manipulation of the transcription can
362 be readily applied to any cell lineage simply by adding specific gRNAs and it will be
363 instrumental in exploring other developmental processes that can be, at least partially, be
364 mimicked *in vitro* with human iPSCs. Some limitations remain in applying CRISPR-mediated

gene activation strategies in long differentiation protocols due to the challenge of detecting high copy numbers of the gRNAs, limiting the possibility of providing the fine statistical correlation needed to predict downstream target genes. Testing different promoters driving the gRNA expression or structural modifications of the gRNA scaffold could result in a more robust expression and allow for correlation analysis.

Finally, using this approach, we have identified the supportive role of IGFBP2, predominantly produced by endothelial cells, which induces transcriptional and metabolic remodelling in association with the induction of *RUNX1* expression, and result in higher hematopoietic progenitors' activity.

Figures

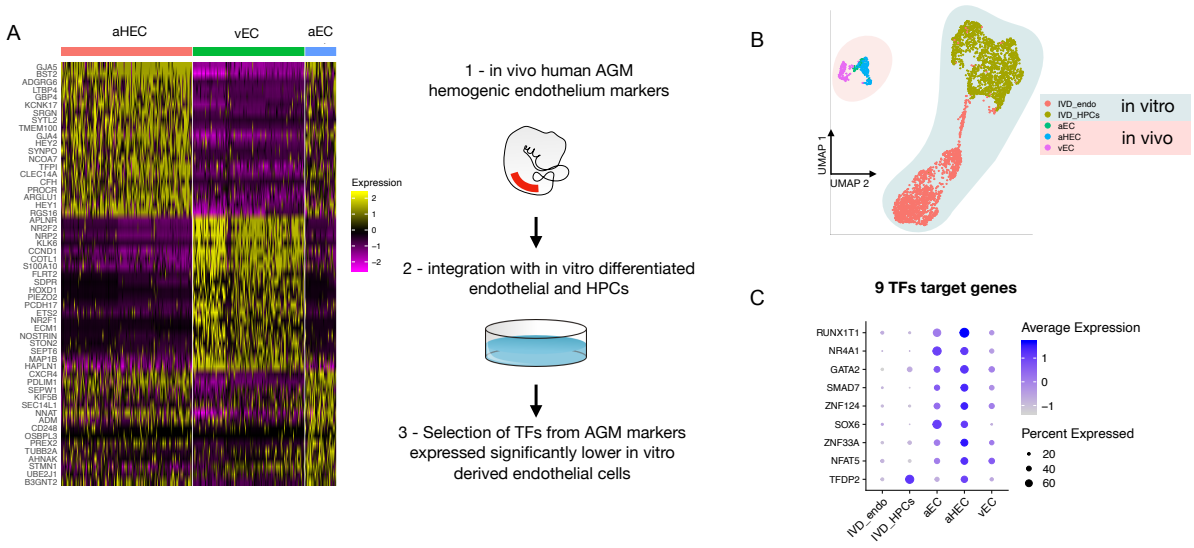

Figure 1 - Comparison of *in vitro* iPSC-derived and *in vivo* AGM-derived endothelial cells identifies 9 differentially expressed transcription factors.

A - Schematic of the analytic pipeline used to identify the target genes. **B** - Integrative analysis of single cell transcriptome of *in vitro* derived endothelial (IVD_Endo) and hematopoietic cells (IVD_HPCs) with *in vivo* developed endothelial cells (venous, vEC; arterial, aEC; arterial hemogenic, HECs) from human embryos (CS12-CS14) visualised on UMAP dimensions**. C** - Target genes expression level showing higher expression in arterial hemogenic endothelium *in vivo* than *in vitro* derived cells.

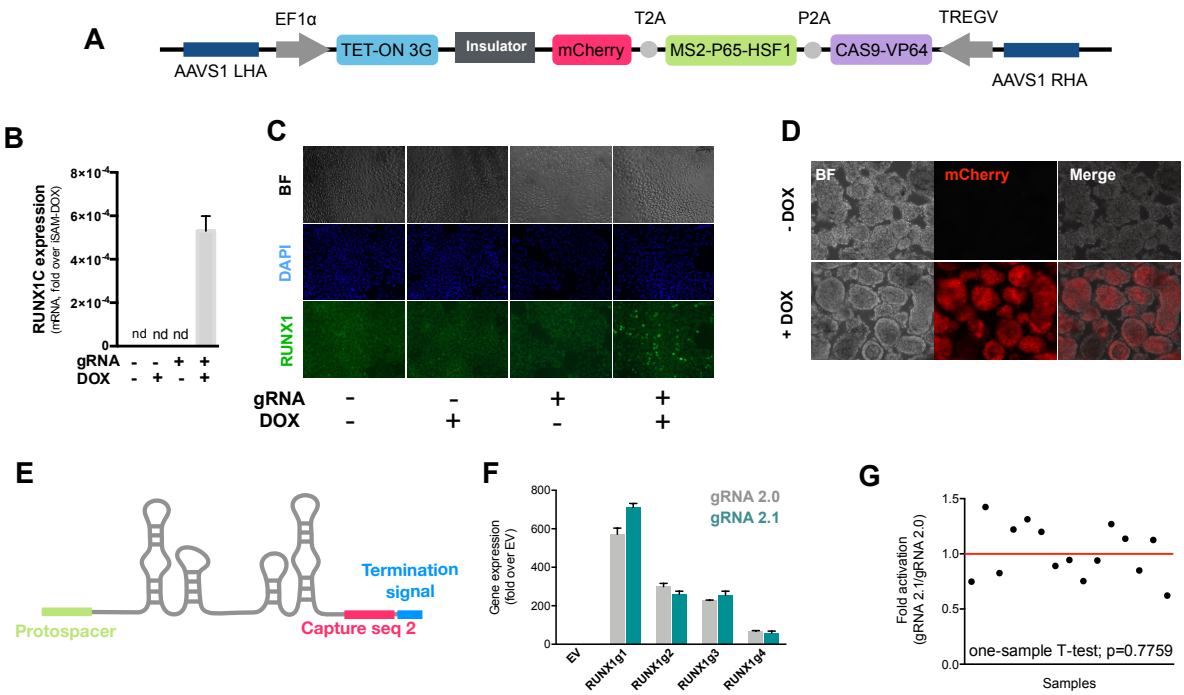

Figure 2 - The inducible iSAM cassette successfully mediates activation of endogenous gene expression upon DOX induction.

A - Schematic of the iSAM cassette containing the TET-on system under the control of EF1α and dCAS9-P2A-MS2-p65-HSF1-T2A-mCherry under the rTTA responsive elements, separated by genetic silencer and flanked by AAVS1 specific homology arms. **B** - *RUNX1C* gene expression activation after transient transfection of the iSAM plasmid and gRNAs in presence or absence of DOX in human iPSC line (n=3 from independent transfections). **C** – RUNX1 protein expression upon iSAM activation after transient transfection of the iSAM plasmid and gRNAs with DOX in human iPSC line detected by immunostaining. **D** Expression of the iSAM cassette reported by mCherry tag during the differentiation protocol, the representative images (bright field – BF, and fluorescence) show embryoid bodies at day 3 of differentiation. **E** - Schematic of the gRNA 2.1 containing the capture sequence for detection during the scRNAseq pipeline. **F** - *RUNX1C* gene activation level obtained using either the gRNA 2.0 or 2.1 backbone (n=3 from independent transfections of the 4 different gRNAs). **G** - Statistical analysis of the gRNAs activation level showing no significant variation following addition of the capture sequence (n=3 for each of the 4 different gRNAs).

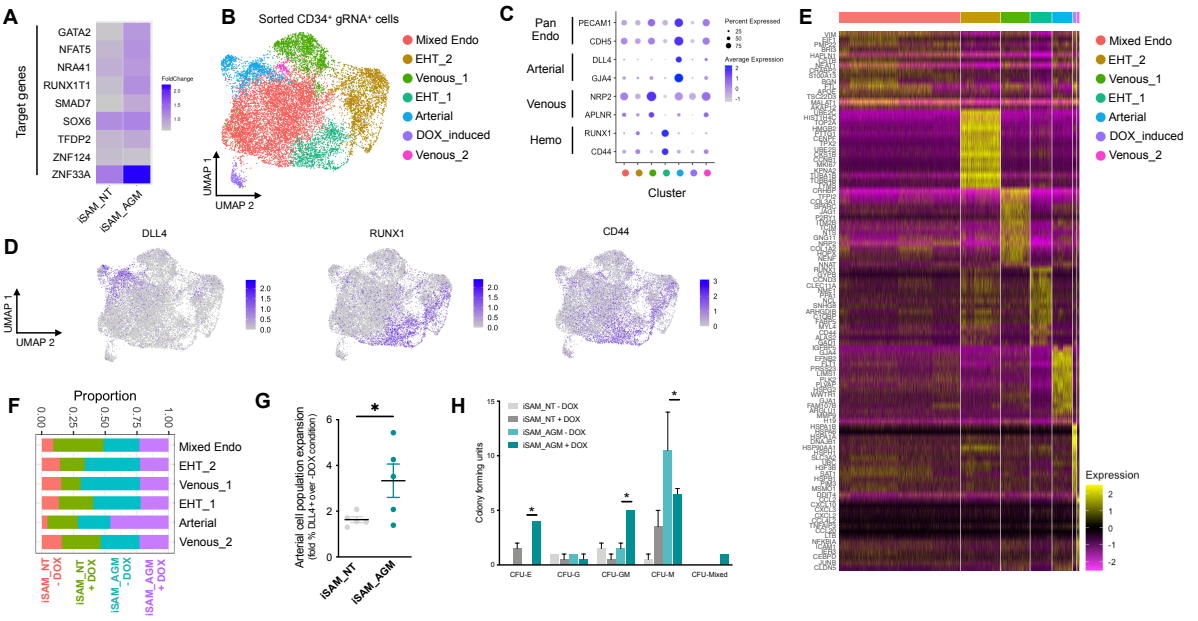

Figure 3 - Single Cell RNA sequencing in combination with CRISPR activation identifies arterial cell type and functional hematopoietic expansion in association with activation of the 9 target genes.

A - Gene expression profile of target genes following target genes' activation, heatmap shows the expression level of the target genes in the iSAM_NT and iSAM_AGM treated with DOX following normalisation on the -DOX control. **B** - Dimension reduction and clustering analysis of the scRNAseq data following activation, filtered on cells where the gRNA expression was detected. **C** - Arterial (*GJA4*, *DLL4*), venous (*NRP2*, *APLNR*) and hemogenic marker (*CD44*, *RUNX1*) expression distribution in the clusters indicated by the colour. **D** - Expression distribution visualised on the UMAP plot showing the location of arterial cells marked by *DLL4*, and hemogenic endothelium marked by *CD44* and *RUNX1*. **E** - Heatmap of the top 15 marker genes for each of the clusters. **F** - Contribution of the different libraries to the clusters showing that arterial cell cluster is overrepresented in the iSAM_AGM treated with DOX, compared to the other libraries. **G** - Expansion of the arterial population assessed by the membrane marker expression of DLL4+ following targets' activation, quantified by flow cytometry at day 8 of differentiation (Data are normalised on the iSAM_NT + DOX sample, n=5 independent differentiations, * p = 0.0417 paired t-test). **H** - Colony forming potential of the suspension progenitor cells derived from the two lines treated with or without DOX following OP9 coculture activation, data show the colony obtained for 104 CD34+ input equivalent (n=3 from independent differentiations * p<0.05, Tukey's two-way ANOVA).

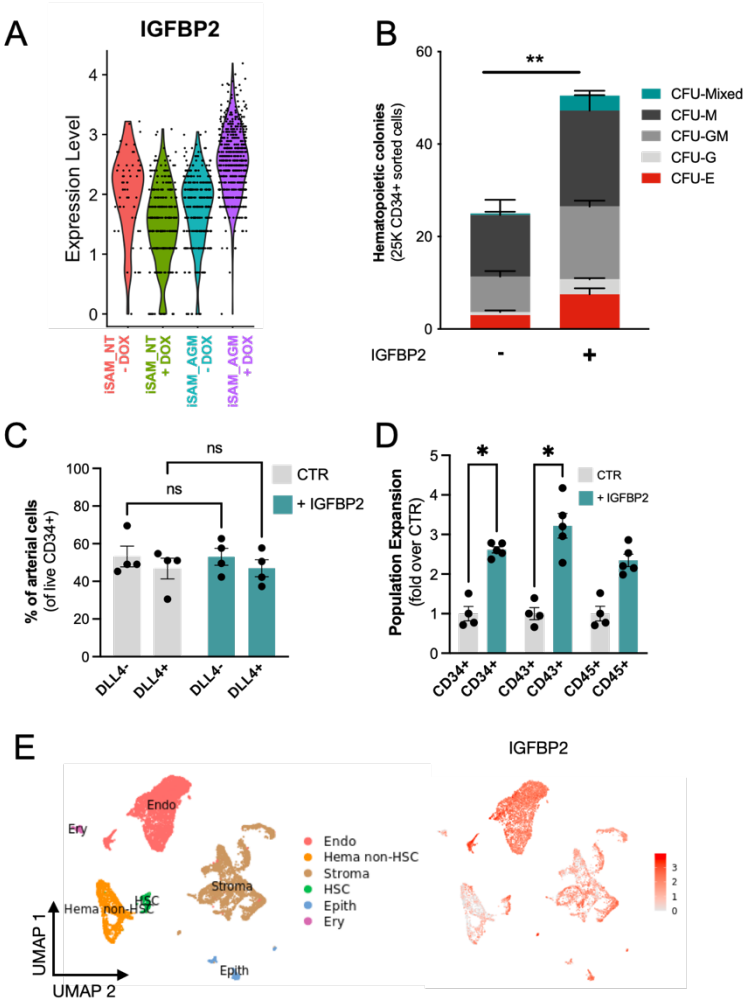

Figure 4 - IGFBP2 addition to the *in vitro* differentiation leads to a higher number of functional hematopoietic progenitor cells. A - Violin plot of *IGFBP2* expression profile in the arterial cells obtained from the different conditions, in the presence or absence of gRNAs and DOX. **B** -Number of hematopoietic colonies obtained after coculture on OP9 in presence or absence of IGFBP2 (n=3-4 from independent differentiations, ** p=0.0080, Sidak's Two way ANOVA). **C** – Percentage of DLL4^{+} arterial cells differentiation within the CD34^{+} compartment analysed by flow cytometry in day 8 EBs (n=4 from independent differentiations, Two-way Anova, ns =p>0.99). **D** – Expansion of hematopoietic progenitors analysed using markers' expression on suspension progenitors derived after coculture of CD34+ cells onto OP9 support (data are expressed as fold over the CTR in the absence of IGFBP2 (n=4 from independent differentiations, * p<0.02, Sidak's Two way ANOVA. **E** - Single-cell transcriptomic analysis of developing AGM collected from human embryos at Carnegie Stages14 and 15 enriched for CD31^{+} and CD34^{+} showing the *IGFBP2* expression profile *in vivo* in the AGM.

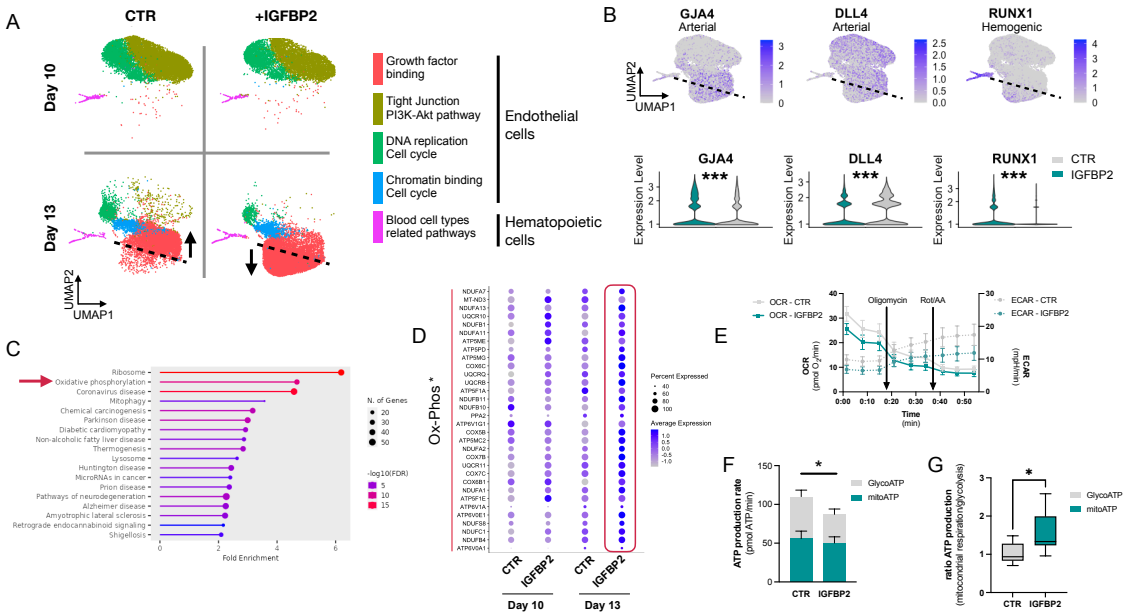

Figure 5 - IGFBP2 alters cell metabolism by inducing a reduction in glycolytic ATP production.

A – Clustering analysis of the single cell transcriptomic time course analysis of differentiating cells at day 10 and day 13 in the absence (CTR) or presence of IGFBP2. Arrows indicate the difference in the clustering due to the addition of IGFBP2 compared to control. **B** – Expression profile of arterial markers, *GJA4* and *DLL4*, and hemogenic marker *RUNX1* (top - the dashed line shows the location of the shift in gene expression of cells treated with IGFBP2) and their expression profile in the endothelial cells cluster marked by Growth factor binding in absence (CTR) and in presence of IGFBP2 (*GJA4* $p=1E^{-54}$, *DLL4* $p=1.2E^{-119}$, *RUNX1* $p=8.2E^{-163}$). **C –** KEGG enrichment analysis of the genes upregulated at day 13 upon IGFBP2 treatment. The arrow shows the ranking of the Oxidative Phosphorylation pathway. **D** - Dot Plot showing the expression profile of the genes coding for the enzyme of the Oxidative Phosphorylation pathway. **E** -Oxygen Consumption Rate (OCR) and Extracellular Acidification Rate (ECAR) profile in cells at day 13 of differentiation reporting mitochondrial respiration and glycolysis, respectively. **F** – ATP production rate divided by that deriving from glycolysis and from mitochondrial respiration, in cells treated with IGFBP2 and controls at day 13. **G** – Ratio of the ATP production between glycolysis and mitochondrial respiration in cells treated with IGFBP2 and controls at day 13.

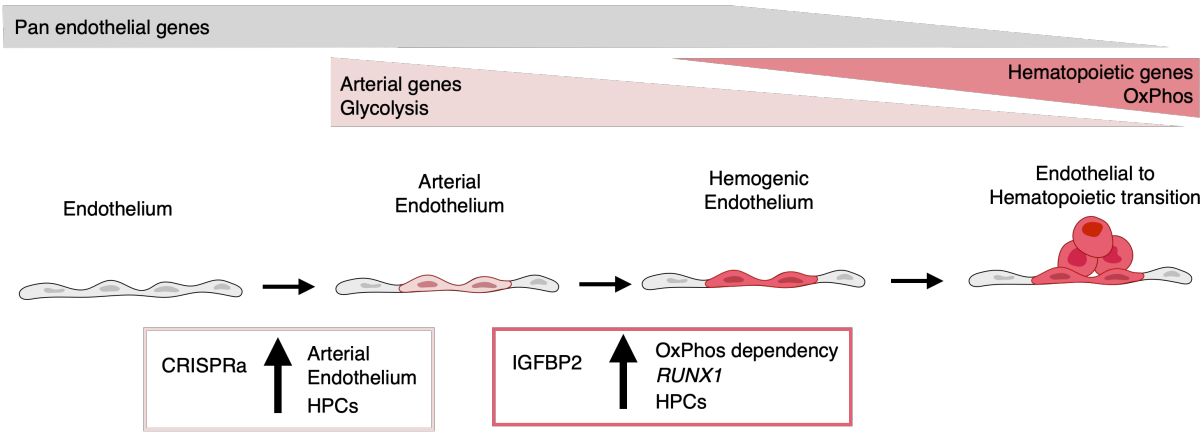

Figure 6 - Model summarising the results.
During development, some endothelial cells undergo arterialisation, as identified by their arterial genes' expression profile (e.g. *DLL4*). Arterial cells, characterised by high dependency on glycolysis, are expanded by our CRISPR activation approach, resulting in more blood production since these cells are the cell-of-origin of the hemogenic endothelium. Once arterial cells commit to hemogenic endothelium fate, they start to express hematopoietic genes (e.g. *RUNX1*); this process is enhanced by IGFBP2 via induction of both *RUNX1* expression and increased dependency on Oxidative Phosphorylation, known to be important for the progression of the endothelial to hematopoietic transition.

482 **Methods**

483

484 **Resource availability**

485 R code is available at https://github.com/afidanza/CRISPRa. Raw and processed data have
486 been deposited to ArrayExpress (E-MTAB-12748 – currently on hold). The AAVS1-iSAM are
487 available on Addgene (Addgene #211495) and the gRNA 2.1 plasmids (Addgene #211496)
488 (both currently on hold). Further information and requests for resources and reagents should
489 be directed to and will be fulfilled by the corresponding author.

490

491 **Pluripotent Stem Cells maintenance**

492 hPSCs were maintained *in vitro* in StemPro hESC SFM (Gibco) with bFGF (R&D) at 20
493 ng/ml. Wells were coated with Vitronectin (ThermoFisher Scientific) at least 1 hour before
494 plating and cells were passaged using the StemPro EZPassage tool (ThermoFisher
495 Scientific). Media change was performed every day and cells passaged every 3–4 days at a
496 ratio of 1:4.

497

498 **Transfection**

499 iPSCs SFCi55 and hESCs RUNX1-GFP were plated at 3×10^5 cells per a well of a 6 well
500 plate and reverse transfected with 2 µg of DNA using the Xfect Transfection reagent
501 (Clontech) and analyzed 2 days later.
502 HeLa cells were cultured in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12
503 (DMEM/F12) with Glutmax and 5% FCS (Gibco) and passaged every few days, at a ratio of
504 1:6. HEL were cultured in Iscove's Modified Dulbecco's Medium (IMDM) with 10% FCS
505 (Gibco) and passaged every few days, at a ratio of 1:4. 2×10^5 cells were plated, transfected
506 at 6–8 hours with 0.75 µg of DNA using Xfect Transfection reagent (Clontech) and then
507 analysed 2 days after.

508

509 **Immunocytochemistry**

510 Cells were fixed in 4% PFA in PBS at room temperature for 10', permeabilized in PBS-T
511 (0.4% Triton-X100) for 20' and blocked in PBS-T with 1% BSA and 3% goat serum for
512 1 hour. Primary antibodies were incubated in blocking solution over night at 4 °C (RUNX1
513 1:200 - ab92336, Abcam). Cells were then washed in PBS-T and incubated with secondary
514 antibodies for 1 hour at room temperature (donkey α-rabbit 1:200 - A-11008 - Thermo
515 Scientific). Cells were washed in PBS-T and counterstained with DAPI. Images were
516 generated using the Zeiss Observer microscope.

517

518 **Gene expression analysis**

519 Total RNA was purified using the RNAeasy Mini Kit (Qiagen) and cDNA synthesized from
520 500 ng of total RNA using the High Capacity cDNA synthesis Kit (Applied Biosystem). 2 ng of
521 cDNA were amplified per reaction and each reaction was performed in triplicate using the
522 LightCycler 384 (Roche) with SYBR Green Master Mix II (Roche). A melting curve was
523 performed and analyzed for each gene to ensure the specificity of the amplification. *β-*
524 *Actin* was used as reference genes to normalize the data [19].
525
526 **Pluripotent Stem Cells differentiation to hematopoietic progenitors**
527 hPSCs were differentiated in a xeno-free composition of SFD medium [14], BSA was
528 substituted with human serum albumin, HSA (Irvine-Scientific). Day 0 differentiation medium,
529 containing 10 ng/ml BMP4 was added to the colonies prior cutting. Cut colonies were
530 transferred to a Cell Repellent 6 wells Plates (Greniner) to form embryoid bodies and
531 cultured for two days. At day 2 media was changed and supplemented with 3 μM CHIR
532 (StemMacs). At day 3, EBs were transferred into fresh media supplemented with 5 ng/ml
533 bFGF and 15 ng/ml VEGF. At day 6 media was changed for final haematopoietic induction in
534 SFD medium supplemented with 5 ng/ml bFGF, 15 ng/ml VEGF, 30 ng/ml IL3, 10 ng/ml IL6,
535 5 ng/ml IL11, 50 ng/ml SCF, 2 U/ml EPO, 30 ng/ml TPO, 10 ng/ml FLT3L and 25 ng/ml
536 IGF1. From day 6 onward, cytokines were replaced every two days.
537
538 **CD34 isolation**
539 CD34+ cells were isolated using CD34 Magnetic Microbeads from Miltenyi Biotec, according
540 to their manufacturing protocol. Briefly, Embryoid bodies were dissociated using Accutase
541 (Life Techonologies) at 37°C for 30'. Cells were centrifuged and resuspended in 150 μl of
542 PBS + 0.5% BSA + 2mM EDTA with 50 μl Fcr blocker and 50 μl of magnetic anti-CD34 at
543 4°C for 30'. Cells were washed using the same buffer and transferred to pre-equilibrated
544 columns, washed three times and eluted. After centrifugation, cells were resuspend in SFD
545 media, counted and plated for OP9 coculture.
546
547 **OP9 coculture and colony assay**
548 OP9 cells were maintained in α-MEM supplemented with 20% serum (Gibco) and sodium
549 bicarbonate (Gibco) and passaged with Trypsin every 3-4 days. The day before the co-
550 culture, 45.000 OP9 cells were plated for each 12 well plates' well in SFD media. The day of
551 the co-culture the 20.000 iSAM cells or 25.000 SFCi55 or H9 were plated in each well and
552 culture in SFD media supplemented with 5 ng/ml bFGF, 15 ng/ml VEGF, 30 ng/ml IL3, 10
553 ng/ml IL6, 5 ng/ml IL11, 50 ng/ml SCF, 2 U/ml EPO, 30 ng/ml TPO, 10 ng/ml FLT3L and 25
554 ng/ml IGF1 and 100ng/ml IGFBP2. Cytokines were replaced twice during one week of

555 coculture. At the end of the coculture, cells were collected by Trypsin and half of the well
556 equivalent was plated in 2 ml of methylcellulose medium (Human enriched H4435, Stemcell
557 Technologies). Cells were incubated in the assay for 14 days and then scored.
558
559 **Laminin EHT culture**
560 Laminin EHT culture was performed as previously described[34]. Briefly, 24 well plates were
561 coated for at least 2 hours with recombinant human Laminin-521 (Thermo-Fisher). Following
562 CD34+ isolation at day 8, 400.000 CD34+ cells were seeded in each well of a pre-coated
563 24-well plate in SFD media supplemented with 5 ng/ml bFGF, 15 ng/ml VEGF, 30 ng/ml IL3,
564 10 ng/ml IL6, 5 ng/ml IL11, 50 ng/ml SCF, 2 U/ml EPO, 30 ng/ml TPO, 10 ng/ml FLT3L and
565 25 ng/ml IGF1 and 100ng/ml IGFBP2.
566
567 **Flow cytometry staining and cell sorting**
568 Embryoid bodies were dissociated using Accutase (Life Techonologies) at 37°C for 30'.
569 Cells were centrifuged and resuspended in PBS + 0.5% BSA + 2mM EDTA, counted and
570 stained at 10^5 cells for a single tube. Cells were stained with antibodies for 30' at room
571 temperature gently shaking. Flow cytometry data were collected using DIVA software (BD).
572 For the sorting experiments, the cells were prepared as above and stained at 10^7cells/ml in
573 presence of the specific antibodies. Sorting was performed using FACSAria Fusion (BD) and
574 cells were collected in PBS + 1% BSA. Data were analysed using FlowJo version 10.4.2.
575
576 **Flow cytometry antibodies**
577 For flow cytometry 10^5 cells per test were stained in 50 μl of staining solution with the
578 following antibodies: CD34 Percp-Efluor710 (4H11 eBioscience, 1:100), CD34 Pe (4H11
579 eBioscience, 1:200), CD43 APC (eBio84-3C1, 1:100), CD45 FITC (2D1 ebioscience, 1:100),
580 DLL4 Pe (MHD4-46 Biolegend, 1:200), CD41 PE (HIP8 Biolegend, 1:200), CD144 APC (16B1
581 eBioscience, 1:100), CD235a FITC (HIR2 BD Bioscience, 1:250).
582
583 **iSAM plasmid generation**
584 The iSAM plasmid was obtained by Gibson assembly of four fragments. The first fragment,
585 the backbone, was a DOX-inducible AAVS1 targeted plasmid expressing an E6-E7-IRES-
586 ZsGreen which was excised by BstBI and NdeI. The second fragment, one of the adapters,
587 was derived from the UniSAM plasmid that we previously generated (Addgene #99866) by
588 PCR with the following primer sets
589 FW_aggggacccggttcgagaaggggctcttcatcactagggccgctagctctagagagcgtcgaatt,

590 RV_ttcgggtcccaattgccgtcgtgctggcggctcttcccacctttctcttcttcttggggctcatggtggcc. The UniSAM
591 cassette was obtained also from the UniSAM plasmid via digestion with BsrGI and BsiWI.
592 Finally, the last fragment consisting of another adapter for the Gibson was custom
593 synthetised and contained overlapping sequences flanking a chicken b-globin insulator that
594 we inserted to prevent cross-activation of the EF1α-promoter and the TRE-GV promoter
595 driving the iSAM. Correct assembly was verified by Sanger sequencing.
596
597 **iSAM cell lines derivation**
598 The iSAM plasmid was used together with ZNFs specific for the AAVS1 locus to mediate
599 specific integration in SFCi55 human iPSCs line [14,20]. Briefly, 10 μg of AAVS1-iSAM with 2.5
600 μg of each ZNFs, left and right, using Xfect (Takara) according to the manufacturer protocol.
601 Cells were selected using Neomycin. Single clones were picked, amplified, and initially
602 screened by mCherry expression upon DOX addition. Clones that expressed the fluorescent
603 tag were screened for specific integration using PCR followed by Sanger sequencing for the
604 correctly integrated clones. 100 ng of genomic DNA was amplified using the EmeraldAmp®
605 MAX HS Takara and specific primer sets (Table 1). For the specific AAVS1 integration site,
606 Sigma_AAVS1 - CGG AAC TCT GCC CTC TAA CG and NeoR -GAT ATT GCT GAA GAG
607 CTT GGC GG were used with the PCR conditions of 95 °C for 7 min, 32 cycles of 95 °C for
608 15 s, 57 °C for 30 s, and 72 °C for 1 min, with the final elongation step at 72 °C for 7 min. For
609 the Wild type locus screening, Sigma_AAVS1 - CGG AAC TCT GCC CTC TAA CG and
610 AAVS1_EXT3_RV – ACA CCC AGA CCT GAC CCA AA were used with the PCR cycling
611 conditions of 95 °C for 7 min, 30 cycles of 95 °C for 15 s, 57 °C for 30 s, and 72 °C for 2 min,
612 with the final elongation step at 72 °C for 6 min.
613
614
615 **Capture sequencing addition to the gRNA backbone**
616 The Capture sequence 2 was added to the gRNA_Puro_Backbone (Addgene #73797) by
617 PCR. Briefly, the capture sequence was added before the termination signal of the gRNA
618 followed by a BamHI site using the following PCR primers:
619 gRNA_FW gagggcctatttcccatgattcct,
620 gRNA_Cap_RV aaaaaaggatccaaaaaaCCTTAGCCGCTAATAGGTGAGCgcaccgactcggtgcc.
621 The gRNA backobone was replaced from the original plasmid via NdeI and BamHI digestion,
622 followed by ligation of the PCR produced following the same digestion. Correct integration of
623 the insert was verified by Sanger sequencing.
624
625 **Target genes identification**

626 Candidate genes were identified by comparing *in vivo* hemogenic endothelium [15]
627 (GSE135202) and *in vitro* iPSCs-derived endothelial cells that we previously generated [14]
628 (E-MTAB-9295). Briefly, the two datasets were merged and normalised using the R
629 package Seurat. Specific markers for hemogenic endothelium were identified and
630 transcription factors were sorted based on their GO annotation. Within those genes we
631 filtered those detected in more than 50% of the *in vivo* hemogenic endothelium and
632 expressed in less than 25% of the *in vitro* derived endothelial cells. This pipeline identified 9
633 target genes *RUNX1T1*, *NR4A1*, *GATA2*, *SMAD7*, *ZNF124*, *SOX6*, *ZNF33A*, *NFAT5*, *TFDP.*
634
635 **AGM gRNAs library preparation**
636 sgRNA design was performed by selecting the top candidates for on-target and off-target
637 score. Between 5 and 7 guides per gene were designed for *RUNX1T1*, *NR4A1*, *GATA2*,
638 *SMAD7*, *ZNF124*, *SOX6*, *ZNF33A*, *NFAT5*, *TFDP2* using the CRISPRpick tool from the
639 Broad Institute (https://portals.broadinstitute.org/gppx/crispick/public) (Table S1). All the
640 guide variants were Golden Gate cloned with the gRNA 2.1 backbone according to the
641 established protocol [49]. The 49 plasmids were pooled together in an equimolar ratio and the
642 library prep was subsequently used to produce lentiviral particles with a second-generation
643 production system. Briefly, the psPAX2 packaging plasmid, pMD2.G envelope, and the AGM
644 vector library were co-transfected using polyethyleneimine (PEI) (Polysciences, Warrington,
645 PA, USA) as previously detailed [50], Lentiviral particles-containing supernatants were
646 harvested 48–72 h post-transfection, concentrated by ultracentrifugation and titered in
647 hiPSCs cells.
648
649 **iSAM_AGM and iSAM_NT cell line derivation**
650 The selected iSAM clone (3.13 internal coding) was infected with viral particles containing
651 either the AGM library or the non targeting gRNA (NT) at a MOI of 10. The iSAM cells were
652 plated the afternoon before at $7x10^6$ cells into a T125 in the presence of 10 μM Rock
653 Inhibitor (Merk) which was maintained until the day following the infection. Cells were
654 infected in presence of 8μg/ml of Polybrene (Merk). Puromycin selection was initiated 36
655 hours post-infection and maintained during their culture until the beginning of the
656 differentiation. Both lines were tested for integration of the gRNAs. Briefly, 100 ng of isolated
657 gDNA was amplified using the PrimeSTAR MAX PCR mix (Takara) using the primers
658 gRNA_screening_FW and gRNA_screening_RV (Table1). Purified amplicons were
659 subjected to Sanger sequencing.
660
661 **Single Cell RNA sequencing**

662 For the iSAM_AGM and iSAM_NT single cell RNA sequencing experiment, embryoid bodies
663 obtained from day 10 of differentiation were dissociated using Accutase (Life Technologies')
664 at 37°C for 30'. For the IGFBP2 experiment, day 10 and day 13 cells from the Laminin EHT
665 culture were detached from the adherent layer using Accutase (Life Technologies') at 37°C
666 for 5'. From both experiments, cells were centrifuged and resuspended in CD34-Pe staining
667 solution at a density of 10^7/ml. CD34+/live/single cells were FAC-sorted in PBS + 0.1 %
668 BSA. Cell viability was also confirmed by Trypan blue stain for an accurate count. Around
669 15000 cells per sample were loaded into the 10X Chromium Controller, and single cell
670 libraries were obtained using the Chromium single cell 3' Reagent Kits v3 (10XGenomics)
671 according to manufacturer protocol. The four libraries were indexed using SI PCR primers
672 with different i7 indexes to allow for demultiplexing of the sequencing data. RNA
673 concentration was obtained using Quibit RNA HS (Thermo-Fisher). Quality of the obtained
674 libraries were verified using LabChip GX (PerkinElmer). Libraries were sequenced using
675 NextSeq 2000 technology (Illumina) at 50.000 reads/cell. Data were aligned to GRCh38
676 using the Cell Ranger dedicated pipeline (10XGenomics). Data filtering, dimension
677 reduction, clustering analysis, differentially expressed genes and cell cycle analysis were
678 obtained using Seurat R package (version 4.1.0)[51]. Cells were subjected to QC and filtering
679 using both the number of genes (1000-7500) and the percentage of mitochondrial genes
680 detected (1-15%), resulting in 9025, 10942, 9468, 13073 cells respectively for the samples
681 iSAM_NT, iSAM_NT+DOX, iSAM_AGM and iSAM_AGM+DOX.Pseudotemporal ordering
682 was performed using Monocle 3 R package [52]. KEGG pathways was performed using
683 ShinyGo[53]. The gRNAs' expression matrix was used to select cells in which the expression
684 of the gRNAs was detected. Briefly, for the libraries derived from the iSAM_NT control
685 containing only the non-targeting gRNA, the filter was set for cells expressing one gRNA,
686 while for the iSAM_AGM libraries was set to more than one. The code is available on
687 GitHub at https://github.com/afidanza/CRISPRa, the raw data have been submitted to Array
688 Express (E-MTAB-12748), and the browsable processed data will be made available at the
689 time of publication on our website containing previous sequencing data at
690 https://lab.antonellafidanza.com.
691
692
693 **ATP production analysis**
694
695 At day 8 of differentiation, 20.000 CD34+ were plated for each precoated well of the
696 Seahorse XFp mini Cell Culture Plates (Agilent) precoated with rhLaminin-521 (Thermo-
697 Fisher). On day 13, the cells' ATP production rate was analysed with the Seahorse XF Real-

698 Time ATP Rate Assay kit (Agilent) according to manufacturer protocol, following
699 confirmation of comparable cell densities across replicates and conditions. Briefly, the media
700 was changed prior to the assay for the XF DMEM medium, pH 7.4, supplemented with 10
701 mM Seahorse XF glucose, 1mM Seahorse XF pyruvate and 2mM Seahorse XF Glutamine
702 (Agilent) and incubated for 45-60 minutes in a non-CO_2 incubator. Oligomycin and
703 Rotenone/AA solutions were prepared and added in the cartridge and finally loaded together
704 with the cells in the Seahorse XF Mini Analyzer using the dedicated software. Data were
705 collected at the end of the run, and the values of ATP production were calculated according
706 to the ATP Production Rate Calculation provided by Agilent. Briefly, OCR ATP was
707 calculated as OCR_{basal}-OCR_{oligo} averaged across the three reads for each well. Mitochondrial
708 ATP was calculated as OCR_{ATP} multiplied by the molecular oxygen consumption rate of 2
709 and by the P/O value of 2.75. For the glycolytic ATP production, we calculated the MitoPER
710 as the OCR_{basal}-OCR_{rot} times the CO_2 contribution Factor of 0.5. The PER was then
711 calculated as the ECAR times the Buffer Factor of 2.6, the volume of reaction of 2.28 μl and
712 the Kvol value of 1.1. The GlycoATP Production Rate was obtained by removing the
713 MitoPER from the PER, and the TotalATP was calculated by adding the GlycoATP and the
714 MytoATP. Mito/Glyco ratio was obtained by dividing their ATP production value.
715
716
717 **Cell cycle analysis**
718 DAPI staining and flow cytometry analysis were performed to verify the proliferation rate of
719 the cells. Briefly, cells were collected from the supernatant by aspiration using a pastette,
720 washed in PBS + 0.5% BSA + 2mM EDTA, counted and stained at 10^5 cells for a single
721 tube. Cells were stained 1:1 v:v with a solution of 1% NP40 and 5 μg/ml DAPI for 2 minutes
722 and acquired using the DIVA software (BD), and analysed using FlowJo version 10.4.2.
723
724
725 **Author contribution**
726
727 AF designed the study, performed experiments and bioinformatic analysis, wrote the paper
728 and led the research. PP, TV, FPL, HM performed experiments. AM, HT, MS provided
729 support to the experiments. NR performed bioinformatic analysis. LF and PM helped the
730 design of the study and the research. All authors provided essential feedback on the
731 experiments and the manuscript.
732
733 **Acknowledgment**

AF and LF acknowledge financial support from the Biotechnology and Biological Sciences Research Council; Grant S002219/1. AF was supported by a European Hematology Association Advanced Research Grant (EHA RAG 2021), and by the American Society for Hematology (Research Global Award). TV and AM were supported by PhD studentships from the Medical Research Council (Precision Medicine) and the College of Medicine and Veterinary Medicine, respectively. FPL was supported by an Erasmus+ Traineeship Program 2016/2017. PM acknowledges financial support from a PERIS program from the Catalan Government and a Retos collaboration project from the MINECO (RTC-2018-4603-1)

Bibliography

748 1. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by
749 the AGM region. *Cell* **86**, 897–906 (1996).
750 2. Palis, J., Robertson, S., Kennedy, M., Wall, C. & Keller, G. Development of erythroid
751 and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development*
752 **126**, 5073–84 (1999).
753 3. Patel, S. H. *et al.* Lifelong multilineage contribution by embryonic-born blood
754 progenitors. *Nature 2022 606:7915* **606**, 747–753 (2022).
755 4. Böiers, C. *et al.* Lymphomyeloid Contribution of an Immune-Restricted Progenitor
756 Emerging Prior to Definitive Hematopoietic Stem Cells. *Cell Stem Cell* **13**, 535–548
757 (2013).
758 5. Hoeffel, G. *et al.* C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give
759 rise to adult tissue-resident macrophages. *Immunity* **42**, 665–78 (2015).
760 6. Jaffredo, T., Gautier, R., Eichmann, A. & Dieterlen-Lièvre, F. Intraaortic hemopoietic
761 cells are derived from endothelial cells during ontogeny. *Development* **125**, 4575–83
762 (1998).
763 7. Zovein, A. C. *et al.* Fate Tracing Reveals the Endothelial Origin of Hematopoietic
764 Stem Cells. *Cell Stem Cell* **3**, 625–636 (2008).
765 8. Bertrand, J. Y. *et al.* Haematopoietic stem cells derive directly from aortic
766 endothelium during development. *Nature* **464**, 108–111 (2010).
767 9. Boisset, J. C. *et al.* In vivo imaging of haematopoietic cells emerging from the mouse
768 aortic endothelium. *Nature* **464**, 116–120 (2010).
769 10. Ottersbach, K. Endothelial-to-haematopoietic transition: an update on the process of
770 making blood. (2019) doi:10.1042/BST20180320.
771 11. Swiers, G. *et al.* Early dynamic fate changes in haemogenic endothelium characterized
772 at the single-cell level. *Nature Communications 2013 4:1* **4**, 1–10 (2013).
773 12. Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel
774 type of cell transition. *Nature* **464**, 112–115 (2010).
775 13. Eilken, H. M., Nishikawa, S. I. & Schroeder, T. Continuous single-cell imaging of
776 blood generation from haemogenic endothelium. *Nature* **457**, 896–900 (2009).
777 14. Fidanza, A. *et al.* Single cell analyses and machine learning define hematopoietic
778 progenitor and HSC-like cells derived from human PSCs. *Blood* **136**, 2893–2904
779 (2020).
780 15. Zeng, Y. *et al.* Tracing the first hematopoietic stem cell generation in human embryo
781 by single-cell RNA sequencing. *Cell Res* **29**, 881–894 (2019).
782 16. Ng, E. S. *et al.* Differentiation of human embryonic stem cells to HOXA+ hemogenic
783 vasculature that resembles the aorta-gonad-mesonephros. *Nat Biotechnol* **34**, 1168–
784 1179 (2016).
785 17. Calvanese, V. *et al.* Mapping human haematopoietic stem cells from haemogenic
786 endothelium to birth. *Nature 2022 604:7906* **604**, 534–540 (2022).
787 18. Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. & Keller, G. Wnt signaling
788 controls the specification of definitive and primitive hematopoiesis from human
789 pluripotent stem cells. *Nat Biotechnol* **32**, 554–561 (2014).
790 19. Fidanza, A. *et al.* An all-in-one UniSam vector system for efficient gene activation. *Sci*
791 *Rep* **7**, 6394 (2017).
792 20. Yang, C.-T. *et al.* Activation of KLF1 Enhances the Differentiation and Maturation of
793 Red Blood Cells from Human Pluripotent Stem Cells. *Stem Cells* **35**, 886–897 (2017).
794 21. Lopez-Yrigoyen, M. *et al.* A human iPSC line capable of differentiating into
795 functional macrophages expressing ZsGreen: A tool for the study and in vivo tracking

796 of therapeutic cells. *Philosophical Transactions of the Royal Society B: Biological*
797 *Sciences* **373**, (2018).
798 22. Lopez-Yrigoyen, M. *et al.* Genetic programming of macrophages generates an in vitro
799 model for the human erythroid island niche. *Nat Commun* **10**, 881 (2019).
800 23. Lopez-Yrigoyen, M. *et al.* A human iPSC line capable of differentiating into
801 functional macrophages expressing ZsGreen: A tool for the study and in vivo tracking
802 of therapeutic cells. *Philosophical Transactions of the Royal Society B: Biological*
803 *Sciences* **373**, (2018).
804 24. May, A. *et al.* Modelling the erythroblastic island niche of dyserythropoietic anaemia
805 type IV patients using induced pluripotent stem cells. *Front Cell Dev Biol* **11**, (2023).
806 25. Kang, S. J., Park, Y. Il, So, B. & Kang, H. G. Sodium butyrate efficiently converts
807 fully reprogrammed induced pluripotent stem cells from mouse partially
808 reprogrammed cells. *Cell Reprogram* **16**, 345–354 (2014).
809 26. Zhang, Z., Xiang, D. & Wu, W. S. Sodium Butyrate Facilitates Reprogramming by
810 Derepressing OCT4 Transactivity at the Promoter of Embryonic Stem Cell–Specific
811 miR-302/367 Cluster. *Cell Reprogram* **16**, 130 (2014).
812 27. Thambyrajah, R. *et al.* HDAC1 and HDAC2 Modulate TGF-β Signaling during
813 Endothelial-to-Hematopoietic Transition. *Stem Cell Reports* **10**, 1369–1383 (2018).
814 28. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA
815 capture and targeted sequencing. *Nat Biotechnol* (2020) doi:10.1038/s41587-020-
816 0470-y.
817 29. Ng, E. S. *et al.* Differentiation of human embryonic stem cells to HOXA+ hemogenic
818 vasculature that resembles the aorta-gonad-mesonephros. *Nat Biotechnol* **34**, 1168–
819 1179 (2016).
820 30. Sturgeon, C. M., Ditadi, A., Awong, G., Kennedy, M. & Keller, G. Wnt signaling
821 controls the specification of definitive and primitive hematopoiesis from human
822 pluripotent stem cells. *Nat Biotechnol* **32**, 554–561 (2014).
823 31. Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular
824 endothelium represent distinct lineages. *Nat Cell Biol* **17**, 580–591 (2015).
825 32. Ventura, T., Egan, E. J., Romanò, N. & Fidanza, A. In Vitro Model of Fetal Human
826 Vessel On-chip to Study Developmental Mechanobiology. *J Vis Exp* **2023**, (2023).
827 33. Fidanza, A. *et al.* Single cell analyses and machine learning define hematopoietic
828 progenitor and HSC-like cells derived from human PSCs. *Blood* **136**, 2893–2904
829 (2020).
830 34. Ventura, T., Egan, E. J., Romanò, N. & Fidanza, A. In Vitro Model of Fetal Human
831 Vessel On-chip to Study Developmental Mechanobiology. *J Vis Exp* **2023**, (2023).
832 35. Azzoni, E. *et al.* The onset of circulation triggers a metabolic switch required for
833 endothelial to hematopoietic transition. *Cell Rep* **37**, (2021).
834 36. Oburoglu, L. *et al.* Pyruvate metabolism guides definitive lineage specification during
835 hematopoietic emergence. *EMBO Rep* **23**, (2022).
836 37. Castaño, J. *et al.* GATA2 Promotes Hematopoietic Development and Represses
837 Cardiac Differentiation of Human Mesoderm. *Stem Cell Reports* **13**, 515–529 (2019).
838 38. de Pater, E. *et al.* Gata2 is required for HSC generation and survival. *Journal of*
839 *Experimental Medicine* **210**, 2843–2850 (2013).
840 39. Ling, K. W. *et al.* GATA-2 Plays Two Functionally Distinct Roles during the
841 Ontogeny of Hematopoietic Stem Cells. *J Exp Med* **200**, 871 (2004).
842 40. McGarvey, A. C. *et al.* A molecular roadmap of the AGM region reveals BMPER as a
843 novel regulator of HSC maturation. *J Exp Med* **214**, 3731–3751 (2017).
844 41. McGrath, K. E. *et al.* A transient definitive erythroid lineage with unique regulation of
845 the β-globin locus in the mammalian embryo. *Blood* **117**, 4600 (2011).

42. Rejeski, K., Duque-Afonso, J. & Lübbert, M. AML1/ETO and its function as a regulator of gene transcription via epigenetic mechanisms. *Oncogene* **40**, 5665–5676 (2021).
43. Huynh, H. D. *et al.* IGF binding protein 2 supports the survival and cycling of hematopoietic stem cells. *Blood* **118**, 3236 (2011).
44. Crosse, E. I. *et al.* Multi-layered Spatial Transcriptomics Identify Secretory Factors Promoting Human Hematopoietic Stem Cell Development. *Cell Stem Cell* **27**, 822 (2020).
45. Hadland, B. *et al.* Engineering a niche supporting hematopoietic stem cell development using integrated single-cell transcriptomics. *Nature Communications 2022 13:1* **13**, 1–17 (2022).
46. Hadland, B. K. *et al.* Endothelium and NOTCH specify and amplify aorta-gonad-mesonephros-derived hematopoietic stem cells. *J Clin Invest* **125**, 2032–2045 (2015).
47. Sandler, V. M. *et al.* Reprogramming human endothelial cells to haematopoietic cells requires vascular induction. *Nature* **511**, 312–318 (2014).
48. Azzoni, E. *et al.* The onset of circulation triggers a metabolic switch required for endothelial to hematopoietic transition. *Cell Rep* **37**, (2021).
49. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
50. Petazzi, P. *et al.* Robustness of Catalytically Dead Cas9 Activators in Human Pluripotent and Mesenchymal Stem Cells. *Mol Ther Nucleic Acids* **20**, 196–204 (2020).
51. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
52. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature 2019 566:7745* **566**, 496–502 (2019).
53. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).