Dissecting infant leukemia developmental origins with a hemogenic gastruloid model

Denise Ragusa[1,2], Chun-Wai Suen[3*], Gabriel Torregrosa Cortés[4*], Liza Dijkhuis[1,9], Connor Byrne[1], Giulia-Andreea Ionescu[1,10], Joana Cerveira[5], Kamil R. Kranc[6], Anna Bigas[7,8], Jordi Garcia-Ojalvo[4], Alfonso Martinez Arias[4], Cristina Pina[1,2¶]*

[1]College of Health, Medicine and Life Sciences, [2]Centre for Genome Engineering and Maintenance, Brunel University London, UK; [3]Department of Genetics, University of Cambridge, UK; [4]Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Spain; [5]Department of Pathology, University of Cambridge, UK; [6]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, UK; [7]Program in Cancer Research, Institut Hospital del Mar d'Investigacions Mèdiques, CIBERONC; [8]Josep Carreras Leukemia Research Institute, Barcelona, Spain

[9]Current address: Sanquin Research Laboratory, Department of Hematopoiesis, Amsterdam, The Netherlands

[10]Current address: Centre for Immunology, Blizard Institute, Queen Mary University of London, UK

*These authors contributed equally to this work

¶Lead contact. Correspondence: cristina.pina@brunel.ac.uk. Brunel University London, Kingston Lane, Uxbridge UB8 3PH, United Kingdom.

ABSTRACT

Current *in vitro* models of developmental blood formation lack spatiotemporal coherence and weakly replicate the hematopoietic microenvironment. Developmentally-appropriate models can enhance understanding of infant acute myeloid leukemia (infAML), which putatively originates *in utero* and has 50% age-unique genetic events, suggesting unique biology. The commonest genetic abnormality unique to infants involves homeobox gene *MNX1*, whose leukemogenic mechanisms remain unknown. Recently, 3D self-organising embryonic stem cell (SC)-based gastruloids have shown promise in recapitulating embryonic events with time/space precision. Herein, we report a hemogenic gastruloid (haemGx) system that captures multi-wave blood formation, progenitor specification from hemogenic endothelium (HE), and approximates generation of hematopoietic SC precursors. Enforced *MNX1*

expression in haemGx promotes HE formation, perturbs endothelial-to-hemogenic transition, and critically achieves transformation, generating myeloid colonies which display *MNX1* AML signatures. By combining functional assays with single-cell transcriptomics, we establish the haemGx as a new model of normal and leukemic embryonic hematopoiesis amenable to mechanistic exploration.

KEYWORDS

Hematopoiesis; Developmental hematopoiesis; Leukemia; Infant Leukemia; Acute Myeloid Leukemia; MNX1; t(7;12); gastruloid; organoid; single-cell RNA-sequencing

INTRODUCTION

The development of organoid systems revolutionized fundamental and translational biology in the last decade (Anonymous, 2018). Organoids are three-dimensional (3D) cultures of primary, or primary-derived cells, which self-organize under defined culture conditions, to recapitulate key structural features of the tissue from which they originate (Clevers, H., 2016; Kim, J. et al., 2020; Lancaster and Knoblich, 2014). In recapitulating structure, organoids recreate elements of their organismal niche, and faithfully establish some of the regulatory physiology that maintains the tissues of origin in their native, *in vivo* environment (Huch et al., 2017). Organoids have been established from a multitude of tissues, primarily neuronal and epithelial (Barker et al., 2010; Gotoh et al., 2014; Greggio et al., 2014; Huch, Dorrell et al., 2013; Huch, Bonfanti et al., 2013; Lancaster et al., 2013; Sato et al., 2009). They are routinely used as 'avatars' of their tissues of origin to test the consequences of genetic perturbation or drug treatment, and can be superior to classical animal pre-clinical models in anticipating therapeutic effect or toxicity (Clevers, Hans C., 2019).

Another category of organoids uses embryonic or pluripotent stem cells to recapitulate developmental processes in a 3D *in vitro* or *ex vivo* space (van den Brink, S C and van Oudenaarden, 2021). Blastoid-type organoids focus on the initial zygotic divisions and can progress through

54 cleavage-like and blastula-like formation (Rivron et al., 2018; Yu et al., 2021). Gastruloids, on the
55 other hand, display elongation and symmetry breaking to mimic gastrulation and early tissue
56 territorialisation for primordial organogenesis (Arias et al., 2022). Multiple variations of gastruloid
57 protocols have captured mesendodermal specification (Turner et al., 2017; van den Brink, S C et al.,
58 2014), posterior and anterior neurulation (Girgin et al., 2021), somitogenesis (van den Brink, S C et
59 al., 2020), as well as fetal heart organogenesis (Rossi et al., 2021) with varying efficiencies. Tissue-
60 specific topography of molecular programs is broadly maintained (Beccari et al., 2018; Moris, Naomi
61 et al., 2020), making gastruloid models superior to embryoid bodies in capturing embryonic
62 organisation *in vitro*.

63 Developmental haematopoiesis progresses through 3 waves of cell specification which produce
64 blood cells in reverse hierarchy, with early emergence of unipotent cells and late stem cell
65 specification (Costa et al., 2012; Dzierzak and Bigas, 2018; Lacaud and Kouskoff, 2017; Medvinsky,
66 Alexander et al., 2011). Embryonic red blood cells and macrophages are initially produced in the
67 yolk sac of the mouse embryo at E7.5 and are generated from angioblasts (Lacaud and Kouskoff,
68 2017). Specification of these early blood cell types does not rely on bilineage or multilineage
69 progenitors, and the red blood cells produced are nucleated and express embryonic globins
70 (Kingsley et al., 2006). Around E8.25, a second, pre-definitive wave of blood production generates
71 erythro-myeloid-megakaryocytic progenitors (EMPs), which eventually differentiate into enucleated
72 erythrocytes, granulocytes, monocytes and megakaryocytes / platelets (McGrath et al., 2015).
73 EMPs, like the subsequent intra-embryonic wave of blood production, are specified from a
74 specialised endothelium – haemogenic endothelium (HE) (Marcelo et al., 2013) – through columnar
75 remodelling and intra-luminal budding of haematopoietic cells, a process known as endothelial-
76 haematopoietic transition (EHT) (Lacaud and Kouskoff, 2017). The third wave of blood production is
77 also HE-based and occurs in the ventral wall of the dorsal aorta and adjacent vessels between E9.5-
78 E11.5. It produces myelo-lymphoid (MLP) and multipotent progenitors (MPP) (Zhu et al., 2020), and
79 a small number of haematopoietic stem cells (HSC) (Medvinsky, A. and Dzierzak, 1996; Medvinsky,
80 A. L. et al., 1993), which expand in the fetal liver and eventually migrate to the bone marrow, where

81 they mature to sustain blood production postnatally and throughout adult life (Hall et al., 2022). EHT
82 generates characteristic clusters of cells, which are a hallmark of definitive and intra-embryonic
83 hematopoiesis (Lacaud and Kouskoff, 2017; Medvinsky et al., 2011).

84 *In vitro* specification of hematopoiesis from embryonic and induced pluripotent stem cells under
85 specified growth factor and stromal co-culture conditions efficiently generate granulocytic-monocytic
86 and erythroid cells, and to a lesser extent, lymphocytes (Sroczynska et al., 2009). However, they
87 cannot produce self-renewing, re-populating HSCs in the absence of genetic manipulation, and even
88 in this case, efficiency and reproducibility are variable (Lis et al., 2017; Sugimura et al., 2017).

89 *In vitro* recapitulation of developmental hematopoiesis has the potential to shed light on the biology
90 of those hematological malignancies that are initiated *in utero* and potentially transform an embryonic
91 cell stage and/or require the embryonic environment for transformation (Cazzola et al., 2021).
92 Indeed, by using an iPS-based model, it has been possible to identify the transient embryonic cell of
93 origin of the most common pediatric malignancy, the pro-B acute lymphoblastic leukemia (ALL)
94 associated with t(11;21) translocation, which generates the *ETV6-RUNX1* fusion (Boiers et al.,
95 2018). ALL is the most common hematological malignancy in children and young adults, while Acute
96 Myeloid Leukemia (AML) dominates in elderly individuals (Britten et al., 2019). However, in the first
97 year of life, AML is at least as frequent as ALL, and is characterized by a distinct set of chromosomal
98 abnormalities, 50% of which are exclusive to this age group (Balgobind et al., 2011; Fornerod et al.,
99 2021). The most common of these is the translocation t(7;12)(q36;p13), a deadly form of AML
100 molecularly characterized by ectopic activation of the *MNX1* gene at 7q36 (Espersen et al., 2018).
101 Its overexpression does not result in leukemic transformation of neonatal cord blood cells or adult
102 mouse bone marrow, but blocks erythroid differentiation and results in cellular senescence
103 (Ingenhag et al., 2019; Waraky et al., 2022; Wildenhain et al., 2010; Wildenhain et al., 2012).
104 Engineering of the t(7;12) translocation in human iPS cells captures some of the transcriptional
105 characteristics of t(7;12) patients and enhances erythroid and myeloid differentiation, but has not
106 been shown to result in transformation (Nilsson et al., 2022). Interestingly, the translocation
107 significantly depletes megakaryocytic signatures in iPS cells, suggesting that this model may not

108 fully capture the spectrum of *MNX1*-rearrangement leukemias, which are mostly undifferentiated or
109 poorly differentiated AML (FAB M0-M2, 70%), but include 15% of megakaryoblastic leukemia
110 (Espersen et al., 2018; Taketani et al., 2008).

111 In this study, we adapt the three-dimensional gastruloid model of mammalian development to
112 capture multi-wave establishment of blood formation in the embryo. Through cytokine-driven
113 maturation of self-organising gastruloids over a 216 hour-period, we observe sequential specification
114 of endothelium, HE, erythro-myeloid progenitor (EMP) and myelo-lymphoid progenitor (MLP)
115 programs and recapitulate the EHT topography with generation of hematopoietic clusters.
116 Interestingly, the gastruloid model can be transformed by introduction of *MNX1*, with sustained serial
117 re-plating of colony-forming cells. By systematically contrasting the gene expression profile of *MNX1*-
118 overexpressing hemogenic gastruloids and MNX1-rearranged leukemias with single-cell signatures
119 obtained from hemogenic gastruloid differentiation, we position the cells targeted by *MNX1* at the
120 HE-to-EMP transition and provide a mechanistic explanation for the strict developmental association
121 of MNX1 with infant AML.

122

123 **RESULTS**

124 **The hemogenic gastruloid protocol captures cellularity and topography of developmental**
125 **blood formation**

126 The original 120h-gastruloid protocol (van den Brink, S C et al., 2014) matches molecular and
127 organizational aspects of embryo development up to day 8 (E8.0) and captures incipient endothelial
128 and erythroid-biased transcriptional signatures which are expected to correspond to yolk-sac based
129 EMP production (Beccari et al., 2018). More recent adaptations of the protocol have used
130 mechanical agitation, ultra-low adherence plates, and/or matrices to extend the gastruloid life to
131 168h (Rossi et al., 2021; van den Brink, S C et al., 2020), or E9.5-10.0, and recapitulate later
132 developmental events such as cardiac specification and somitogenesis, both of which are critically
133 associated with maturation of the intra-embryonic hematopoietic system. We used ultralow
134 adherence multi-well plates (see STAR Methods) and sought to promote hemato-endothelial

135 specification by combining the axes-inducing WNT pulse with the BMP pathway inducer activin (Fig.
136 1A). The addition of activin was required for activation of the endothelial *Vegfr2/Kdr/Flk-1* locus (Fig.
137 S1A), a step which we consistently noted as critical for the subsequent detection of hematopoietic
138 cells. We extended the culture beyond 168h to 216h (Fig. 1A) to attempt to capture HSC-producing
139 AGM hematopoiesis at E10.5-E11.0. We used Flk1-GFP mouse ES cells (Jakobsson et al., 2010)
140 to visually monitor endothelial specification (Fig. 1B) and added VEGF and basic FGF (FGF-2) after
141 the activin pulse to promote development of HE (Sroczynska et al., 2009). Based on paracrine
142 signals critical to AGM maturation (Souilhol et al., 2016), we tested the time-dependent effect of Shh
143 in the cultures and observed that its addition in a 24h-pulse at 144h only, increased the fraction of
144 endothelial VE-Cadherin (CD144)$^+$ and candidate VE-Cadherin$^+$CD45$^+$ pre-HSC (Rybtsov, S. et al.,
145 2011; Taoudi et al., 2008) at 192h (Fig. S1B-C), suggestive of a recapitulation of the *in vivo* effect
146 on pre-HSC specification from HE. Addition of SCF, Flt3-ligand and TPO in the last 48h of the
147 gastruloid culture increased the fraction of hematopoietic cells at end-point, as measured by %CD45$^+$
148 cells (Fig. S1D). The presence of FGF-2 in the same period did not affect the final %CD45$^+$ cells
149 (Fig. S1E), and we omitted the cytokine from 168h onwards to simplify the protocol. We typically
150 cultured *Flk1-GFP* mES cells in serum with leukemia inhibiting factor (LIF) and pre-treated the cells
151 in '2i' (GSK3b inhibitor – Chiron – and MEK inhibitor – PD) + LIF prior to gastruloid assembly (Fig.
152 1A). However, the pre-treatment step may be dispensable if the mES cultures have a compact
153 pluripotent morphology, with minimum cell differentiation.

154 We used flow cytometry to monitor the timing of emergence of phenotypic endothelium, HE,
155 hemogenic progenitors, and candidate pre-HSC/HSC on the basis of Flk1-GFP (Fehling et al., 2003),
156 c-Kit (Marcelo et al., 2013), CD41 (Mikkola et al., 2003) and CD45 (Rybtsov et al., 2011), respectively
157 (Fig. 1C). The first endothelial cells were detectable at 96h (Fig. 1C), with c-Kit$^+$ HE first apparent at
158 120h (Fig. 1C). A transient wave of CD41$^+$ cells followed at 144h (Fig. 1C and S1F), which also
159 included double CD41$^+$CD43$^+$ cells (Fig. S1G-H), compatible with early pre-HSC (Rybtsov, Stanislav
160 et al., 2014). This was followed by the emergence of CD45$^+$ cells at 192h, which was consolidated
161 at 216h (Fig. 1C and S1I). We confirmed that sequential emergence of CD41$^+$ and CD45$^+$ cells could

162 be observed in the more widely used E14TG2a (E14) mES cells (Hooper et al., 1987) (Fig. S2A-C),
163 with end-point gastruloids initiated with mES cells from different genetic backgrounds generating
164 similar levels of CD45$^+$ cells (Fig. S2D). From 120h, gastruloids contained cells with hematopoietic
165 progenitor potential, evidenced in colony-forming cell (CFC) assays (Fig. 1D). Erythro-myeloid
166 progenitor numbers peaked at 144-168h, with a downward trend at the 2 latest timepoints, which
167 also showed a bias towards myeloid colonies (Fig. 1D). Notably, CD45$^+$ cells at 216h (Fig. 1E and
168 S2E) were observed in small clusters budding from a Flk1-GFP / CD31 endothelium reminiscent of
169 hematopoietic emergence in the dorsal aorta. Together with the ordered emergence of HE, CD41$^+$
170 progenitors, and candidate CD45$^+$ pre-HSC, the cluster-like arrangement of hematopoietic cells
171 configures the hemogenic gastruloid as a faithful *in vivo* model of developmental hematopoiesis
172 amenable to cellular and molecular dissection of blood cell specification, including the role of the
173 hematopoietic niche.

174

175 **Hemogenic gastruloid single-cell trajectories reveal pre-definitive and definitive waves of**
176 **hematopoiesis and support generation of pre-HSC**

177 In order to characterize the extent and progression of developmental hematopoiesis in the
178 hemogenic gastruloid model, we performed a single-cell RNA-sequencing (scRNA-seq) time-course
179 analysis of gastruloid cells in reference to the mouse embryo. As summarized in Fig. 2A, we sorted
180 cells from 2 independent gastruloid cultures at 120, 144, 168, 192 and 216h and profiled a total of
181 846 cells using the Smart-Seq2 protocol (Picelli et al., 2014). In line with the flow cytometric
182 phenotyping, we sorted c-Kit$^+$ cells at 144 and 192h to capture endothelial and HE cells at 2 distinct
183 points of blood production, as well as 144h-enriched CD41$^+$ cells, and the CD45$^+$ cells emergent at
184 192 and 216h. We also profiled live single cells obtained at the different timepoints without selection
185 on hematopoietic markers, with the goal of understanding the microenvironment in which gastruloid
186 hematopoiesis is specified. Library preparation and sequencing generated an average of 120000
187 reads/cell, which were mapped to an average of 4000 genes/cell, with almost no cells showing signs
188 of stress or dying as seen by the mitochondrial DNA fraction (Fig. S3A). Read and gene counts were

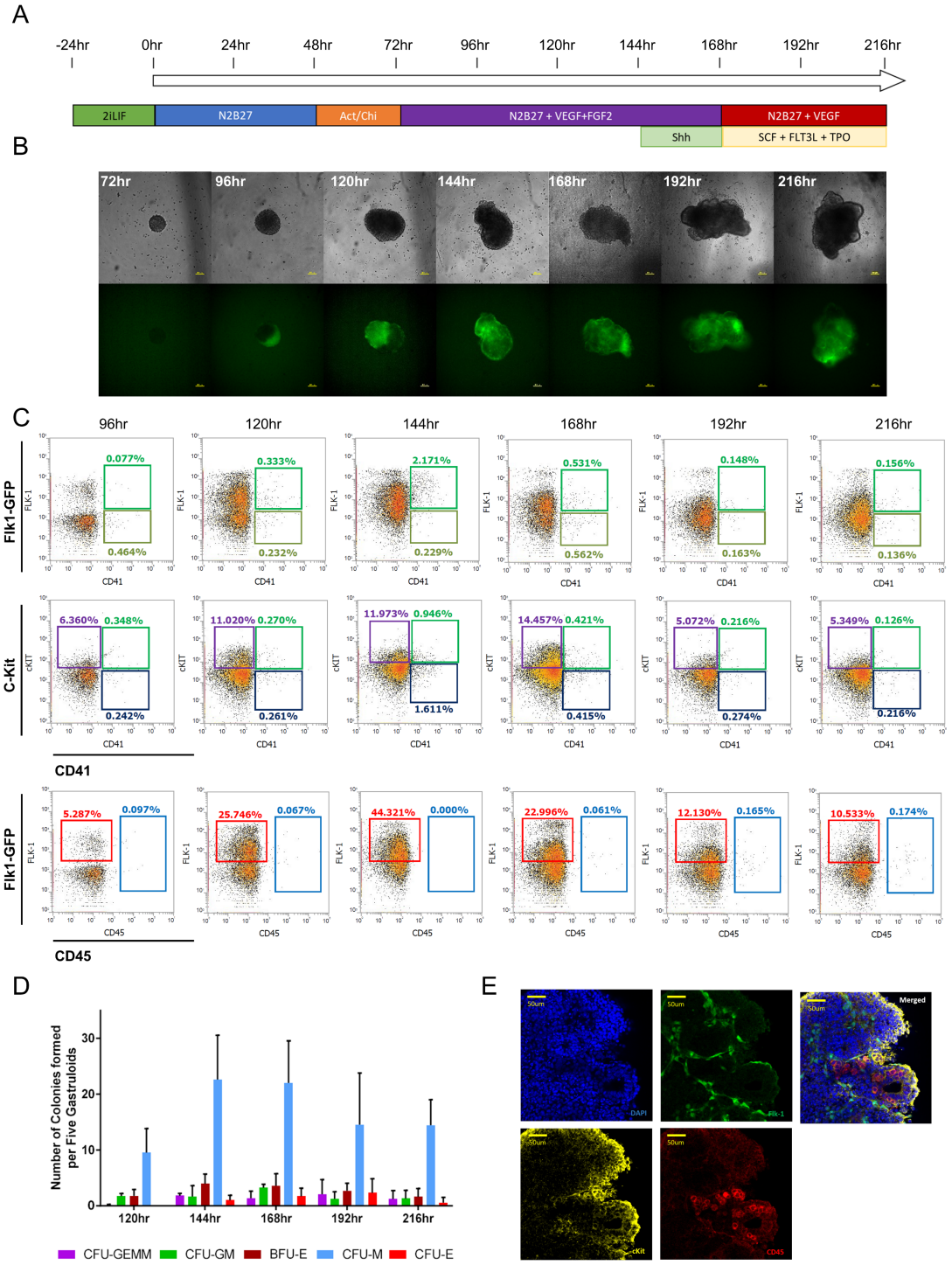


Figure 1 – Hemogenic gastruloids produced from mES cells promote hemato-endothelial specification with spatio-temporally accurate ontogeny. (A) Timeline of mES cells assembly and culture into gastruloids over a 216h time period with the addition of appropriate factors for the promotion of hemato-endothelial specification. **(B)** Imaging of hemogenic gastruloids over time from 72h to 216h at 10x magnification, showing the assembly and growth of the 3D structures and the polarization of the Flk-1-GFP marker from 96h; scale bar: 100μm **(C)** Flow cytometry analysis of hemogenic gastruloids for expression of c-Kit, CD41, and CD45 markers from 96h to 216h, assessing the emergence of c-Kit-endothelium, hemogenic progenitors, and pre-HSC/HSC at each time point. **(D)** Colony-forming unit (CFU) assay of disassembled gastruloids assessing the ability to form hematopoietic colonies in multipotential methylcellulose-based medium. GEMM: granulocyte-erythroid-monocyte-megakaryocyte; GM: granulocyte-monocyte; M: monocyte; E: erythroid; BFU-E: burst-forming unit erythroid. CFU frequency of 5 gastruloids, n=3, mean±SD **(E)** Immunostaining of whole individual gastruloids at 216h showing the localized expression of Flk-1 (green), c-Kit (yellow), and CD45 (red) and DAPI (blue) nuclear staining, in a topologically accurate configuration; scale bar: 50μm.

202 similar between biological replicates, cell types and at different time points, with the sole exception
203 of the 120h unsorted sample which was sequenced to same depth (similar average read count), but
204 mapped to twice the number of genes (Fig. S3B), potentially reflecting multi-gene program priming
205 at the onset of hemogenic specification. We selected highly varying genes before principal
206 component analysis (PCA) dimensionality reduction and retained the most relevant dimensions. In
207 the PCA reduced space we constructed a KNN graph and used uniform manifold approximation and
208 projection (UMAP) to visualize the data on 2-dimensions. We looked for cell communities using
209 Leiden clustering on the graph (Fig. 2B). We made use of the transcriptional identity of the cells to
210 describe chronological and biological aspects of multi-step hemogenic specification. We identified
211 12 cell clusters of which 2 (clusters 0 and 5) almost exactly mapped to c-Kit$^+$ cells, 2 (clusters 1 and
212 8) contained CD45$^+$ sorted cells, and cluster 4 uniquely captured the CD41$^+$ population of cells
213 observed at 144h. Although some unsorted cells could be observed overlapping with the hemogenic
214 cell clusters (Fig. 2B), most cells occupied different transcriptional spaces, in line with the relatively
215 low frequency of hemogenic cells, particularly at the later time points (Fig. 1C).

216 To explore the tissue or lineage affiliation of the different clusters, we performed differential gene
217 expression between each cluster and the remainder cells using Wilcoxon ranking test, and
218 established classifier gene lists for each cluster (Supplemental File S1). The gene lists were
219 compared with the PanglaoDB (Franzen et al., 2019) (Fig. 2C) and Descartes (Cao et al., 2020) (Fig.
220 S3C) repositories of scRNA-seq expression profiles through the EnrichR gene set enrichment
221 analysis tool suite (https://maayanlab.cloud/Enrichr/) (Chen, E. Y. et al., 2013; Kuleshov et al., 2016)
222 to identify enriched cell type representation within the clusters (Supplemental File S2). Clusters 5
223 and 0, which were mostly populated by c-Kit$^+$ cells at 144 and 192h, respectively (Fig. 2B and S3D)
224 had widespread expression of *Cdh5* (VE-cadherin), with some cells also expressing *Gata2* or *Runx1*,
225 compatible with hematopoietic specification from HE (Fig. 2D) (Chen, M. J. et al., 2009; de Pater et
226 al., 2013). Accordingly, their gene expression signatures captured the program of endothelial cells,
227 including those of the aorta, with some evidence of HSC-enriched genes (Fig. 2C). The cells
228 configure distinct clusters at the individual time points (Fig. S3D: 144h cluster 1; 192h cluster 5) (also

A

	120H	144H	168H	192H	216H
all	P1	P2	P5	P6	
Kit+ (Sca1+)		P3		P7	
CD41+		P4			
CD45+				P8	P9

(P = single-cell deposited plate for Smart-Seq2 scRNAseq analysis))

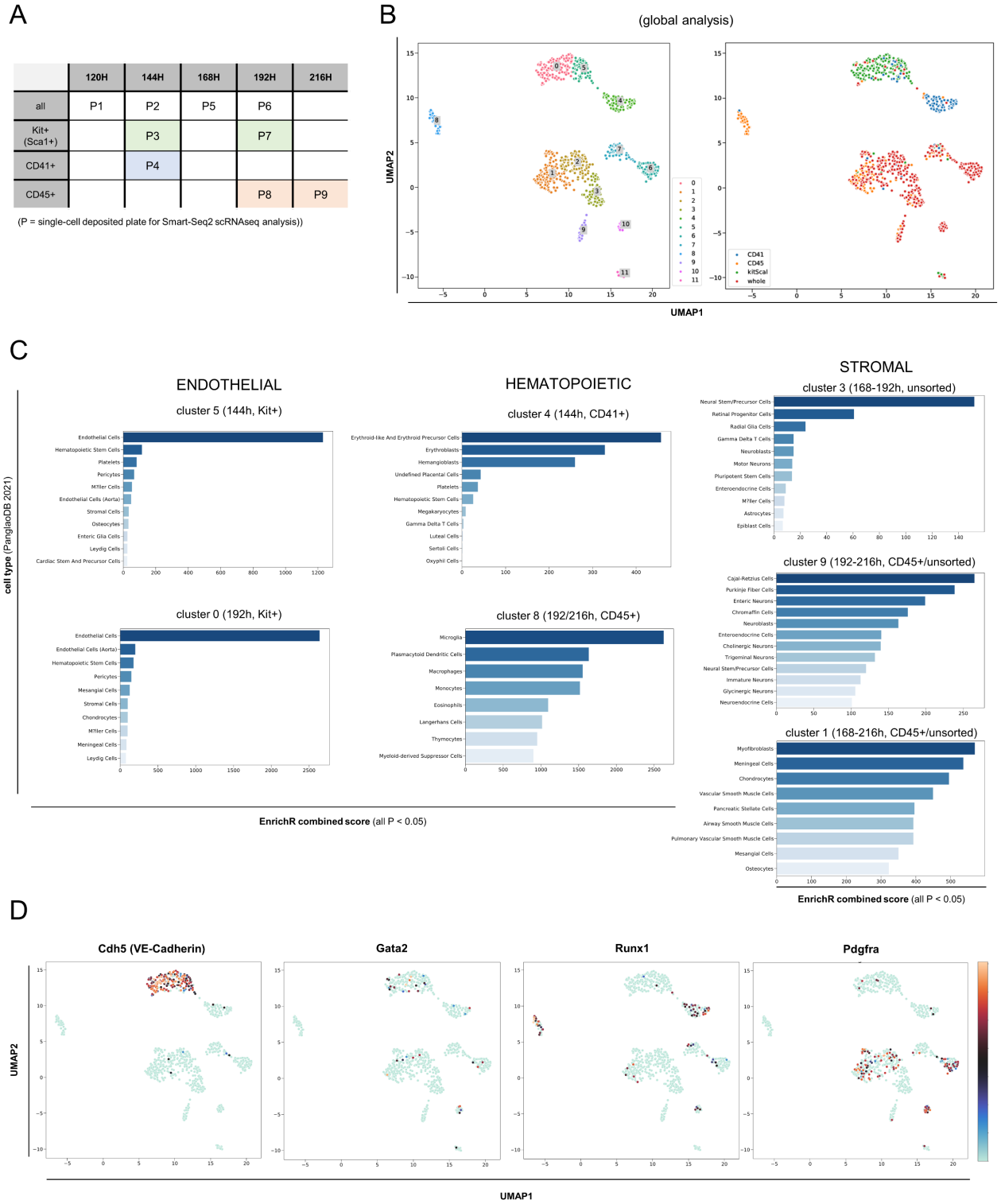


Figure 2 – Time-resolved analysis of scRNA-seq of hemogenic gastruloids captures successive waves of hematopoietic specification. (A) Summary of plating strategy for scRNA-seq analysis of gastruloids at 120h, 144h, 168h, 192h, and 216h without selection of surface markers ('all') or on the basis of expression of c-Kit/ScaI (green shading), CD41 (blue), or CD45 (orange); P = plate. **(B)** UMAP projection of all sequenced cells colored by annotated clusters (left) and by positivity to sorting markers c-Kit/ScaI, CD41, and CD45, or mapping to the unselected sort ('whole' corresponding to 'all' in panel A). **(C)** Cell type enrichment analysis of cluster classifier genes, extracted by differential expression in comparison to all other clusters, using the PanglaoDB and Descartes repositories. The statistical power of representation of individual cell types is expressed in EnrichR combined score with a p-value threshold of < 0.05. **(D)** Expression of key lineage markers *Cdh5* (VE-Cadherin), *Gata2*, *Runx1*, and *Pdgfra* in individual cells projected onto UMAP plots of global analysis shown in **A**. Colour scale represents Z-score gene expression from 0 (blue, bottom) to maximum (orange, top).

240 Supplemental File S1), and occupy minimally overlapping transcriptional spaces, suggesting
241 different waves of HE specification or maturation. Compatible with wave-like specification of
242 hematopoietic cell types from HE, 144h-enriched HE cluster 5 was immediately adjacent to CD41^{+}-
243 enriched (Fig. 2B) *Runx1*$^{+}$*Cdh5*$^{-}$ (Fig. 2D) cluster 4, which exhibits an erythroid/megakaryocytic cell
244 signature suggestive of pre-definitive EMPs (Supplemental File S1) (Fig. 2C and S3D). The EMP-
245 like space was devoid of cells, including hematopoietic CD45^{+} cells (Fig. 2B), at the later 192 and
246 216h-time points (Fig. S3D), which instead occupied a second discrete *Runx1*$^{+}$*Cdh5*$^{-}$ transcriptional
247 space (Fig. 2D) of myeloid-lymphoid affiliation (Fig. 2C and S3D) (Supplemental File S1), compatible
248 with MLP-like cells. We sought to align gastruloid-derived hemogenic cells with developmental
249 hematopoiesis in the embryo by projecting our data onto recent scRNA-seq studies which dissected
250 endothelial maturation and HE-to-hematopoietic transitions (Fadlullah et al., 2022), and pre-HSC
251 and HSC emergence (Vink et al., 2020) in the AGM. We reanalyzed the scRNA-seq datasets using
252 our own pipeline. UMAP-based clustering and dimensionality reduction representations broadly
253 recapitulated the clustering and representations of the original publications (Fig. 3A-C, and S4A-B),
254 with the exception of minor clusters (clusters 6, 9, and 10, Fig. 3A; 4, 5, 7, and 9, Fig. 3B), which are
255 not represented or are captured distinctly in the original probably due to filtering criteria, but which
256 configure distinct hematopoietic lineage affiliations (Supplemental File S3) (Fig. S4A-B). Compatible
257 with our cell type classification, gastruloid endothelial clusters 0 and 5 projected onto embryonic
258 clusters of endothelial, aortic endothelial, and HE affiliation (Fig. 3D-E), with rare cells exhibiting pre-
259 HSC characteristics (Fig. 3E). Candidate EMP-like (gastruloid cluster 4, Fig. 2B) and MLP-like
260 (gastruloid cluster 8, Fig. 2B) clusters, on the other hand, projected onto clusters of similar affiliation
261 in the embryo datasets (Fig. 3A, D, cluster 6; Fig. 3C, E, clusters 4 and 5). Interestingly, some of the
262 cells occupied the transcriptional space of HE-to-hematopoietic (gastruloid cluster 4, Fig. 3A, D
263 cluster 7), and pre-HSC-to-HSC transition (gastruloid cluster 8, Fig. 3C, E clusters 0 and 6), in
264 support of incipient definitive hematopoiesis representation in the gastruloid model. Importantly, the
265 gastruloid model also recapitulates differentiation of stromal / niche elements critical for
266 hematopoiesis specification in the embryo. These include: *Pdgfra*$^{+}$ stromal cells (gastruloid cluster

1, Fig. 2B and Fig. 2D), which project onto the PDGFRa⁺ mesenchyme in Fadlullah et al. (2022) (Fadlullah et al., 2022) (Fig. 3A,D, embryo cluster 5); and cells with candidate autonomic nervous system identity (gastruloid clusters 3 and 9, Fig. 2B) (Fig. 2C and S3C) which are not captured in the Fadlullah et al. (2022) and the Vink et al. (2020) studies, but have been reported to sustain HSC specification (Fitch et al., 2012; Kapeni et al., 2022). Both these niche cell types are specifically present at 192h (Fig. S3D, 192h clusters 1-3) and probably emerge 1 day earlier (Fig. S3D, 168h), thus configuring a timeframe compatible with their reported support of pre-HSC. The remaining

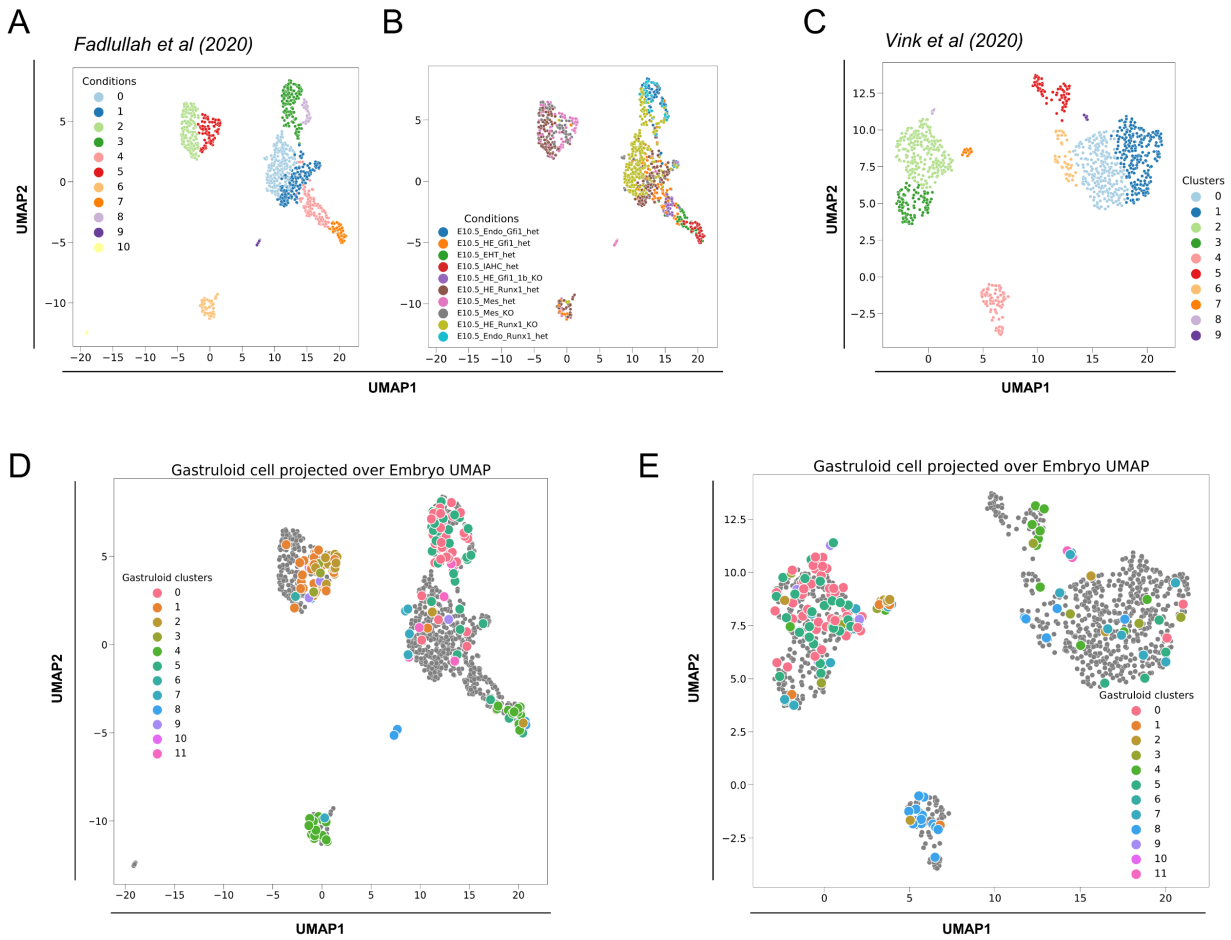


Figure 3 – Single cell profiles of hemogenic gastruloids project onto hemato-endothelial differentiation programs in the mouse AGM. (A-B) Re-analyzed UMAP projection of scRNA-seq dataset from Fadlullah et al. (2020) showing annotated clusters **(A)** and sequenced conditions **(B)** through endothelial to hematopoietic maturation. **(C)** Re-analyzed UMAP projection and cluster annotation of scRNA-seq dataset from Vink et al. (2020) capturing pre-HSC and HSC emergence (clusters 0 and 1), endothelial cells (clusters 2 and 3), and other hematopoietic-affiliated populations (clusters 4-9). **(D-E)** Projection of hemogenic gastruloids single-cell profiles from this study (colored according to their annotated clusters in Fig. 2B) over the UMAP from Fadlullah et al. (2020) **(D)** and Vink et al. (2020) **(E)**.

283 gastruloid clusters have unclear lineage affiliations (Supplemental File S2). They do not project onto
284 AGM-derived cells (Fig. 3D-E), probably reflecting their early gastruloid time signatures (Fig. S3D,
285 120-144h) or a distinct topography which is not strictly hemogenic.

286 Altogether, the data support the notion that the gastruloid model captures developmental blood
287 formation with cellular and temporal precision, and faithfully recapitulates the multi-wave nature of
288 pre-definitive and definitive hematopoiesis. Projection onto the embryo places the 216h-end of the
289 gastruloid protocol at the pre-HSC transition, suggesting that it constitutes a good model to
290 understand intrinsic and extrinsic regulation of definitive hematopoiesis.

291

292 **Infant leukemia gene *MNX1* favours early steps of hemogenic specification**

293 We explored the utility of the hemogenic gastruloid model in understanding the cellular
294 consequences of oncogenic events putatively initiated during development. We elected to
295 investigate *MNX1*-driven AML, the most common of infant-unique forms of AML, which we and
296 others (Ingenhag et al., 2019; Ragusa et al., 2022; Waraky et al., 2022) have shown to be incapable
297 of transforming adult cells.

298 We overexpressed *MNX1* (MNX1-OE) in *FLK1-GFP* mouse ES cells by lentiviral transduction (Fig.
299 S5A) and compared MNX1-OE and control (empty vector, EV) cell performance in the hemogenic
300 gastruloid model (Fig. 4A). We used a human *MNX1* cDNA (Fig. 4B) to distinguish from the
301 endogenous gene, but the degree of homology is nevertheless high (84%), supporting functional
302 equivalence. MNX1-OE gastruloids activated polarised *FLK1-GFP* expression and elongated with
303 similar kinetics to EV (Fig. 4A), but consistently produced larger gastruloids (Fig. 4C) denoting
304 increased cellularity (Fig. S5B). From 192h onwards, MNX1-OE gastruloids had a higher frequency
305 of spontaneously contractile structures (Fig. 4D and Supplemental Movie), compatible with
306 mesodermal cardiac specification. We interrogated the cellularity of gastruloids at the critical
307 hemogenic timepoints of 144h and 192h using markers c-Kit, CD41 and CD45 in flow cytometry. We
308 observed a significant expansion of the c-Kit^{+} compartment specifically at 144h (Fig. 4E and S5C),
309 with relative reduction of CD41^{+} cells (Fig. 4F and Fig. S5D) at the same time point. No changes in

310 proportion of any of the cell types were observed at 192h (Fig. 4E-G and S5C-D), although absolute
311 cell numbers may be higher in MNX1-OE gastruloids (Fig. S5B).

312 In order to understand the molecular programs downstream of *MNX1* which promote cardiogenic,
313 endothelial and potentially hematogenic fates, we initially performed bulk RNA-sequencing of end-
314 point gastruloids and their starting mouse ES cells with and without *MNX1* overexpression (Fig. S6A-
315 B). We verified expression of the human *MNX1* transgene in the sequenced reads (Fig. S6C).
316 Interestingly, we could not detect endogenous *Mnx1* in any of the samples (Fig. S6C), suggesting
317 that it does not normally play a role at the hematogenic stage of embryonic development, or indeed
318 in pluripotent cells. We detected 464 differentially overexpressed genes in MNX1-OE gastruloids (p-
319 value<0.05, FDR adjusted p-value<0.10; Supplemental File S4), and used cell type enrichment
320 (PanglaoDB, Fig. 5A) and GO Biological Process (Fig. 5B-C and S6D) to identify enriched functional
321 categories. MNX1-OE hemogenic gastruloids had enrichment in cardiogenic programs (Fig. 5A-C;
322 5A – cardiomyocytes; 5B-C – muscle contraction, heart development, muscle organ development),
323 in agreement with our observations of autonomous contractile activity (Fig. 4D). There was also
324 enrichment of programs associated with endodermal derivatives, including the gut (Fig. 5A –
325 enterocytes, which may further enhance heart development (Silva et al., 2021). Other enriched
326 functional categories were predominantly metabolic, including lipid metabolism (Fig. 5B, D; S6D-E),
327 reminiscent of some of the transcriptional consequences of *MNX1* in the context of leukemia
328 (Ragusa et al., 2022). Compatible with the absence of relative enrichment of hemogenic cells in late
329 gastruloids, we did not find an enrichment in endothelial or hematopoietic-affiliated programs in
330 MNX1-OE gastruloids. In contrast, *MNX1*-expressing mouse ES cells had clear evidence of
331 activation of blood (Fig. 5D – platelets) and angiogenic signatures (Fig. 5E-F – angiogenesis),
332 including key markers and regulators such as *c-Kit*, *Vegfa*, *Sox17* (Clarke et al., 2013; Nobuhisa et
333 al., 2014), and *Runx1* (Fig. 5F) (Supplemental File S4), which are required for the establishment of
334 definitive hematopoiesis. Likewise, incipient cardiogenic programs are also enriched in MNX1-OE

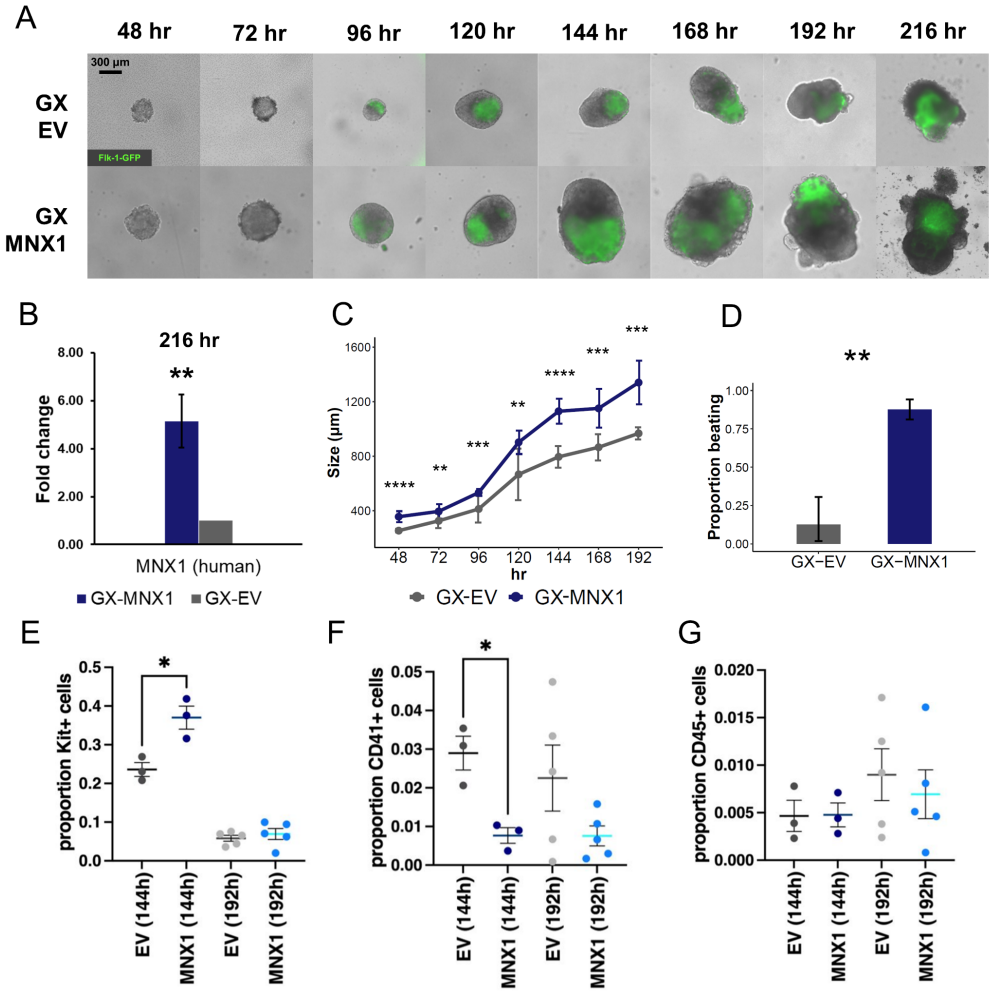


Figure 4 – Hemogenic gastruloids with *MNX1* overexpression have increased cellularity and enhanced cardiogenic and hemogenic endothelial potential. (A) Imaging of hemogenic gastruloids with *MNX1* overexpression and EV controls at 10x magnification, showing appropriate assembly and polarization of the Flk1-GFP marker. scale bar: 300µm. **(B)** Quantitative (q)RT-PCR confirming *MNX1* overexpression in hemogenic gastruloids at endpoint (216h). Fold change in expression was calculated by normalization to *HPRT1*. Statistical difference was calculated by Student's t-test, $p < 0.001$ (**). Error bars indicate ± SD of 3 replicates. **(C)** Size of gastruloids at each timepoint, determined by the distance of the longest extremes in *µ*m. Statistical difference was calculated by Student's t-test, p value < 0.05 (*), 0.001 (**), 0.0001 (***), and 0.00001 (****). Error bars represent standard deviation (SD) of n=3. **(D)** Proportion of gastruloids exhibiting spontaneous contraction at 192h; error bars show SD of n=3. **(E-G)** Flow cytometry quantification of positive MNX1 and EV gastruloids cells for c-Kit **(E)**, CD41 **(F)**, and CD45 **(G)** at 144h and 192h. Statistical difference was calculated by Student's t-test, $p < 0.05$ (*).

mouse ES cells (Fig. 5D – cardiac precursor cells) prior to cardiac cell specification in gastruloids. It is important to note that MNX1-OE mouse ES cells can be maintained as pluripotent cells in the presence of LIF and do not exhibit morphological signs of spontaneous differentiation, suggesting that the enhancement of hemogenic and cardiogenic signatures denotes priming rather than full-scale activation of lateral mesoderm lineage programs.

353 Taken together, functional and transcriptional analysis of MNX1-OE gastruloids suggests that *MNX1*
354 ectopically promotes the specification of lateral mesodermal lineages, including activation of hemato-
355 endothelial programs and early expansion of hemogenic endothelium, compatible with the
356 hypothesis of an early developmental origin of *MNX1* infant leukemia.

357

358 ***MNX1* targets the HE-to-EMP transition to initiate leukemic transformation**

359 In an attempt to deconvolute the cellular heterogeneity underlying bulk RNA-seq differential gene
360 expression of MNX1-OE hemogenic gastruloids, we interrogated the cluster-specific signatures
361 inferred from our single-cell time-course analysis of gastruloid differentiation (Fig. 6A; refer to Fig.
362 2B), including the clustering structure at individual timepoints (Fig. 6B; refer to Fig. S3D), to infer
363 differential cell type representation. In support of enhanced hemato-endothelial specification driven
364 by *MNX1*, we observed enrichment of HE cluster signatures, comprising both clusters 5 and 0 (144
365 and 192h, respectively), as well as of EMP-like cluster 4 (144h) and MLP/pre-HSC-like cluster 8
366 (192/216h) signatures (Fig. 6A), altogether compatible with increased representation of HE and
367 hematopoietic cells in MNX1-OE gastruloids. The signature for clusters 10 (120h), which has less
368 clear lineage-affiliations and include pluripotency genes such as *Zfp42*, *Nanog*, and *Sall4*
369 (Papatsenko et al., 2015), was also enriched, which could suggest persistence of primed ES cells.
370 Significantly, clusters 5 and 0 HE signatures were also enriched in RNA-seq data of infAML patients
371 (Balgobind et al., 2011) (TARGET, https://ocg.cancer.gov/programs/target) carrying an *MNX1*-
372 overexpressing chromosomal rearrangement (MNX1-r) [t(7;12) translocation] (Fig. 6A-B), thus
373 implicating a characteristic hemogenic developmental stage in infAML biology, and validating the
374 mechanistic role of *MNX1* in this form of leukemia. It is noteworthy that, unlike HE signatures, EMP
375 and MLP/pre-HSC signatures were not enriched in MNX1-r/t(7;12) infAML (Fig. 6A-B).

376 The t(7;12) RNA-seq signature was calculated in comparison with other forms of pediatric leukemia
377 (Supplemental File S5) (see STAR Methods for detail), which could dilute hematopoietic progenitor
378 signatures. However, RNA-seq data for the most common form of infAML, driven by the t(9;11)
379 *KMT2A-MLLT3* fusion (https://ocg.cancer.gov/programs/target), shows a significant enrichment of

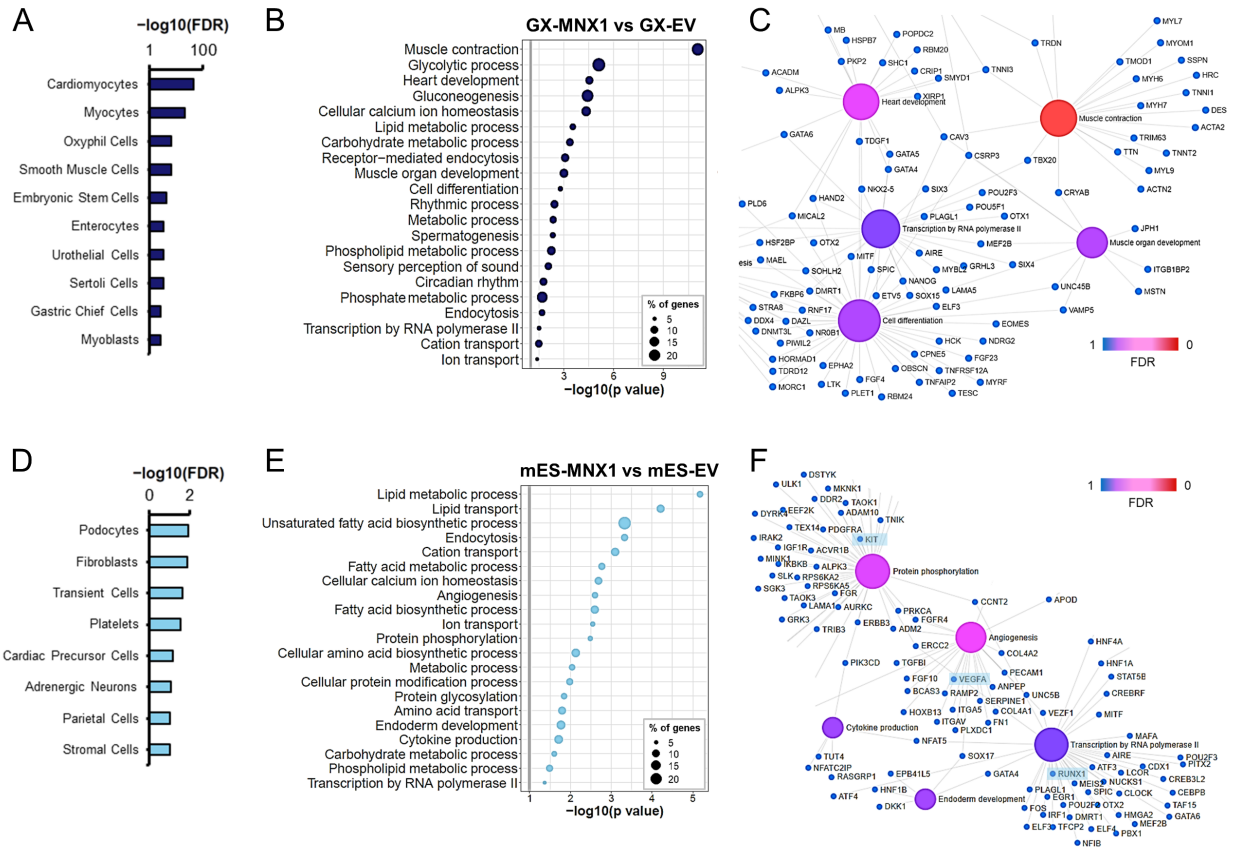


Figure 5 – *MNX1* overexpression promotes lateral mesodermal expansion in hemogenic gastruloids and primes angio- and hemogenic programs in mES. (A) Enrichment of Gene Ontology (GO) terms in differentially upregulated genes in GX-MNX1 compared to GX-EV (filtered by FDR ≤ 0.1) relating to biological processed (PANTHER database). The grey intercept shows p = 0.05 threshold in –log10. **(B)** Cell type analysis of differentially upregulated genes in GX-MNX1 using the Panglao DB 2021 database. **(C)** Bipartite network plot of upregulated genes in GX-MNX1 mapping to PANTHER biological processes filtered, constructed on ExpressAnalyst (www.expressanalyst.ca). Full network shown in Supplementary Figure S6D. **(D)** Biological processes (PANTHER) enriched in differentially upregulated genes in overexpressing mES compared to EV control. **(E)** Cell type analysis corresponding to upregulated genes in MNX1 overexpressing mES. **(F)** Bipartite network of upregulated genes in MNX1-mES mapping to biological processes from PANTHER. Full network in Supplemental Figure S6E.

the MLP/pre-HSC-like cluster 8 (Fig. S7A-B), compatible with the myeloid progenitor affiliation of *KMT2A*-rearranged AML. This re-enforces the notion that the HE stage may indeed be critical in t(7;12) *MNX1* leukemia biology. To verify the significance of the individual genes enriched in each of the signatures, we considered the leading-edge genes (LEGs) in MNX1-OE and t(7;12) enrichments (Fig. 6C and Supplemental File S5). Over 40% of t(7;12) LEGs overlapped with MNX1-OE gastruloid LEGs, with less reverse overlap as expected, given the heterogenous cellular composition of gastruloids (Fig. 6C). Again, common LEGs, as well as a subset t(7;12)-unique genes aligned with endothelial cells (Fig. 6D and S7C), firming the link with HE. Similar to cluster enrichments (Fig. 6A-

400 B), association with erythroid cell types was exclusive to MNX1-OE (Fig. 6D), putatively separating
401 leukemia and non-leukemia-specific roles of MNX1. Conversely, a t(7;12)-unique association with
402 mesodermal / stromal cell types (Fig. 6D and Fig. S7C), encapsulated by the *Pdgfra*$^{+}$ cluster 1 (Fig.
403 6A and 2D), is not captured in MNX1-OE and EV gastruloids (Fig. 6B).

404 Although heterogeneous in terms of hematopoietic-lineage affiliation, *MNX1* t(7;12) leukemias are
405 most often of moderately-differentiated myeloid morphology (Espersen et al., 2018; Wildenhain et
406 al., 2010), suggesting progression beyond the HE stage. To more closely investigate later
407 hematopoietic stages, we performed scRNA-seq on MNX1-OE and EV cells from 216h gastruloids
408 broadly sorted on CD41^{+} and/or CD45^{+} gates to encompass all hematopoietic cells. The sequencing
409 quality control confirmed robust detection of an average of 120000 reads/cell, corresponding to 4000
410 aligned genes/cell (EV and MNX1-OE, respectively), with no stress signatures. Clustering analysis
411 (Fig. S7D and Supplemental File S6) distinguished 6 clusters with representation of both EV and
412 MNX1-OE cells, albeit with enrichment of MNX1-OE cells in cluster 0 and of EV cells in cluster 1
413 (Fig. S7D-E). Separate clustering and projection of each condition identified 3 clusters each in EV
414 and MNX1-OE cells, which were more clearly discrete in EV compared to MNX1-OE gastruloid cells
415 (Fig. 6E), putatively suggesting distinct degrees of cell specification and/or lineage resolution.
416 Accordingly, cluster-defining molecular signatures were more statistically robust in EV cells
417 (Supplemental File S6). More than half of EV cells were found in cluster 1, which has a gene
418 expression signature enriched for targets of key hematopoietic regulators PU.1 (SPI1), RUNX1 and
419 GATA2 (Fig. 6F), with evidence of myeloid, lymphoid and erythroid lineage programs (Fig. 6G)
420 (Supplemental File 6). Hematopoietic affiliations are more evenly spread across MNX1-OE clusters,
421 particularly 0 and 2, which capture enrichments for GATA2 and PU.1 targets, respectively (Fig. 6F).
422 Accordingly, lineage-affiliated programs were present in the 3 clusters (Fig. 6G), with a bias towards
423 endothelial or hemato-endothelial signatures enriched in cluster 0, and lymphoid and myelo-
424 lymphoid signatures in cluster 2, consistent with the enriched transcriptional regulators (Fig. 6H-I).
425 Relative to EV, MNX1-OE has more extensive erythroid priming, particularly of hemoglobin chains
426 (Fig. 6G), as well as of endothelial signatures, including co-expression with erythro-myeloid

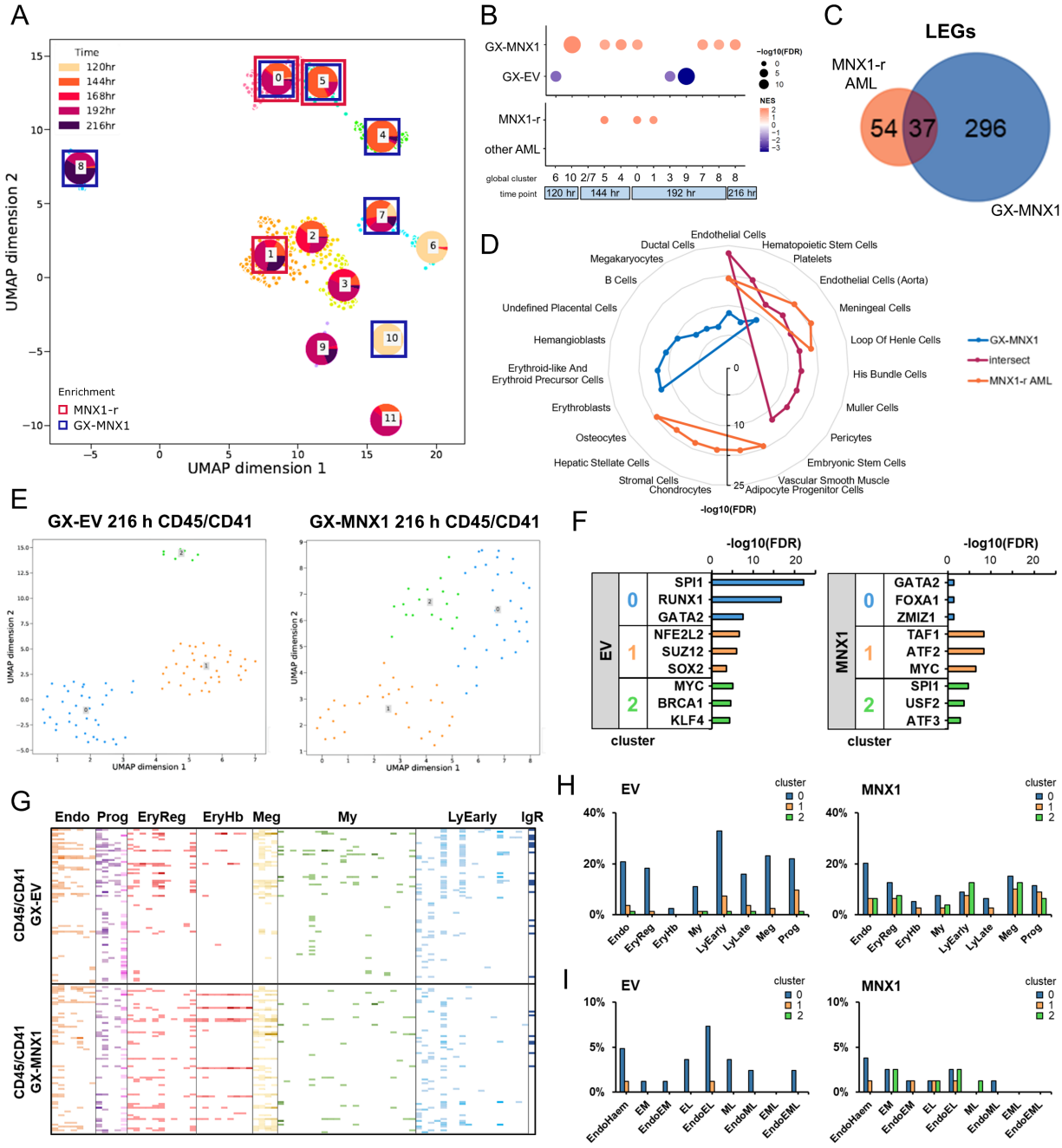
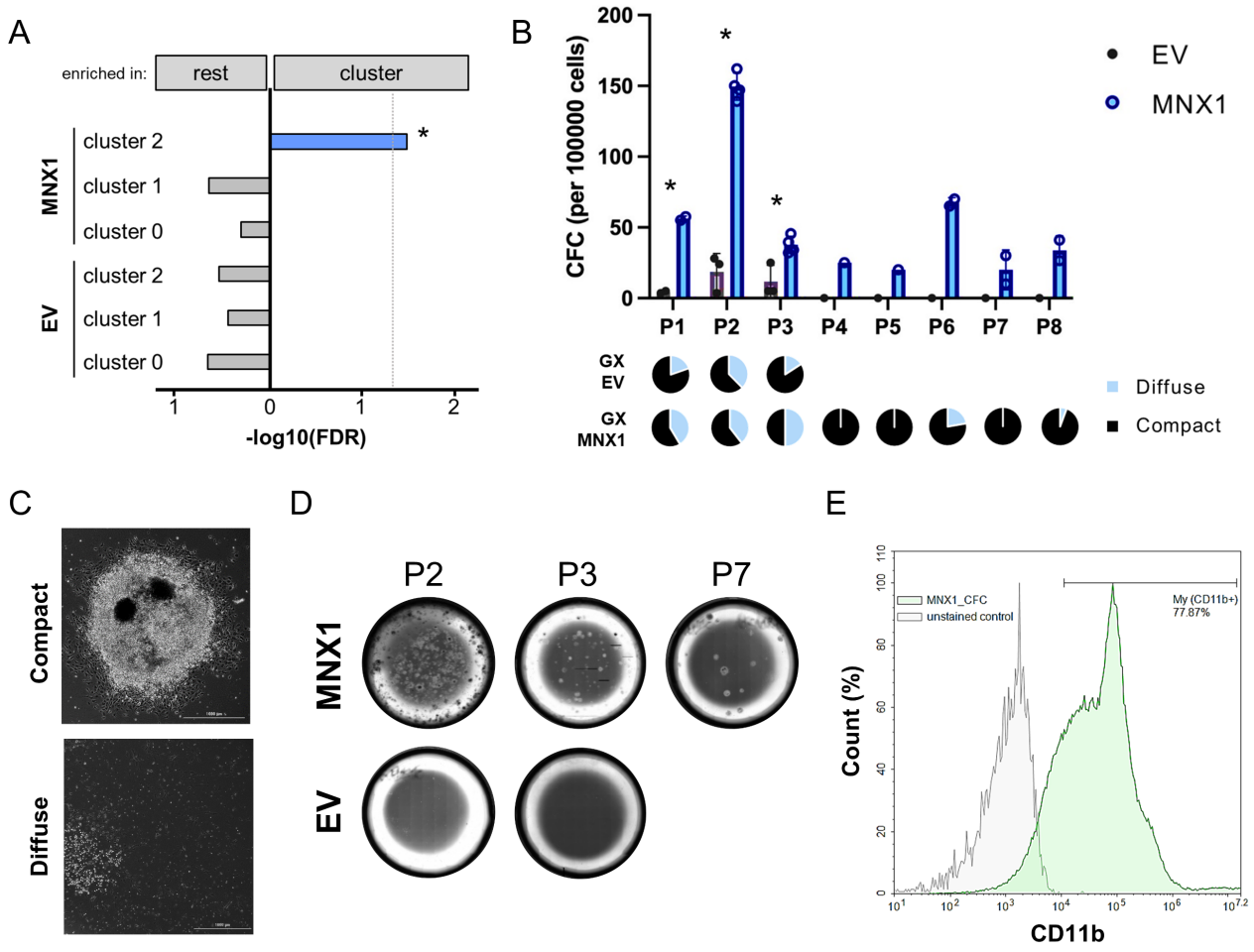


Figure 6 – Hemogenic gastruloids overexpressing MNX1 capture transcriptional features of MNX1-r AML. (A) UMAP of time-resolved global clustering of scRNA-seq of hemogenic gastruloids showing the enriched clusters in GX-MNX1 (blue boxes) and MNX1-r AML (red boxes) determined by GSEA. Cluster numbers as per Fig. 2B. **(B)** Bubble plot of GSEA NES values and statistical significance by –log10(FDR) of enrichments in specific time clusters and corresponding global clusters in GX-MNX1 compared to GX-EV, and MNX1-r AML samples compared to other pediatric AML. **(C)** Intersect between GSEA leading edge genes (LEGs) enriched in MNX1-r AML and GX-MNX1 against time-resolved scRNA-seq gastruloids clusters. **(D)** Radar chart mapping enriched cell types (analyzed in EnrichR using the Panglao DB database) using LEGs shown in **C. (E)** UMAP of scRNA-seq of sorted CD45+ and/or CD41+ MNX1 and EV gastruloids cells at 216h. **(F)** Transcription factor binding site enrichment on top 100 scRNA-seq cluster identifier genes (filtered by p value ≤ 0.05) using the ENCODE and ChEA Consensus TFs from ChIP database on EnrichR. **(G)** Heatmap showing the expression of hematopoietic lineage markers (listed in Methods) for each cell in CD45⁺ and/or CD41⁺ sorted GX-EV and GX-MNX1 at 216h. **(H)** Percentage of cells for each lineage identity (determined by co-expression of at least 2 lineage genes) by cluster in CD45⁺/CD41⁺ EV (left) and MNX1 (right) gastruloids at 216h. **(I)** Percentage of multi-lineage priming (determined by co-existence of lineage signatures) by cluster in CD45⁺/CD41⁺ EV (left) and MNX1 (right) gastruloids at 216h.

443 signatures (Fig. 6I). In contrast, EV cells had a higher frequency of expression of recombinant Ig
444 chains, as well as of hematopoietic progenitor regulatory signatures (Fig. 6G). Altogether, the data
445 are suggestive of perturbed hematopoietic progression upon MNX1-OE, with persistence or
446 expansion of HE cells at the HE-to-EMP transition, and relative depletion of more progressed myelo-
447 lymphoid progenitors and pre-HSC. Crucially, MNX1-OE cells in cluster 2 were singularly enriched
448 in t(7;12) AML patient-signature genes (Fig. 7A), suggesting that this cluster may capture myeloid
449 leukemia transformation. To test transformation potential of MNX1-OE cells, we dissociated EV and



450

451 **Figure 7 – MNX1 overexpression transforms hemogenic gastruloids to gain re-plating capacity of myeloid**
452 **colonies. (A)** Enrichment of MNX1-r AML patient signatures from Ragusa et al. (2022) in clusters of CD45+/CD41+ MNX1
453 and EV gastruloids, determined by GSEA by comparison of each cluster against the others ('rest'). Statistical significance
454 is by FDR; *p<0.05. **(B)** Serial colony-forming assay (CFC) of disassembled unsorted MNX1 and EV gastruloids at 216h
455 on methylcellulose-based medium. 'P' indicates the order of plating. Mean±SD of n=2-4 replicate experiments; *p<0.05,
456 unpaired t-test with Welch correction. Underneath, pie charts indicate the proportion of colony types at each replating. **(C)**
457 Representative images of colony types identified as 'compact' (top) and 'diffuse' (bottom) at 20x magnification. Scale bar
458 = 1000 μm. **(D)** Representative images of methylcellulose plates at early (P2-P3) and late platings (P7). **(E)** Flow cytometry
459 plot of late plating colonies at P8 for the myeloid marker CD11b; data are overlayed with the respective unstained control
460 sample, with normalized counts.

461

462 MNX1-OE gastruloids, placed them in CFC progenitor assays containing multi-lineage cytokines,
463 and assessed colony frequency and serial re-plating capacity as an *in vitro* measure of
464 transformation. From the initial plating, MNX1-OE cells generated more colonies (Fig. 7B), with
465 progressive selection of a characteristic compact colony morphology depicted in Fig. 7C. EV-derived
466 colonies were extinguished after 3 platings, while MNX1-OE colony-forming cells persisted at even
467 frequencies for at least 5 additional platings (Fig. 7D), and generated cells with a myeloid surface
468 phenotype, as per flow cytometry detection of early myelo-monocytic marker CD11b (Fig. 7E).
469 Altogether, the data indicate that overexpression of *MNX1* during developmental hematopoietic
470 specification, expands HE and favours a distinct and potentially heterogenous pathway of myeloid
471 differentiation with leukemic self-renewal potential.

472

473 **DISCUSSION**

474 In this study, we have developed and explored a 3-dimensional gastruloid model of hemogenic
475 mouse development. Through sequential utilisation of extrinsic cues for symmetry breaking,
476 mesodermal induction, vascular development, and hematopoietic maturation, we extended the self-
477 organising properties of the gastruloid system to recapitulate key aspects of blood formation with
478 spatiotemporal accuracy. Namely, we observed the successive specification of hemogenic
479 endothelium (HE), and of pre-definitive erythro-myeloid progenitors (EMP), definitive myelo-
480 lymphoid progenitors (MLP) and pre-HSC in coherent temporal progression, and we captured the
481 formation of discrete micro-aggregates of hematopoietic cells in intimate association with
482 endothelial-like lumina, reminiscent of mid-gestation HSC-generating aortic clusters. Hemato-
483 endothelial development was accompanied by generation of stromal components critical for HSC
484 emergence, including PDGFRA^{+} mesenchyme (Chandrakanthan et al., 2022) and sympathetic
485 neurons (Fitch et al., 2012; Kapeni et al., 2022), suggesting coordinated organisation of elements
486 of a hemogenic niche. Importantly, the *in vitro* gastruloid system responded to a relevant oncogenic
487 event unique to infant AML (infAML) and putatively characteristic of the developmental period, i.e.

488 *MNX1* overexpression (MNX1-OE). MNX1-OE gastruloids displayed perturbation of hemogenic
489 trajectories and selection of a transformed myeloid progenitor, which not only recapitulated the
490 molecular signatures of human patients, but critically shed light on previously uncharacterised
491 mechanisms of transformation associated with *MNX1*. Thus, recapitulation of physiological stages
492 of blood development and clinical-relevant responsiveness to a pathological leukemia-initiating event
493 uniquely associated with embryonic development, together configure the hemogenic gastruloid as a
494 faithful, robust, and accessible *in vitro* model to dissect blood formation.

495 Our hemogenic gastruloid system extends recent findings by (Rossi et al., 2022), who observed
496 early erythro-myeloid specification in a gastruloid protocol previously developed for modelling of
497 primordial heart development (Rossi et al., 2021). Together, our data highlight the power of
498 gastruloid models in faithfully capturing mid-development specification of lateral mesoderm-derived
499 programs. They also suggest that modelling of post-gastrulation mesodermal lineage development
500 progresses modularly, with protocol optimization preferentially supporting individual cell fates. Our
501 protocol is not without its limitations, namely (1) it does not organise a networked vascular system,
502 and (2) there is no production of functional hematopoietic stem cells (HSC). In line with the notion of
503 modularity of lineage development, our protocol does not consistently produce spontaneously
504 contractile structures, i.e. a heart primordium, a process intimately associated with HSC formation.
505 Despite the ability of gastruloids to execute transcriptional programs and configure temporally-
506 accurate regulatory networks that sustain specification of cell identity and enable the correct
507 assembly of network-directed micro-topographies, they are unable to fully capture the mechanical
508 and chemical cues that direct and remodel macro-topographies and ultimately specify organismal
509 physiology. Likely, their optimal cell number at initiation gives gastruloids an advantage relative to,
510 for example, much larger embryoid bodies in which micro-topographies may quickly become
511 redundant and disorganised, and, in the case of blood formation, require cell dissociation and
512 additional stages of cell culture to reach similar stages of hematopoietic specification, with loss of
513 the temporal resolution (Garcia-Alegria et al., 2018; Pearson et al., 2015; Sugimura et al., 2017). At
514 a fundamental level, gastruloid models can be seen as capturing the interface between intrinsic and

515 extrinsic regulation, and become limiting at the critical transition where intrinsic cell identity is
516 superseded by extrinsic tissue or organ coordination enabled by long-distance chemical and physical
517 signalling. Other single-tissue organoid models employ matrices that provide the extracellular
518 support required to organise a degree of macro-topography. Gastruloids themselves, elegantly
519 organise somitogenesis (van den Brink, S C et al., 2020), or approximate a central nervous system
520 (Girgin et al., 2021), in the presence of matrices, but precise combinations of chemical and
521 mechanical cues are tissue-specific and successful 3-dimensional *in vitro* organisation are
522 necessarily modular. In the case of blood, different cell identities are specified independently, at
523 different times and in different places, during embryonic development (Costa et al., 2012; Dzierzak
524 and Bigas, 2018; Lacaud and Kouskoff, 2017; Medvinsky et al., 2011), and the micro-topographical
525 accuracy of the gastruloid model assists in faithfully capturing the successive hematopoietic waves
526 in intimate association with hemogenic endothelium and with contemporaneous specification of at
527 least some relevant stromal cells. Absence of *bona fide* HSC, i.e. transplantable cells capable of
528 long-term reconstitution of the hematopoietic system may be due to insufficient developmental time
529 at the end of the protocol – 216h may be short of HSC emergence – or insufficient extrinsic macro-
530 topographical regulation – with the disorganised vascular network and the absence of coordinated
531 contractile activity key omissions. The protocols that have achieved production of *bona fide* HSC
532 from pluripotent sources (Sugimura et al., 2017) required ectopic expression of multiple transgenes
533 to enforce key transcriptional networks and signalling pathways, and have been difficult to reproduce
534 in other laboratories, suggesting a requirement of yet undefined factors. It has also been proposed
535 that circulating blood flow contributes to HSC specification and shedding from the intra-aortic clusters
536 (North et al., 2009), compatible with the sequential nature of initiation of heart contractility and blood
537 flow, and intra-embryonic re-localization of blood formation and HSC production. Protocols with
538 reported success at HSC formation (Lis et al., 2017; Sugimura et al., 2017) generate HSC at the
539 stromal co-culture step which may expose the culture surface to a certain level of haphazard flow
540 stress. In a proportion of gastruloids, we have observed the generation of cluster-like structures at
541 the surface, rather than internally, re-enforcing the notion of a contribution from mechanical cues.

542 Several recent reports have demonstrated the utility of organ-on-a-chip (Ooc) solutions for the
543 recreation of adult bone marrow environments in physiological (Sugimura et al., 2020; Torisawa et
544 al., 2014) and pathological (Abdullah Obaid Khan et al., 2022; Chou et al., 2020), i.e. leukemia
545 settings. In the future, it is worth considering adapting Ooc platforms to support gastruloid growth,
546 providing exposure to bespoke matrices and levels of flow at specific stages, to promote HSC
547 development.

548 The hemogenic gastruloid protocol described in this study does nevertheless faithfully capture pre-
549 definitive EMP-like cells and definitive MLP-like cells and putative pre-HSC which are
550 transcriptionally aligned with E8.5-E10.5 stages of mouse development. Interestingly, the gastruloid
551 hemogenic endothelial cells that temporally associate with EMP-like and definitive hematopoietic
552 progenitors occupy adjacent but largely non-overlapping clusters, in agreement with independent
553 waves of blood formation. The physiological alignment of gastruloid HE is highlighted by the
554 successful modelling of *MNX1*-rearranged (*MNX1*-r) AML, which had so far remained elusive.
555 Several reports by us and others have used ectopic *MNX1-ETV6* cDNA (Wildenhain et al., 2010),
556 *MNX1* overexpression (Ingenhag et al., 2019), or genomic engineering of the t(7;12) rearrangement
557 (Nilsson et al., 2022; Ragusa et al., 2022) in adult mouse or human HSC and progenitors in an
558 attempt to model *MNX1*-r leukemia, but consistently failed to achieve transformation *in vitro* or *in*
559 *vivo*. Similar to this study, (Ingenhag et al., 2019) observed a relative expansion of erythroblasts
560 upon *MNX1* overexpression, consistently supporting a unique effect in the erythroid lineage, but one
561 which is independent of the potential for transformation. While these observations could bring into
562 question the value of *MNX1* overexpression as a surrogate of *MNX1*-r, the recent engineering of the
563 t(7;12) rearrangement in human iPS cells (Nilsson et al., 2022) also captured an expansion of
564 erythroid progenitors, re-aligning the functional consequences of *MNX1* and its targeting
565 translocation. Interestingly, t(7;12) engineering in iPS cells also promoted differentiation of myeloid
566 progenitors, providing a more accurate alignment with the lineage phenotype of the leukemia,
567 putatively favoured by the expression of embryonic transcriptional programs, although with no
568 evidence of transformation (Nilsson et al., 2022). Systematic comparison of *MNX1*-r / t(7;12) with

569 other pediatric AML (Balgobind et al., 2011; Fornerod et al., 2021; Ragusa et al., 2022) identified a
570 unique transcriptional program enriched for lipid metabolism and cell adhesion signatures, but with
571 limited affiliation to hematopoietic lineages. Instead, t(7;12) signatures include endothelial genes
572 such as *KDR*, *VWF* or *EDIL3*, as well as *LIN28B*, a negative regulator of fetal-to-adult HSC
573 maturation via repression of *let-7* (Wang et al., 2022). Detailed mapping of gastruloid hematopoietic
574 development against *MNX1*-r / t(7;12) leukemia signatures unveils their affiliation with HE, which we
575 found to be expanded by MNX1-OE at the EMP-like stage suggestive of an early, putatively pre-
576 definitive, target cell of *MNX1* ectopic expression. The later HE signature at 192h was also enriched
577 in both end-point MNX1-OE gastruloids and t(7;12) AML samples, although no relative expansion of
578 Kit+ cells was observed at that time point of gastruloid differentiation, suggesting that the 192h
579 enrichment may convey the HE lineage affiliation rather than precisely reflect the developmental
580 time. Indeed, the dispersed profiles of MNX1-OE hemogenic cells by scRNA-seq are compatible with
581 differentiation heterogeneity, asynchrony and/or delayed resolution of developmental programs, also
582 apparent through detection of myeloid differentiation signatures in 2 neighbouring clusters, 0 and 2,
583 which reflect distinct weak GATA2-centric and PU.1 / SPI1-centric regulatory networks. The latter
584 likely corresponds to the early-transformed cell, which we were able to select through serial replating
585 and is phenotypically aligned with the myeloid classification of *MNX1*-r / t(7;12) AML. Our study also
586 detects the erythroid affiliation bias previously described in *MNX1* overexpressing BM cells, which
587 diverges from t(7;12) AML-associated signatures. Significantly, this erythroid bias does not configure
588 a distinct MNX1-OE gastruloid cluster upon single-cell analysis, but instead corresponds to pervasive
589 priming of erythroid signatures, which co-exist with or are superimposed on, endothelial, myeloid,
590 and lymphoid signatures. Notably, the endothelial priming is more extensive, but the lymphoid
591 programs are less progressed than in control cells, and there is lower frequency of expression of
592 pre-HSC associated transcription factors. This suggests expansion and/or selection of a cell at the
593 HE-to-hematopoietic transition, which is developmentally closer to the pre-definitive EMP stage
594 and/or uniquely tolerant of ectopic erythroid expression, a property not shared by adult BM cells,
595 which couple MNX1-expanded erythroid expression with senescence. In contrast, MNX1-OE

596 gastruloid cells persist and expand, and can transform into a leukemia which is myeloid in nature,
597 and carries unique residual endothelial, but not erythroid, programs, effectively dissociating
598 transformation from erythroid specification or bias. It is possible that the MNX1-driven oncogenic
599 program requires a cell that is tolerant of erythroid gene priming, but does not progress through
600 hematopoietic lineage differentiation, and consequently dispenses full activation, or full repression
601 of the erythroid signature. Such tolerance may be explained by chromatin plasticity and/or an
602 adaptive checkpoint response, either as a consequence of MNX1 activity, or as a result of additional
603 genetic events. In t(7;12) AML patients, additional copies of chromosomes 19 or 8 are frequently
604 encountered (Espersen et al., 2018), whose origin and significance are currently unexplored. Further
605 exploration of the gastruloid model and dissection of the corresponding stages in the embryo will
606 refine these findings, and pinpoint the cell-of-origin and precise developmental window in which
607 MNX1 is active, as well as suggest specific vulnerabilities that can be targeted therapeutically.

608 In conclusion, the hemogenic gastruloid model constitutes a faithful *in vitro* model of embryonic and
609 fetal blood specification and development. Its cellular and temporal resolution has the potential to
610 add to the mechanistic understanding of developmental blood disorders, offering a medium-
611 throughput platform for therapeutic compound testing. Refinement of the mouse gastruloid model
612 and extension to the human will further strengthen the translatability of its findings.

613

614 **ACKNOWLEDGEMENTS**

615 This project was funded by a start-up grant and a BRIEF award from Brunel University London to
616 CP, and by ERC Advanced Grant (MiniEmbryoBlueprint 834580) to AMA. GTC was funded by grant
617 FPU18/05091 from the Spanish Ministry of Universities. CP was also funded by a KKLF Intermediate
618 Fellowship (KL888), a Leuka John Goldman Fellowship for Future Science (2017-2019), and a
619 Wellcome Trust / ISSF Bridge Funding award at the University of Cambridge (2019). JGO
620 acknowledges financial support from the Spanish Ministry of Science and Innovation and FEDER
621 (grant PGC2018-101251-B-I00), by the Maria de Maeztu Programme for Units of Excellence in R&D
622 (grant CEX2018-000792-M), and by the Generalitat de Catalunya (ICREA Academia programme).

Library preparation and next-generation sequencing for single-cell RNA-seq analysis were performed by the Single Cell Genomics Group at the National Centre for Genomic Analysis – Centre for Genomic Regulation (CNAG-CRG), Barcelona. The Authors wish to thank Susanne van den Brink (UPF, Barcelona) for helpful discussions. The Authors also wish to acknowledge Tina Balayo, Ana Filipa Domingues, Oliver Davies, Kristen Place and Remisha Gurung's technical support at different stages of the project.

AUTHOR CONTRIBUTIONS

Conceptualization: CP, AMA; Methodology: CWS, DR, GTC, KRK, JGO, AMA, CP; Software: GTC, JGO; Validation: CWS, DR, CP; Investigation: CWS, DR, GTC, LD, CB, G-AI, CP; Formal Analysis: GTC, DR, JGO, CP; Resources: GTC, JC, JGO, AMA; Data curation: DR, GTC, JGO; Writing – Original Draft: CP, DR, GTC; Writing – Review and editing: CP, DR, GTC, AB, JGO, AMA; Visualisation: DR, GTC, CWS, CP; Supervision: CP; Project administration: CP, AMA, JGO; Funding acquisition: CP, AMA.

DECLARATION OF INTERESTS

AMA and CP are co-inventors in the patent application PCT/GB2019/052668: Polarised Three-Dimensional Cellular Aggregates. The other authors have no interests to declare.

REFERENCES

Anonymous (2018). Method of the Year 2017: Organoids. Nature Methods *15,* 1.

Abdullah Obaid Khan, Michela Colombo, Jasmeet S Reyat, Guanlin Wang, Antonio Rodriguez-Romera, Wei Xiong Wen, Lauren Murphy, Beata Grygielska, Christopher Mahony, Andrew Stone*, et al.* (2022). Human bone marrow organoids for disease modelling, discovery and validation of therapeutic targets in hematological malignancies. bioRxiv

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31,* 166-169.

650 Arias, A.M., Marikawa, Y., and Moris, N. (2022). Gastruloids: Pluripotent stem cell models of mammalian gastrulation and
651 embryo engineering. Dev. Biol. *488,* 35-46.

652 Baillie-Johnson, P., van den Brink, S C, Balayo, T., Turner, D.A., and Martinez Arias, A. (2015). Generation of Aggregates
653 of Mouse Embryonic Stem Cells that Show Symmetry Breaking, Polarization and Emergent Collective Behaviour In Vitro.
654 J. Vis. Exp. *(105). doi,* 10.3791/53252.

655 Balgobind, B.V., Van den Heuvel-Eibrink, M M, De Menezes, R.X., Reinhardt, D., Hollink, I.H., Arentsen-Peters, S.T., van
656 Wering, E.R., Kaspers, G.J., Cloos, J., de Bont, E.S.*, et al.* (2011). Evaluation of gene expression signatures predictive of
657 cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. Haematologica *96,* 221-230.

658 Barker, N., Huch, M., Kujala, P., van de Wetering, M., Snippert, H.J., van Es, J.H., Sato, T., Stange, D.E., Begthel, H., van
659 den Born, M.*, et al.* (2010). Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro.
660 Cell. Stem Cell. *6,* 25-36.

661 Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.C., Lutolf, M.P., Duboule, D., and Arias, A.M.
662 (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature *562,* 272-276.

663 Boiers, C., Richardson, S.E., Laycock, E., Zriwil, A., Turati, V.A., Brown, J., Wray, J.P., Wang, D., James, C., Herrero, J.*,*
664 *et al.* (2018). A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction as Developmental Susceptibility to B
665 Acute Lymphoblastic Leukemia-Associated ETV6-RUNX1. Dev. Cell. *44,* 362-377.e7.

666 Britten, O., Ragusa, D., Tosi, S., and Kamel, Y.M. (2019). MLL -Rearranged Acute Leukemia with t(4;11)(q21;q23)-Current
667 Treatment Options. Is There a Role for CAR-T Cell Therapy? Cells *8,* 1341.

668 Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R.,
669 Zhang, F.*, et al.* (2020). A human cell atlas of fetal gene expression. Science *370,* 10.1126/science.aba7721.

670 Cazzola, A., Cazzaniga, G., Biondi, A., Meneveri, R., Brunelli, S., and Azzoni, E. (2021). Prenatal Origin of Pediatric
671 Leukemia: Lessons From Hematopoietic Development. Front. Cell. Dev. Biol. *8,* 618164.

672 Chandrakanthan, V., Rorimpandey, P., Zanini, F., Chacon, D., Olivier, J., Joshi, S., Kang, Y.C., Knezevic, K., Huang, Y.,
673 Qiao, Q.*, et al.* (2022). Mesoderm-derived PDGFRA(+) cells regulate the emergence of hematopoietic stem cells in the
674 dorsal aorta. Nat. Cell Biol. *24,* 1211-1225.

675 Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr:
676 interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14,* 128-128.

677 Chen, M.J., Yokomizo, T., Zeigler, B.M., Dzierzak, E., and Speck, N.A. (2009). Runx1 is required for the endothelial to
678 haematopoietic cell transition but not thereafter. Nature *457,* 887-891.

679 Chou, D.B., Frismantas, V., Milton, Y., David, R., Pop-Damkov, P., Ferguson, D., MacDonald, A., Vargel Bolukbasi, O.,
680 Joyce, C.E., Moreira Teixeira, L.S.*, et al.* (2020). On-chip recapitulation of clinical bone marrow toxicities and patient-
681 specific pathophysiology. Nat. Biomed. Eng. *4,* 394-406.

682 Clarke, R.L., Yzaguirre, A.D., Yashiro-Ohtani, Y., Bondue, A., Blanpain, C., Pear, W.S., Speck, N.A., and Keller, G. (2013).
683 The expression of Sox17 identifies and regulates haemogenic endothelium. Nature Cell Biology *15,* 502-510.

684 Clevers, H. (2016). Modeling Development and Disease with Organoids. Cell *165,* 1586-1597.

685 Clevers, H.C. (2019). Organoids: Avatars for Personalized Medicine. Keio Journal of Medicine *68,* 95.

686 Costa, G., Kouskoff, V., and Lacaud, G. (2012). Origin of blood cells and HSC production in the embryo. Trends in
687 Immunology *33,* 215-223.

688 Dalby, A., Ballester-Beltran, J., Lincetto, C., Mueller, A., Foad, N., Evans, A., Baye, J., Turro, E., Moreau, T., Tijssen, M.R.,
689 and Ghevaert, C. (2018). Transcription Factor Levels after Forward Programming of Human Pluripotent Stem Cells with
690 GATA1, FLI1, and TAL1 Determine Megakaryocyte versus Erythroid Cell Fate Decision. Stem Cell. Reports *11,* 1462-
691 1478.

692 de Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper,
693 S.A., Speck, N., and Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. J. Exp. Med. *210,* 2843-2850.

694 Dzierzak, E., and Bigas, A. (2018). Blood Development: Hematopoietic Stem Cell Dependence and Independence. Cell.
695 Stem Cell. *22,* 639-651.

696 Espersen, A.D.L., Noren-Nystrom, U., Abrahamsson, J., Ha, S.Y., Pronk, C.J., Jahnukainen, K., Jonsson, O.G., Lausen,
697 B., Palle, J., Zeller, B., Palmqvist, L., and Hasle, H. (2018). Acute myeloid leukemia (AML) with t(7;12)(q36;p13) is
698 associated with infancy and trisomy 19: Data from Nordic Society for Pediatric Hematology and Oncology (NOPHO-AML)
699 and review of the literature. Genes Chromosomes Cancer *57,* 359-365.

700 Fadlullah, M.Z.H., Neo, W.H., Lie-A-Ling, M., Thambyrajah, R., Patel, R., Mevel, R., Aksoy, I., Do Khoa, N., Savatier, P.,
701 Fontenille, L.*, et al.* (2022). Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation
702 marked by ACE. Blood *139,* 343-356.

703 Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G., and Kouskoff, V. (2003). Tracking mesoderm
704 induction and its specification to the hemangioblast during embryonic stem cell differentiation. Development *130,* 4217-
705 4227.

706 Fitch, S.R., Kimber, G.M., Wilson, N.K., Parker, A., Mirshekar-Syahkal, B., Gottgens, B., Medvinsky, A., Dzierzak, E., and
707 Ottersbach, K. (2012). Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence
708 during embryogenesis. Cell. Stem Cell. *11,* 554-566.

709 Fornerod, M., Ma, J., Noort, S., Liu, Y., Walsh, M.P., Shi, L., Nance, S., Liu, Y., Wang, Y., Song, G.*, et al.* (2021). Integrative
710 Genomic Analysis of Pediatric Myeloid-Related Acute Leukemias Identifies Novel Subtypes and Prognostic Indicators.
711 Blood Cancer. Discov. *2,* 586-599.

712 Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J.,
713 Armstrong, J.*, et al.* (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. *47,*
714 D766-D773.

715 Franzen, O., Gan, L.M., and Bjorkegren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human
716 single-cell RNA sequencing data. Database (Oxford) *2019,* 10.1093/database/baz046.

717 Garcia-Alegria, E., Menegatti, S., Fadlullah, M.Z.H., Menendez, P., Lacaud, G., and Kouskoff, V. (2018). Early Human
718 Hemogenic Endothelium Generates Primitive and Definitive Hematopoiesis In Vitro. Stem Cell. Reports *11,* 1061-1074.

719 Girgin, M.U., Broguiere, N., Mattolini, L., and Lutolf, M.P. (2021). Gastruloids generated without exogenous Wnt activation
720 develop anterior neural tissues. Stem Cell. Reports *16,* 1143-1155.

721 Gotoh, S., Ito, I., Nagasaki, T., Yamamoto, Y., Konishi, S., Korogi, Y., Matsumoto, H., Muro, S., Hirai, T., Funato, M.*, et al.*
722 (2014). Generation of alveolar epithelial spheroids via isolated progenitor cells from human pluripotent stem cells. Stem
723 Cell. Reports *3,* 394-403.

724 Greggio, C., De Franceschi, F., Figueiredo-Larsen, M., and Grapin-Botton, A. (2014). In vitro pancreas organogenesis
725 from dispersed mouse embryonic progenitors. J. Vis. Exp. *(89). doi,* 10.3791/51725.

726 Guibentif, C., Rönn, R.E., Böiers, C., Lang, S., Saxena, S., Soneji, S., Enver, T., Karlsson, G., and Woods, N. (2017).
727 Single-Cell Analysis Identifies Distinct Stages of Human Endothelial-to-Hematopoietic Transition. Cell Reports *19,* 10-19.

728 Gulino, G.M., Bruno, F., Sturiale, V., Brancato, D., Ragusa, D., Tosi, S., Saccone, S., and Federico, C. (2021). From FISH
729 to Hi-C: The Chromatin Architecture of the Chromosomal Region 7q36.3, Frequently Rearranged in Leukemic Cells, Is
730 Evolutionary Conserved. Int. J. Mol. Sci. *22,* 10.3390/ijms22052338.

731 Hall, T.D., Kim, H., Dabbah, M., Myers, J.A., Crawford, J.C., Morales-Hernandez, A., Caprio, C.E., Sriram, P., Kooienga,
732 E., Derecka, M.*, et al.* (2022). Murine fetal bone marrow does not support functional hematopoietic stem and progenitor
733 cells until birth. Nature Communications *13,* 5403.

734 Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos
735 derived from germline colonization by cultured cells. Nature (London) *326,* 292-295.

736 Huch, M., Bonfanti, P., Boj, S.F., Sato, T., Loomans, C.J., van de Wetering, M., Sojoodi, M., Li, V.S., Schuijers, J., Gracanin,
737 A.*, et al.* (2013). Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis.
738 Embo J. *32,* 2708-2721.

739 Huch, M., Dorrell, C., Boj, S.F., van Es, J.H., Li, V.S., van de Wetering, M., Sato, T., Hamer, K., Sasaki, N., Finegold, M.J.,
740 *et al.* (2013). In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. Nature *494,* 247-250.

741 Huch, M., Knoblich, J.A., Lutolf, M.P., and Martinez-Arias, A. (2017). The hope and the hype of organoid research.
742 Development *144,* 938-941.

743 Ingenhag, D., Reister, S., Auer, F., Bhatia, S., Wildenhain, S., Picard, D., Remke, M., Hoell, J.I., Kloetgen, A., Sohn, D.*, et*
744 *al.* (2019). The homeobox transcription factor HB9 induces senescence and blocks differentiation in hematopoietic stem
745 and progenitor cells. Haematologica *104,* 35-46.

746 Jakobsson, L., Franco, C.A., Bentley, K., Collins, R.T., Ponsioen, B., Aspalter, I.M., Rosewell, I., Busse, M., Thurston, G.,
747 Medvinsky, A., Schulte-Merker, S., and Gerhardt, H. (2010). Endothelial cells dynamically compete for the tip cell position
748 during angiogenic sprouting. Nat. Cell Biol. *12,* 943-953.

749 Kapeni, C., Nitsche, L., Kilpatrick, A.M., Wilson, N.K., Xia, K., Mirshekar-Syahkal, B., Chandrakanthan, V., Malouf, C.,
750 Pimanda, J.E., Gottgens, B.*, et al.* (2022). p57Kip2 regulates embryonic blood stem cells by controlling sympathoadrenal
751 progenitor expansion. Blood *140,* 464-477.

752 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of
753 transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14,* R36-r36.

754 Kim, J., Koo, B.K., and Knoblich, J.A. (2020). Human organoids: model systems for human biology and medicine. Nat.
755 Rev. Mol. Cell Biol. *21,* 571-584.

756 Kingsley, P.D., Malik, J., Emerson, R.L., Bushnell, T.P., McGrath, K.E., Bloedorn, L.A., Bulger, M., and Palis, J. (2006).
757 "Maturational" globin switching in primary primitive erythroid cells. Blood *107,* 1665-1672.

758 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik,
759 K.M., Lachmann, A.*, et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic
760 Acids Res. *44,* 90.

761 Lacaud, G., and Kouskoff, V. (2017). Hemangioblast, hemogenic endothelium, and primitive versus definitive
762 hematopoiesis. Exp. Hematol. *49,* 19-24.

763 Lancaster, M.A., and Knoblich, J.A. (2014). Organogenesis in a dish: modeling development and disease using organoid
764 technologies. Science *345,* 1247125.

765 Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson,
766 A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501,* 373-
767 379.

768 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9,* 357-359.

769 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome
770 Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25,* 2078-
771 2079.

772 Lis, R., Karrasch, C.C., Poulos, M.G., Kunar, B., Redmond, D., Duran, J.G.B., Badwe, C.R., Schachterle, W., Ginsberg,
773 M., Xiang, J.*, et al.* (2017). Conversion of adult endothelium to immunocompetent haematopoietic stem cells. Nature *545,*
774 439-445.

775 Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with
776 DESeq2. Genome Biol. *15,* 550-8.

777 Marcelo, K.L., Sills, T.M., Coskun, S., Vasavada, H., Sanglikar, S., Goldie, L.C., and Hirschi, K.K. (2013). Hemogenic
778 endothelial cell specification requires c-Kit, Notch signaling, and p27-mediated cell-cycle control. Dev. Cell. *27,* 504-515.

779 McGrath, K.E., Frame, J.M., Fegan, K.H., Bowen, J.R., Conway, S.J., Catherman, S.C., Kingsley, P.D., Koniski, A.D., and
780 Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells
781 in the Mammalian Embryo. Cell. Rep. *11,* 1892-1904.

782 Medvinsky, A.L., Samoylina, N.L., Muller, A.M., and Dzierzak, E.A. (1993). An early pre-liver intraembryonic source of
783 CFU-S in the developing mouse. Nature *364,* 64-67.

784 Medvinsky, A., and Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. Cell *86,*
785 897-906.

786 Medvinsky, A., Rybtsov, S., and Taoudi, S. (2011). Embryonic origin of the adult hematopoietic system: advances and
787 questions. Development (Cambridge) *138,* 1017-1031.

788 Mikkola, H.K., Fujiwara, Y., Schlaeger, T.M., Traver, D., and Orkin, S.H. (2003). Expression of CD41 marks the initiation
789 of definitive hematopoiesis in the mouse embryo. Blood *101,* 508-516.

790 Moris, N., Edri, S., Seyres, D., Kulkarni, R., Domingues, A.F., Balayo, T., Frontini, M., and Pina, C. (2018). Histone
791 Acetyltransferase KAT2A Stabilizes Pluripotency with Control of Transcriptional Heterogeneity. Stem Cells *36,* 1828-1838.

792 Moris, N., Anlas, K., van den Brink, Susanne C, Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A.,
793 and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature
794 (London) *582,* 410-415.

795 Niakan, K.K., Ji, H., Maehr, R., Vokes, S.A., Rodolfa, K.T., Sherwood, R.I., Yamaki, M., Dimos, J.T., Chen, A.E., Melton,
796 D.A., McMahon, A.P., and Eggan, K. (2010). Sox17 promotes differentiation in mouse embryonic stem cells by directly
797 regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes Dev. *24,* 312-326.

798 Nilsson, T., Waraky, A., Ostlund, A., Li, S., Staffas, A., Asp, J., Fogelstrand, L., Abrahamsson, J., and Palmqvist, L. (2022).
799 An induced pluripotent stem cell t(7;12)(q36;p13) acute myeloid leukemia model shows high expression of MNX1 and a
800 block in differentiation of the erythroid and megakaryocytic lineages. Int. J. Cancer *151,* 770-782.

801 Nobuhisa, I., Osawa, M., Uemura, M., Kishikawa, Y., Anani, M., Harada, K., Takagi, H., Saito, K., Kanai-Azuma, M., Kanai,
802 Y., Iwama, A., and Taga, T. (2014). Sox17-Mediated Maintenance of Fetal Intra-Aortic Hematopoietic Cell Clusters.
803 Molecular and Cellular Biology *34,* 1976-1990.

804 North, T.E., Goessling, W., Peeters, M., Li, P., Ceol, C., Lord, A.M., Weber, G.J., Harris, J., Cutting, C.C., Huang, P.,
805 Dzierzak, E., and Zon, L.I. (2009). Hematopoietic Stem Cell Development Is Dependent on Blood Flow. Cell *137,* 736-748.

806 Papatsenko, D., Darr, H., Kulakovskiy, I., Waghray, A., Makeev, V., MacArthur, B., and Lemischka, I. (2015). Single-Cell
807 Analyses of ESCs Reveal Alternative Pluripotent Cell States and Molecular Mechanisms that Control Self-Renewal. Stem
808 Cell Reports *5,* 207-220.

809 Pearson, S., Cuvertino, S., Fleury, M., Lacaud, G., and Kouskoff, V. (2015). In vivo repopulating activity emerges at the
810 onset of hematopoietic specification during embryonic stem cell differentiation. Stem Cell. Reports *4,* 431-444.

811 Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from
812 single cells using Smart-seq2. Nat. Protoc. *9,* 171-181.

813 Pina, C., May, G., Soneji, S., Hong, D., and Enver, T. (2008). MLLT3 regulates early human erythroid and megakaryocytic
814 cell fate. Cell. Stem Cell. *2,* 264-273.

815 Pina, C., Fugazza, C., Tipping, A.J., Brown, J., Soneji, S., Teles, J., Peterson, C., and Enver, T. (2012). Inferring rules of
816 lineage commitment in haematopoiesis. Nature Cell Biology *14,* 287-294.

817 Pina, C., Teles, J., Fugazza, C., May, G., Wang, D., Guo, Y., Soneji, S., Brown, J., Edén, P., Ohlsson, M., Peterson, C.,
818 and Enver, T. (2015). Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. Cell
819 Reports *11,* 1503-1510.

820 Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., and Zanini, F. (2022). Analysing high-throughput sequencing data in
821 Python with HTSeq 2.0. Bioinformatics *38,* 2943-2945.

822 Ragusa, D., Ciciro, Y., Federico, C., Saccone, S., Bruno, F., Saeedi, R., Sisu, C., Pina, C., Sala, A., and Tosi, S. (2022).
823 Engineered model of t(7;12)(q36;p13) AML recapitulates patient-specific features and gene expression profiles.
824 Oncogenesis *11,* 50-2.

825 Rivron, N.C., Frias-Aldeguer, J., Vrij, E.J., Boisset, J.C., Korving, J., Vivie, J., Truckenmuller, R.K., van Oudenaarden, A.,
826 van Blitterswijk, C.A., and Geijsen, N. (2018). Blastocyst-like structures generated solely from stem cells. Nature *557,* 106-
827 111.

828 Rossi, G., Broguiere, N., Miyamoto, M., Boni, A., Guiet, R., Girgin, M., Kelly, R.G., Kwon, C., and Lutolf, M.P. (2021).
829 Capturing Cardiogenesis in Gastruloids. Cell. Stem Cell. *28,* 230-240.e6.

830 Rossi, G., Giger, S., Hubscher, T., and Lutolf, M.P. (2022). Gastruloids as in vitro models of embryonic blood development
831 with spatial and temporal resolution. Sci. Rep. *12,* 13380-1.

832 Rybtsov, S., Sobiesiak, M., Taoudi, S., Souilhol, C., Senserrich, J., Liakhovitskaia, A., Ivanovs, A., Frampton, J., Zhao, S.,
833 and Medvinsky, A. (2011). Hierarchical organization and early hematopoietic specification of the developing HSC lineage
834 in the AGM region. J. Exp. Med. *208,* 1305-1315.

835 Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., and Medvinsky, A. (2014). Tracing the
836 Origin of the HSC Hierarchy Reveals an SCF-Dependent, IL-3-Independent CD43– Embryonic Precursor. Stem Cell
837 Reports *3,* 489-501.

838 Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters,
839 P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.
840 Nature *459,* 262-265.

841 Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods
842 *9,* 671-675.

843 Silva, A.C., Matthys, O.B., Joy, D.A., Kauss, M.A., Natarajan, V., Lai, M.H., Turaga, D., Blair, A.P., Alexanian, M., Bruneau,
844 B.G., and McDevitt, T.C. (2021). Co-emergence of cardiac and gut tissues promotes cardiomyocyte maturation within
845 human iPSC-derived organoids. Cell. Stem Cell. *28,* 2137-2152.e6.

846 Souilhol, C., Gonneau, C., Lendinez, J.G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S.,
847 Antonchuk, J., Zhao, S., and Medvinsky, A. (2016). Inductive interactions mediated by interplay of asymmetric signalling
848 underlie development of adult haematopoietic stem cells. Nat. Commun. *7,* 10784.

849 Sroczynska, P., Lancrin, C., Pearson, S., Kouskoff, V., and Lacaud, G. (2009). In vitro differentiation of mouse embryonic
850 stem cells as a model of early hematopoietic development. Methods Mol. Biol. *538,* 317-334.

851 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub,
852 T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting
853 genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102,* 15545-15550.

854 Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., da Rocha, E.L., Lu, Y.F., Goettel, J.A., Serrao, E., Rowe, R.G.,
855 Malleshaiah, M.*, et al.* (2017). Haematopoietic stem and progenitor cells from human pluripotent stem cells. Nature *545,*
856 432-438.

857 Sugimura, R., Ohta, R., Mori, C., Li, A., Mano, T., Sano, E., Kosugi, K., Nakahata, T., Niwa, A., Saito, M.K., and Torisawa,
858 Y.S. (2020). Biomimetic aorta-gonad-Mesonephros-on-a-Chip to study human developmental hematopoiesis. Biomed.
859 Microdevices *22,* 34-2.

860 Swiers, G., Baumann, C., O'Rourke, J., Giannoulatou, E., Taylor, S., Joshi, A., Moignard, V., Pina, C., Bee, T., Kokkaliaris,
861 K.D.*, et al.* (2013). Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nature
862 Communications *4,* 2924.

863 Taketani, T., Taki, T., Sako, M., Ishii, T., Yamaguchi, S., and Hayashi, Y. (2008). MNX1-ETV6 fusion gene in an acute
864 megakaryoblastic leukemia and expression of the MNX1 gene in leukemia and normal B cell lines. Cancer Genet.
865 Cytogenet. *186,* 115-119.

866 Taoudi, S., Gonneau, C., Moore, K., Sheridan, J.M., Blackburn, C.C., Taylor, E., and Medvinsky, A. (2008). Extensive
867 Hematopoietic Stem Cell Generation in the AGM Region via Maturation of VE-Cadherin+CD45+ Pre-Definitive HSCs. Cell
868 Stem Cell *3,* 99-108.

869 Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L., and Mi, H. (2022). PANTHER: Making genome-
870 scale phylogenetics accessible to all. Protein Science *31,* 8-22.

871 Torisawa, Y.S., Spina, C.S., Mammoto, T., Mammoto, A., Weaver, J.C., Tat, T., Collins, J.J., and Ingber, D.E. (2014). Bone
872 marrow-on-a-chip replicates hematopoietic niche physiology in vitro. Nat. Methods *11,* 663-669.

873 Turner, D.A., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., Glodowski, C.R., Hayward, P.C., Collignon,
874 J., Gustavsen, C., Serup, P.*, et al.* (2017). Anteroposterior polarity and elongation in the absence of extraembryonic tissues
875 and spatially localised signalling in Gastruloids , mammalian embryonic organoids. Development (Cambridge) *144,* 3894-
876 3906.

877 van den Brink, S C, Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivie, J., Baillie-Johnson, P., Nichols, J.,
878 Sonnen, K.F., Martinez Arias, A., and van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal
879 somitogenesis in gastruloids. Nature *582,* 405-409.

880 van den Brink, S C, Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.K., Nowotschin, S., Turner, D.A., and Martinez Arias,
881 A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem
882 cells. Development *141,* 4231-4242.

883 van den Brink, S C, and van Oudenaarden, A. (2021). 3D gastruloids: a novel frontier in stem cell-based in vitro modeling
884 of mammalian gastrulation. Trends Cell Biol. *31,* 747-759.

885 Vink, C.S., Calero-Nieto, F.J., Wang, X., Maglitto, A., Mariani, S.A., Jawaid, W., Gottgens, B., and Dzierzak, E. (2020).
886 Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells. Cell. Rep. *31,*
887 107627.

888 Wang, D., Tanaka-Yano, M., Meader, E., Kinney, M.A., Morris, V., Lummertz da Rocha, E., Liu, N., Liu, T., Zhu, Q., Orkin,
889 S.H.*, et al.* (2022). Developmental maturation of the hematopoietic system controlled by a Lin28b-let-7-Cbx2 axis. Cell
890 Reports (Cambridge) *39,* 110587.

891 Waraky, A., Ostlund, A., Nilsson, T., Weichenhan, D., Lutsik, P., Bahr, M., Hey, J., Adamsson, J., Hamdy, M., Morsy, A.*,*
892 *et al.* (2022). Mnx1 Induces Leukemia Transformation Through Altering Histone Methylation in a Model of Pediatric t(7;12)
893 Acute Myeloid Leukemia. BiorXiv

894 Wildenhain, S., Ingenhag, D., Ruckert, C., Degistirici, O., Dugas, M., Meisel, R., Hauer, J., and Borkhardt, A. (2012).
895 Homeobox protein HB9 binds to the prostaglandin E receptor 2 promoter and inhibits intracellular cAMP mobilization in
896 leukemic cells. J. Biol. Chem. *287,* 40703-40712.

897 Wildenhain, S., Ruckert, C., Rottgers, S., Harbott, J., Ludwig WD FASchuster, F R, Schuster, F.R., Beldjord, K., Binder,
898 V., Slany, R., Hauer, J., and Borkhardt, A. (2010). Expression of cell-cell interacting genes distinguishes HLXB9/TEL from
899 MLL-positive childhood acute myeloid leukemia. Leukemia *24,* 1657-1660.

900 Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L.,
901 Kropiwnicki, E., Jagodnik, K.M., Jeon, M., and Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. Curr.
902 Protoc. *1,* e90.

903 Yu, L., Wei, Y., Duan, J., Schmitz, D.A., Sakurai, M., Wang, L., Wang, K., Zhao, S., Hon, G.C., and Wu, J. (2021).
904 Blastocyst-like structures generated from human pluripotent stem cells. Nature *591,* 620-626.

905 Zhu, Q., Gao, P., Tober, J., Bennett, L., Chen, C., Uzun, Y., Li, Y., Howell, E.D., Mumau, M., Yu, W.*, et al.* (2020).
906 Developmental trajectory of prehematopoietic stem cell formation from endothelium. Blood *136,* 845-856.

907

908 **KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse c-Kit (CD117) clone 2B8 - APC-Cy7	Biolegend	Cat. #105825
Anti-mouse CD41 clone MWReg30 - PE-Dazzle594	Biolegend	Cat. #133935
Anti-mouse CD45 clone 30F11 - APC	Biolegend	Cat. #103111
Anti-mouse CD144 (VE-cadherin) clone BV13 - APC	Biolegend	Cat. #138011
Anti-mouse CD43 clone S11 - PE	Biolegend	Cat. #143205
CD11b anti-mouse/human clone M1/70 - biotin	Biolegend	Cat. #101203
Mouse Anti-Mouse CD45.2 clone 104 - PE	BD Biosciences	Cat. #560695
Rat Anti-Mouse CD31 clone MEC13.3 - biotin	BD Biosciences	Cat. #553371
Goat Anti-Mouse c-Kit/CD117	R&D Systems	Cat. #AF1356SP
Bacterial and virus strains		
E. coli: NEB 5-alpha Competent *E. coli*	New England Biolabs (NEB)	Cat. #C2987H
Chemicals, peptides, and recombinant proteins		
Murine LIF	Peprotech	Cat. #250-02
StemMACS PD0325901	Miltenyi Biotec	Cat. #130-106-5411
Chiron (CHIR99021)	Biogems	Cat. #2520691
Activin A PLUS	QKine	Cat. #Qk005
Murine VEGF 165	Peprotech	Cat. #450-32
Murine FGF-basic	Peprotech	Cat. #450-33
Murine Sonic Hedgehog (Shh)	Peprotech	Cat. #315-22
Murine TPO	Peprotech	Cat. #315-14
Murine Flt3-Ligand	Peprotech	Cat. #250-31L
Murine SCF	Peprotech	Cat. #250-03
N2B27 medium (NDiff 227)	Takara Bio	Cat. #Y40002
Critical commercial assays		
Mouse Methylcellulose Complete Media	R&D Systems	Cat. #HSC007
Deposited data		
Single cell RNA-seq of hemogenic gastruloids at timepoints	This paper	ArrayExpress: E-MTAB-12148
Bulk RNA-seq of hemogenic gastruloids with MNX1 overexpression at 216 h	This paper	ArrayExpress: E-MTAB-12173
Single cell RNA-seq of hemogenic gastruloids with MNX1 overexpression at 216 h, sorted CD45+/CD41+	This paper	ArrayExpress: E-MTAB-12149
Experimental models: Cell lines		
Mus musculus: Flk-1-GFP mouse embryonic stem cell line	Alfonso Martinez Arias Lab	(Jakobsson et al., 2010)
Mus musculus: Sox17-GFP mouse embryonic stem cell line	Alfonso Martinez Arias Lab	(Niakan et al., 2010)
Mus musculus: T/Bra::GFP mouse embryonic stem cell line	Alfonso Martinez Arias Lab	(Fehling et al., 2003)
Mus musculus: E14Tg2a mouse embryonic stem cell line	MMRRC, University of California Davis, US	(Hooper et al., 1987)
Homo sapiens: HEK293T	ATCC	CRL-3216
Oligonucleotides		
MNX1 (forward 5-GTTCAAGCTCAACAAGTACC-3; reverse 5- GGTTCTGGAACCAAATCTTC-3)	(Gulino et al., 2021)	n/a
Ppia (forward 5-TTACCCATCAAACCATTCCTTCTG-3; reverse 5-AACCCAAAGAACTTCAGTGAGAGC-3)	(Moris, N. et al., 2018)	n/a
Recombinant DNA		

pWPT-LSSmOrange-PQR	Ghevaert Lab, WT-MRC Cambridge Stem Cell Institute, UK	(Dalby et al., 2018)
pWPT-LSSmOrange-PQR-MNX1	Cloned by Biomatik Corporation (Kitchener, Canada)	n/a
Software and algorithms		
GSEA software v4.2.3	(Subramanian et al., 2005)	https://www.gsea-msigdb.org/gsea/index.jsp
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
Bowtie2	(Langmead and Salzberg, 2012)	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools	(Li et al., 2009)	http://samtools.sourceforge.net/
Tophat2	(Kim, D. et al., 2013)	https://ccb.jhu.edu/software/tophat/index.shtml
HTseq	(Anders et al., 2015)	https://htseq.readthedocs.io/en/master/
DEseq2	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ExpressAnalyst	n/a	https://www.expressanalyst.ca/
PANTHER	(Thomas et al., 2022)	http://www.pantherdb.org/
EnrichR	(Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021)	https://maayanlab.cloud/Enrichr/
Panglao BP	(Franzen et al., 2019)	https://panglaodb.se
Descartes	(Cao et al., 2020)	https://descartes.brotmanbaty.org/
TARGET (Therapeutically Applicable Research to Generate Effective Treatments) Database	https://ocg.cancer.gov/programs/target https://xenabrowser.net/	GDC TARGET-AML
Single-cell RNA sequencing analysis scripts	DockerHub: dsblab/single_cell_analysis:0.5	https://github.com/dsb-lab/blood_gastruloids
Other		
CELLSTAR cell-repellent cell culture plate, 96 well, U-bottom	Greiner (Bio-One)	Cat. #650970

909

910 **METHODS**

911 **Cell culture**

912 Flk-1-GFP (Jakobsson et al., 2010), Sox17-GFP (Niakan et al., 2010), T/Bra::GFP (Fehling et al.,
913 2003), and E14Tg2A (Hooper et al., 1987) mouse embryonic stem cells (mES) lines were cultured
914 in ES+LIF medium in gelatinized (0.1% gelatin) with daily medium change, as previously described
915 (Turner et al., 2017). The ES+LIF medium contained 500 ml Glasgow MEM BHK-21 (Gibco), 50 ml
916 of fetal bovine serum (FBS, Embryonic Stem Cells tested, biosera, Nuaillé, France), 5 ml GlutaMAX
917 supplement (Gibco), 5 ml MEM Non-Essential Amino Acids (Gibco), 5 ml Sodium pyruvate solution
918 (100 mM, Gibco), and 1 ml 2-Mercaptoethanol (50mM, Gibco). Murine LIF (Peprotech) was added
919 at 1000 U/ml. HEK293T cells were grown in Dulbecco's modified Eagle medium (Gibco)
920 supplemented with 10% FBS (Gibco) and 1% penicillin/streptomycin (Gibco). All cultures were kept
921 at 37°C and 5% CO2.

922

923 **Hemogenic gastruloids assembly, culture, and dissociation**

924 mES were maintained in ES+LIF medium and transferred to 2i+LIF (containing Chiron and MEK
925 inhibitor PD03) for 24 hours prior to the assembly into gastruloids. 250 cells were seeded in each
926 well of a U-bottom, cell-repellent 96-well plate (Greiner Bio-One, Stonehouse, UK) in 40 µl of N2B27
927 medium (Takara Bio). The plate was centrifuged at 750 rpm for 2 minutes to promote deposition and
928 aggregation of the cells and was then incubated at 37°C, 5% CO_2 for 48 hours. After 48 hours, 150
929 µl of N2B27 medium supplemented with 100 ng/ml Activin A (QKine, Cambridge, UK) and 3 µM
930 chiron (Peprotech) was added to each well. At 72 hours, 150 µl of medium were removed, without
931 disrupting the gastruloids in the wells. 100 µl of N2B27 with 5 ng/ml of Vegf and Fgf2 each
932 (Peprotech) were added to each well. From 72 h to 144 h, each day 100 µl of medium were removed
933 and replaced with N2B27 + Vegf + Fgf2. At 144 h, the medium was further supplemented with Shh
934 at 20 ng/ml. From 168 h to 216 h, the medium was N2B27 + 5 ng/ml Vegf, plus 20 ng/ml mTpo, 100
935 ng/ml mFlt3l, and 100 ng/ml mScf (Peprotech). To dissociate cells from the gastruloid structures,
936 medium was removed and individual gastruloids were collected using a pipette and precipitated at

937 the bottom of a microcentrifuge tube. The remaining medium was aspirated and the bulk of
938 gastruloids was washed in PBS. 50 µl of TrypLE express was added to pelleted gastruloids to be
939 incubated at 37°C for 2 minutes to dissociate cells.
940

941 **Methylcellulose colony forming assays (CFC)**

942 Disassembled gastruloids cells were plated on Mouse Methylcellulose Complete Media (R&D
943 Systems). Cells were first suspended in Iscove's Modified Dulbecco's Medium (IMDM, Gibco) with
944 20% FBS (Gibco) before addition to the methylcellulose medium. Cells were plated in duplicate 35
945 mm dishes with $2x10^5$ cells/plate. Plates were incubated at 37°C and 5% CO2 for 10 days, when
946 colonies were scored. For serial replating experiments, cells in methylcellulose were collected and
947 washed in phosphate buffer saline (PBS) to achieve single-cell suspensions and replated as
948 described above.
949

950 **Immunofluorescence**

951 Immunostaining of whole gastruloids was performed as described before (Baillie-Johnson et al.,
952 2015). Briefly, gastruloids were fixed in 4% paraformaldehyde (PFA) dissolved in PBS for 4 hours at
953 4°C on orbital shaking and permeabilised in PBSFT (10% FBS and 0.2% Triton X-100), followed by
954 one hour blocking in PBSFT at 4°C on orbital shaking. Antibody dilutions were made in PBSFT at
955 1:200 for primary and 1:500 for secondary antibodies. Antibody incubations were performed
956 overnight at 4°C on orbital shaking, and subjected to optical clearing overnight in ScaleS4 clearing
957 solution. Individual gastruloids were then mounted on glass coverslips by pipetting as droplets in
958 ScaleS4 and DAPI nuclear stain.
959

960 **Imaging**

961 Images of cultured gastruloids and CFC plates were captured using the Cytation 5 Cell Imaging
962 Multi-Mode Reader (Biotek) plate reader using bright field and FITC channels. ImageJ (Schneider et
963 al., 2012) was used for gastruloid size quantification. Confocal microscopy was performed on

964 LSM700 on a Zeiss Axiovert 200 M with Zeiss EC Plan-Neofluar 10x/0.30 M27 and Zeiss LD Plan-
965 Neofluar 20x/0.4 M27 objective lens.
966

967 **Lentiviral vector packaging and transduction**

968 The lentiviral overexpression vector pWPT-LSSmOrange-PQR was used to clone the *MNX1* gene
969 cDNA. The viral packaging vectors pCMV and pMD2.G, described in (Pina, C. et al., 2008), were
970 used to assemble lentiviral particles using HEK293T cells via transfection using TurboFect Reagent
971 (Invitrogen). Transduction of mES cells was performed overnight by addition of lentivirus to culture
972 medium and washed the following day (Moris et al., 2018). Transduced cells were sorted for positivity
973 to LSSmOrange.
974

975 **Flow cytometry**

976 Surface cell marker analysis was performed by staining using the antibodies listed in Key Resources
977 Table. Disassembled gastruloids cells were resuspended in PBS, 2% FBS and 0.5 mM EDTA and
978 stained at a dilution of 1:100 for primary antibodies for 20 minutes at 4°C. When indicated,
979 streptavidin was added at a dilution of 1:200. Analysis was performed on ACEA Novocyte (Agilent)
980 or AttuneNxT (Thermo) analyzers, using the respective software packages. Cell sorting was
981 performed using a CS&T calibrated BDFACS Aria III system (488 nm 40 mW, 633 nm 20 mW, 405nm
982 30 mW, and 561 nm 50 mW), set with the 100μm nozzle at 20PSI and a 4-way purity mask. Single-
983 cell deposition in 96-well plates was performed using single-cell sorting mode. Intact cells were gated
984 on FSC-A vs SSC-A plot, followed by doublet exclusion on FSC-A vs FSC-H and SSC-A vs SSC-H,
985 prior to gating on fluorescent parameters for the markers described in the results.
986

987 **Single cell RNA sequencing of time-resolved hemogenic gastruloids**

988 Gastruloids were collected at different timepoints of the protocol, disassembled and FACS deposited
989 into 96-well plates, either as unsorted (global) or sorted by CD45, CD41 and c-Kit/ScaI markers
990 (sorted). RNA from the cells were extracted from single cells using Smart-seq2 technology at 500Kb

991 and depth of 151 bases. Sequencing reads were quality-checked using FastQC (v0.11.9). We
992 trimmed the samples using trimGalore! (0.6.6) with a cutoff of 30, clipping 15 base pairs and retaining
993 reads of more than 100 bases. Alignment was performed on STAR (2.7.8a) and *Mus muluscus*
994 annotations from GENCODE vM26. Aligned BAM files were annotated using featureCounts (v2.0.1)
995 and count matrices were computed in python by directly accessing the BAM files with gtfparse
996 packages and collapsing lanes.
997

998 **Single cell RNA sequencing of MNX1 overexpressing hemogenic gastruloids at endpoint**

999 MNX1 overexpressing gastruloids and empty vector (EV) control gastruloids were collected at 216
1000 h, sorted by positivity to CD45 or CD41, and deposited into 96-well plates for sequencing. Processing
1001 of reads was performed as described above, but retaining reads over 75 bases to account for the
1002 sequencing depth of 101 bases.
1003

1004 **Single cell RNA sequencing analysis**

1005 For quality control, based on the histogram of counts and multimodality distributions, we set a
1006 minimum count threshold of 200000 counts, a minimum threshold of 1000 expressed genes, and a
1007 maximum threshold of 20% mitochondrial fraction per cell. We performed the same procedure of the
1008 second dataset with a more restrictive threshold of 400000 counts and similar expressed genes and
1009 mitochondrial thresholds. Cells that did not pass the quality control metrics were omitted from
1010 analysis. We normalized the cells to the mean count number per dataset and applied a plus-one-log
1011 transformation of the data before proceeding to the downstream analysis.
1012 Dimensionality reduction was performed on feature selection of the gene space using the function
1013 *scanpy.highly_varying_genes* with default parameters. Selection of principal components in principal
1014 component analysis (PCA) was performed by heuristic elbow method. Nearest neighbor analysis
1015 was constructed by KNN graphs using a correlation metric and 10 nearest neighbors. Data was
1016 projected for low dimensional visualization using the UMAP algorithm with default parameters as
1017 implemented in *scanpy.tl.umap*. We used the leiden algorithm as implemented in *scanpy.tl.leiden* to

1018 partition the data into clusters. In order to assess the election of the resolution parameter we used
1019 *Newman-Girvan* modularity as a metric of clustering quality. Differential expressed genes were
1020 computed comparing each cluster against the rest using the Wilcoxon test with Benjamini-Hochberg
1021 correction.
1022

1023 **Annotation and projection to additional datasets**

1024 Raw counts matrices from Fadlullah et al. (2022) and Vink et al. (2020) were downloaded and
1025 processed following the same pipeline as described above for scRNA-seq of gastruloids to generate
1026 UMAP and clustering. We implemented the scmap algorithm to compare our scRNA sequencing of
1027 gastruloids with the available datasets. We reduced the dimensionality of the space by selecting
1028 highly varying genes from the annotated dataset. Then, we constructed a KNN classifier with
1029 correlation metric and computed the nearest neighbors of the target data. If neighbors with
1030 correlation metrics below 0.7 default standards, the projected cells were not projected onto any cell
1031 from the annotated dataset. To visualize the cells over the UMAP plots of the other datasets, we
1032 constructed a KNN regressor with three neighbors and a correlation metric.
1033

1034 **Bulk RNA sequencing**

1035 Total RNA was extracted from disassembled gastruloid cells at 216 hours. Sequencing was
1036 performed on NovaSeq PE150 platform, at 20M paired-end reads per sample. Tophat2 with Bowtie2
1037 were used to map paired-end reads to the reference *Mus musculus* genome build GRCm39 (Kim et
1038 al., 2013; Langmead and Salzberg, 2012) GENCODE Release M30 (Frankish et al., 2019) was used
1039 as the reference mouse genome annotation. Aligned reads were filtered by quality using samtools
1040 (Li et al., 2009) with a minimum threshold set at 30 (q30). Transcript assembly and quantification
1041 was achieved using htseq (Putri et al., 2022). Differential expression between sample and control
1042 was performed by collapsing technical replicates for each condition using Deseq2 (Love et al., 2014)
1043 in R environment (*Deseq2* library v 1.32.0). The differential expression was expressed in the form of
1044 log2 fold change and filtered by false discovery rate (FDR) of 0.1.
1045

1046 **Real-time polymerase chain reaction (qPCR)**

1047 Extracted RNA was reverse-transcribed into complementary DNA (cDNA) using High-Capacity RNA-
1048 to-cDNA Kit (Applied Biosystems). QPCR was performed using FastGene 2x IC Green Universal
1049 qPCR Mix (Nippon Genetics, Duren, Germany) using primers for human *MNX1* (forward 5-
1050 GTTCAAGCTCAACAAGTACC-3; reverse 5- GGTTCTGGAACCAAATCTTC-3) (Gulino et al., 2021)
1051 and *Ppia* for endogenous control (forward 5- TTACCCATCAAACCATTCCTTCTG-3; reverse 5-
1052 AACCCAAAGAACTTCAGTGAGAGC-3) (Moris et al., 2018). Differential gene expression was
1053 calculated using the delta delta Ct (ΔΔCt) method.

1054

1055 **Gene list enrichment analyses**

1056 Gene ontology (GO) analysis was performed in ExpressAnalyst (available at
1057 www.expressanalyst.ca) using the PANTHER Biological Process (BP) repository. GO terms and
1058 pathways were filtered by false discovery rate (FDR) with a cut-off of ≤ 0.1 for meaningful association.
1059 EnrichR (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021) was used for cell type analysis
1060 using the Panglao DB (Franzen et al., 2019) and Descartes (Cao et al., 2020) databases using a
1061 FDR threshold of ≤ 0.1 or p value ≤ 0.01, where specified, as well as transcription factor (TF) binding
1062 site enrichment using the ENCODE and ChEA Consensus TFs from ChIP database.

1063

1064 **Gene Set Enrichment Analysis (GSEA)**

1065 Custom gene signatures were used as gene sets for GSEA analysis (Subramanian et al., 2005) on
1066 the GSEA software v4.2.3 on RNA sequencing expression values in counts units GSEA was ran in
1067 10000 permutations on gene set using the weighted Signal2Noise metric. Enrichment metrics are
1068 shown as normalized enrichment score (NES) and filtered by FDR ≤ 0.05. Leading edge genes
1069 (LEGs) are genes with a “Yes” values for core enrichment. For AML patient analysis, clinical
1070 phenotype and expression data (in counts units) were extracted from the GDC TARGET-AML
1071 cohorts in the Therapeutically Applicable Research to Generate Effective Treatments project
1072 (TARGET, https://ocg.cancer.gov/programs/target), downloaded from the University of California

1073 Santa Cruz (UCSC) Xena public repository (last accessed 31st August 2022). Patient samples were
1074 selected according to the reported karyotype to include t(7;12), inv(16), *MLL*, normal karyotype, and
1075 t(8;21). GSEA was performed comparing RNA sequencing counts of t(7;12) samples against pooled
1076 AML subtypes (inv(16), MLL, normal karyotype and t(8;21)) as "other AML".
1077

1078 **Lineage analysis from scRNA-seq**

1079 Genes with detectable expression in 216-hour CD41/CD45+ single MNX-OE or EV gastruloid cells
1080 were manually scanned for lineage-affiliated markers and regulators conventionally associated with
1081 the endothelial and hematopoietic lineages (Guibentif et al., 2017; Pina, Cristina et al., 2012; Pina,
1082 Cristina et al., 2015; Swiers et al., 2013). Erythroid genes were: *Gata1*, *Klf1*, *Zfpm1*, *Tal1*, *Gfi1b* and
1083 *Nfe2* (transcription factors, TF), *Epor* (growth factor receptor, GFR), and *Epb42*, *Slc4a1*/Band3,
1084 *Gypa* and *Gypc* (membrane-associated proteins, MAP), in addition to alpha (*Hba*) and beta globin
1085 (*Hbb*) chains. Myeloid genes were *Spi1*, *Cebpa*, *Cebpe* and *Egr2* (TF), *Csf1r*, *Csf2ra*, *Csf2rb*, *Csf3r*
1086 and *Il3ra* (GFR), *Fcgr1*, *Fcgr2b*, *Anpep*/Cd13, *Cd14*, *Cd33*, *Ly6c*, *Ly6g* and *Ly6a* (MAP), and
1087 enzymes *Mpo*, *Lyz1* and *Lyz2*. Lymphoid genes were *Notch1*, *Gata3*, *Ebf1*, *Ebf2* and *Ikfz1* (TF),
1088 *Il2ra*, *Il2rg* and *Il7r* (GFR), *Cd5*, *Cd7*, *Cd8a* and *Cd79a* (MBP), as well as surrogate (*Igll1*, *Vpreb1*)
1089 and mature Ig chains (*Ighm*, *Ighd*, *IghJ*); variable regions of *Igl* and *Igh* chains were considered
1090 separately as present or absent. Endothelial (*Flt1*, *Kdr*/Flt2, *Flt4*, *Cdh5*, *Esam*, *Epas1* and *Sox17*),
1091 stem/progenitor (*Runx1*, *Cbfb*, *Myb* and *Gata2*, in addition to the HSC/erythroid-associated *Mllt3*),
1092 and megakaryocytic genes (*Mpl*, *Fli1*, *Itga2b*/Cd41, *Itgb3*/Cd61) were also annotated. Lineage
1093 signatures were called if 2 or more lineage genes were expressed. Erythroid lineage was separated
1094 into marker/regulator (Reg) and Hemoglobin signatures, the latter requiring at least one *Hbb* and
1095 one *Hba* chain. Priming for multiple lineages (erythroid, myeloid or lymphoid ± endothelial) was
1096 quantified separately based on co-existence of lineage signatures. Megakaryocytic lineages were
1097 not considered for multi-lineage priming due to systematic co-expression of markers in progenitor
1098 and erythroid cells. Co-expression of endothelial and progenitor signatures were called endoHaem
1099 if no or a single lineage signature were present.

1100

1101 **Statistical analysis**

1102 Experiments were performed at least in triplicates, unless specified otherwise. Data are plotted to
1103 include standard deviation (+/- SD) between replicates. Statistical significance was set at a threshold
1104 of p value < 0.05. Statistical analysis was performed in R environment (version 4.1.3) or using
1105 GraphPad Prism 8.0 software.

1106

1107 **Data availability**

1108 Raw data as well as processed count matrices and post-processed files from single-cell RNA-seq
1109 for the time-resolved data is available at E-MTAB-12148. Single-cell RNA-seq for the MNX1
1110 overexpression experiment is available at Array Express with accession code E-MTAB-12149. Bulk
1111 RNA-seq of MNX1 overexpressing gastruloids is available at Array Express with accession code E-
1112 MTAB-12173. The post-processing was performed in Python on DockerHub:
1113 dsblab/single_cell_analysis:0.5. Scripts are available in https://github.com/dsb-
1114 lab/blood_gastruloids and Zenodo (https://doi.org/10.5281/zenodo.7053423). The results published
1115 here are partly based upon data generated by the Therapeutically Applicable Research to Generate
1116 Effective Treatments (TARGET) (https://ocg.cancer.gov/programs/target) initiative, of the Acute
1117 Myeloid Leukemia (AML) cohort GDC TARGET-AML. The data used for this analysis are available
1118 at https://portal.gdc.cancer.gov/projects and https://xenabrowser.net/.