Dissecting infant leukemia developmental origins with a hemogenic gastruloid model

Denise Ragusa[1,2], Chun-Wai Suen[3*], Gabriel Torregrosa-Cortés[4*], Fabio Pastorino[5*], Ayona Johns[1,2], Ylenia Cicirò[1,2], Liza Dijkhuis[1,12], Susanne van den Brink[6,7], Michele Cilli[8], Connor Byrne[1,13], Giulia-Andreea Ionescu[1,14], Joana Cerveira[9], Kamil R. Kranc[10,11], Victor Hernandez-Hernandez[1,2], Mirco Ponzoni[5], Anna Bigas[6,7], Jordi Garcia-Ojalvo[4], Alfonso Martinez Arias[4], Cristina Pina[1,2,15]¶*

[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]Laboratory of Experimental Therapies in Oncology, IRCCS Istituto Giannina Gaslini, Genova, Italy; [6]Program in Cancer Research, Hospital de Mar Research Institute, CIBERONC, Barcelona, Spain; [7]Josep Carreras Leukemia Research Institute, Barcelona, Spain; [8]Animal Facility, IRCCS Policlinico San Martino, Genova, Italy; [9]Department of Pathology, University of Cambridge, UK; [10]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, UK; [11]Centre for In Vivo Modelling, Institute of Cancer Research, Sutton, UK.

[12]Current address: Sanquin Research, Landsteiner Laboratory, Amsterdam University Medical Center, Amsterdam, The Netherlands

[13]Current address: Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

[14]Current address: Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, Henry Wellcome building, CF14 4XN, UK

[15]Current address: Sanquin Research, Amsterdam, The Netherlands

*These authors contributed equally to this work

¶Lead contact. Correspondence: c.pina@sanquin.nl. Sanquin Research, Plesmanlaan 125, 1066 CX, Amsterdam, The Netherlands.

ABSTRACT

Current in vitro models of developmental blood formation lack spatio-temporal accuracy and weakly replicate successive waves of hematopoiesis. Herein, we describe a mouse embryonic stem cell (SC)-derived 3D hemogenic gastruloid (haemGx) that captures multi-wave blood formation, progenitor specification from hemogenic endothelium (HE), and generates hematopoietic progenitors capable of short-term engraftment of immunodeficient mice upon maturation in an in vivo niche. We took advantage of the haemGx model to interrogate the origins of infant acute myeloid leukemia (infAML). We focused on MNX1-driven leukemia, representing the commonest genetic abnormality unique to the infant group. Enforced MNX1 expression in haemGx promotes the expansion and in vitro transformation of yolk sac-like erythroid-myeloid progenitors at the HE-to-hematopoietic transition to faithfully recapitulate patient transcriptional signatures. By combining phenotypic, functional and transcriptional profiling, including at the single-cell level, we establish the haemGx as a useful new model for the study of normal and leukemic embryonic hematopoiesis.

INTRODUCTION

Blood formation in the embryo is a multi-stage process that develops across multiple cellular niches, reflecting a complex hierarchy of tissue interactions. It involves sequential production of distinct cell types in so-called hematopoietic waves, separated in time and space (Lacaud and Kouskoff, 2017; Costa, Kouskoff and Lacaud, 2012; Medvinsky, Rybtsov and Taoudi, 2011; Dzierzak and Bigas, 2018). The first wave of hematopoiesis produces unipotent red blood cell and macrophage precursors in the yolk sac (YS) from an angioblast precursor which can also form endothelium; it occurs at mouse embryo day (E)7-E7.5 (Fujimoto *et al.,* 2001; McGrath and Palis, 2005; Moore and Metcalf, 1970). The subsequent wave, known as definitive, produces: first, oligopotent erythro-myeloid progenitors (EMPs) in the YS (E8-E8.5); and later myelo-lymphoid progenitors (MLPs – E9.5-E10), multipotent progenitors (MPPs – E10-E11.5), and hematopoietic stem cells (HSCs – E10.5-E11.5), in the aorta-gonad-mesonephros (AGM) region of the embryo proper (Ivanovs *et al.,* 2011; McGrath *et al.,* 2015; Palis *et al.,* 1999; Medvinsky and Dzierzak, 1996; Medvinsky, Rybtsov and Taoudi, 2011). Definitive blood waves are characteristically specified from a specialised hemogenic endothelium (HE) through a process of endothelial-to-hematopoietic transition (EHT) (Zovein *et al.,* 2008; Li, W. *et al.,* 2005). HSC specification from HE is the last event in hematopoietic development and occurs in characteristic cell clusters found on the ventral wall of the dorsal aorta (Ivanovs *et al.,* 2011; Boisset *et al.,* 2010). All blood cell types enter circulation from their original locations and migrate to the fetal liver (FL) (Ciriza *et al.,* 2013; Ghiaur *et al.,* 2008). Most embryonic blood progenitors are transient and sustain blood production exclusively during embryonic development, through differentiation. In contrast, HSCs make little or no contribution to embryonic and fetal blood: instead, they proliferate (i.e. self-renew) to expand their numbers. At least a subset of HSC migrates to the bone marrow (BM) niche at the end of gestation and balances self-renewal and differentiation to maintain blood production throughout the entire post-natal life (Bowie *et al.,* 2006; Kikuchi and Kondo, 2006; Ganuza *et al.,* 2022; Yokomizo *et al.,* 2022).

Blood progenitors and HSCs can be targeted by genetic alterations that confer clonal advantage and/or drive initiation and development of leukemia. Leukemia is the most common malignancy and the most common cause of cancer-related death in the paediatric age group (Steliarova-Foucher *et al.,* 2017; Dong *et al.,* 2020). Several

74 paediatric leukemias originate *in utero*, and a subset of these are driven by mutations
75 exclusive to the paediatric group (Masetti *et al.,* 2015; Cazzola *et al.,* 2021). Such
76 paediatric-exclusive mutations target blood cell types restricted to development or
77 depend on signals that are specific to embryonic hematopoietic niches, effectively
78 defining distinct biological entities to adult malignancies (Cazzola *et al.,* 2021; Bolouri
79 *et al.,* 2018; Chaudhury *et al.,* 2018). Dissection of these forms of developmental
80 leukemia is confounded by the inability to robustly identify and isolate their embryonic
81 cells of origin, and/or to fully recapitulate the disease in post-natal cells, which leads
82 to failure in generating relevant targeted therapies (Alexander and Mullighan, 2021).
83 One such example is t(7;12)(q36;p13) acute myeloid leukemia (AML) driven by
84 rearrangement and overexpression of the *MNX1* gene locus: *MNX1*-rearranged
85 (MNX1-*r*) AML is restricted to infants under the age of 2 and carries a high risk of
86 relapse and poor survival (Ragusa *et al.,* 2023). *MNX1* is a developmental gene
87 required for specification of motor neurons and pancreatic acinar cells (Thaler *et al.,*
88 1999; Harrison *et al.,* 1999). MNX1 is not known to participate in embryonic
89 hematopoiesis, however it has been discordantly described to be expressed in some
90 subsets of adult hematopoietic cells (Deguchi and Kehrl, 1991; Nagel *et al.,* 2005;
91 Wildenhain *et al.,* 2012; Taketani *et al.,* 2008). It is ectopically expressed in AML cells
92 carrying one of a variable set of t(7;12) translocations which place the normally silent
93 *MNX1* gene under the control of *ETV6* regulatory elements (Weichenhan *et al.,* 2023;
94 Bousquets-Muñoz *et al.,* 2024; Ragusa *et al.,* 2023; Ballabio *et al.,* 2009). *ETV6* is
95 required for hematopoietic cell specification from HE via VEGF signalling (Ciau-Uitz *et*
96 *al.,* 2010), potentially positioning *MNX1* within a critical embryonic hemogenic
97 regulatory network.

98 Faithful *in vitro* recapitulation of developmental blood cell types and their embryonic
99 niches has the potential to dissect the cellular and molecular mechanisms presiding
100 at initiation of developmental leukemia, reveal prognostic biomarkers, and indicate
101 therapeutic vulnerabilities which are otherwise challenging to test in the embryo.
102 Despite widespread use of embryonic stem (ES) cell and induced pluripotent stem
103 (iPS) cell-based models of blood specification, a system which recapitulates the spatial
104 and temporal complexity of embryonic blood formation within the respective time-
105 dependent niches is still lacking (Dijkhuis *et al.,* 2023). In the last few years, gastruloid
106 models have emerged as robust avatars of early development, self-organising

107 processes of symmetry breaking, elongation, multi-axis formation, somitogenesis and
108 early organogenesis, with remarkable similarity to embryo processes in space and
109 time (van den Brink, Susanne C and van Oudenaarden, 2021; Beccari *et al.,* 2018;
110 van den Brink, Susanne C *et al.,* 2020). Gastruloids are grown individually in a multi-
111 well format that facilitates tracking and screening of developmental processes.
112 Relevant to blood formation, a variant gastruloid culture system, which successful
113 organises a heart primordium (Rossi *et al.,* 2021), was also able to recapitulate YS
114 blood formation with erythroid and myeloid cellularity (Rossi *et al.,* 2021; Rossi *et al.,*
115 2022).

116 Herein, we describe a new protocol of gastruloid formation, i.e. hemogenic gastruloids
117 (haemGx), that captures putative YS-like and AGM-like waves of definitive blood
118 progenitor specification from HE. The earlier wave of putative YS-like hematopoiesis
119 includes EMP-like cells which further mature in the gastruloid culture. HaemGx AGM-
120 like hematopoiesis captures the emergence of intra-aortic cluster-like cells, including
121 cells capable of short-term multi-lineage hematopoietic engraftment of the spleen and
122 BM of immunodeficient mice. We harnessed the embryonic hematopoietic context
123 provided by haemGx to attest its utility in modelling leukemias with developmental
124 origins, specifically by recapitulating MNX1-*r* infant leukemia by proxy of *MNX1*
125 overexpression. MNX1-overexpressing haemGx showed expansion of cells at the HE-
126 to-EMP transition, which became susceptible to transformation *in vitro* and
127 recapitulated MNX1-*r* AML patient signatures, placing *MNX1*'s leukemogenic effects
128 at a very early stage of developmental blood formation.

129

130 **RESULTS**

131 **HaemGxs capture time-dependent emergence of endothelium and**
132 **hematopoietic cells**

133 With the aim of recapitulating embryonic hematopoiesis with spatial and temporal
134 accuracy, we modified existing gastruloid formation protocols to (1) promote the
135 specification of hemogenic mesoderm derivatives, and (2) extend developmental time
136 beyond the E8.0-equivalent of early developmental gastruloids (Van den Brink,
137 Susanne C *et al.,* 2014) and the E9.0 timepoint of cardiogenic gastruloids (Rossi *et*
138 *al.,* 2022). We established a 216h haemGx protocol (**Fig. 1A**) using a *Flk1-GFP* mouse
139 ES cell line (Jakobsson *et al.,* 2010), which allows the tracking of early hemato-

140 endothelial specification by expression of the *Flk1* (i.e. *Kdr*) *locus*. The haemGx
141 protocol includes an indispensable pulse of Activin A to induce hemato-endothelial
142 programmes via TGF-β signalling (visible by proxy of GFP-tagged *Flk1* activation), in
143 addition of the characteristic gastruloid patterning via WNT activation by CHI99021
144 supplementation at 48h (**Fig. 1S1A-B**). We promoted hemato-endothelial programs
145 through addition of VEGF and FGF2 at 72h (Sroczynska *et al.,* 2009) (**Fig. 1A**). We
146 observed extension of a polarised *Flk1-GFP*-expressing branched endothelial network
147 (**Fig. 1B**). Flow cytometry analysis revealed the onset of VE-cadherin$^+$ C-Kit$^+$ cells
148 suggestive of EHT, with a peak at 120-144h (**Fig. 1C-D**). This was followed by a surge
149 in CD41$^+$ candidate hematopoietic cells at 144h (**Fig. 1C-D**). In our experience,
150 observation of early polarised activation of the *Flk1 locus* at 96h (**Fig. 1B**) was
151 predictive of CD41$^+$ detection 2 days later, suggesting that TGF-β activation in early
152 patterning is required for later hematopoietic development, putatively by modifying
153 mesoderm specification. We incorporated a 24-hour pulse of sonic hedgehog (Shh)
154 between 144-168h to mimic the aortic patterning that precedes cluster formation (**Fig.
155 1A**) (Rybtsov *et al.,* 2014). Continued exposure to VEGF after the Shh pulse resulted
156 in specification of CD45$^+$ hematopoietic cells progenitors (**Fig. 1C-E**), which were
157 apparent after 168h. The proportion of CD45$^+$ cells was enhanced by the addition of
158 SCF, FLT3L and TPO (**Fig. 1S1C-D**), a combination of cytokines routinely used for
159 HSC maintenance; conversely, it did not require continued addition of FGF2 after 168h
160 (**Fig. 1S1E**), which was omitted. CD45$^+$ cells, some of which co-expressed C-Kit, were
161 present in a small number of 'cluster'-like structures adjacent to Flk1-GFP$^+$ Kit$^+$
162 endothelium (**Fig. 1E**). The process is overall reminiscent of hematopoietic progenitor
163 cell budding from the aortic HE, suggesting recapitulation of hemogenic tissue
164 architecture. Time-dependent specification of C-Kit$^+$, CD41$^+$ and CD45$^+$ cells in
165 haemGx could be obtained with other mouse ES cell lines, including commonly used
166 E14Tg2a (E14) cells (**Fig. 1S1F-H**), with successful specification of comparable
167 frequencies of CD45$^+$ progenitors (**Fig. 1S1I**), attesting to the reproducibility of the
168 protocol across cell lines of different genetic backgrounds.

169 Overall, our extended haemGx model could capture time-dependent specification of
170 CD41$^+$ and CD45$^+$ hematopoietic cells, with recapitulation of the cluster-like spatial
171 arrangement of CD45$^+$ cell emergence from HE, thus showing initial promise as a tool
172 for the exploration of normal and perturbed developmental hematopoiesis.

173
174
175
176
177 **HaemGx specify cells with phenotypic and functional hematopoietic progenitor**
178 **characteristics**

179 We confirmed the induction of endothelial patterning from the Flk1-GFP population by
180 immunofluorescence staining showing the branched vascular-like network of Flk1 co-
181 expressing CD31 (**Fig. 2A**). Endothelial identity of Flk1-GFP$^+$ populations was also
182 detectable by flow cytometry through the presence of endothelial markers CD31 and
183 CD34 (**Fig. 2S1A-B**). Given the co-expression of CD31 and CD34 with hematopoietic
184 progenitors emerging from HE, we inspected CD41$^+$ and CD45$^+$ cells for the presence
185 of these markers. Indeed, we found that subsets of CD41$^+$ and CD45$^+$ hematopoietic
186 cells co-express CD31 and CD34 (**Fig. 2B, Fig. 2S1C-D**), with increased relative
187 representation of hemato-endothelial markers within CD41$^+$ cells at 144h and CD45$^+$
188 at 216h (**Fig. 2C**), consistent with two successive waves of hemogenic emergence
189 from HE. CD45$^+$ cells from 216h haemGx also co-expressed C-Kit, Flk1-GFP and VE-
190 cadherin, a phenotype enriched in emergent hematopoietic stem and progenitor cells
191 (HSPC) in the AGM (**Fig. 2S1E-F**). The phenotype is detected in 1.8% of CD45$^+$ cells
192 on average (**Fig. 2S1G**), corresponding to an estimated 8 cells per haemGx. In
193 agreement, analysis of the haemGx progenitor content at 216h using multipotential
194 CFC assays (**Fig. 2D-E**) revealed the presence of >50 CFC/haemGx, of which an
195 average of 7 have mixed-lineage granulocytic-monocytic-erythroid (GEM) potential,
196 numerically matching the estimated number of HSPC. Other progenitors were
197 predominantly granulo-monocytic (GM) and erythroid (E) progenitors (**Fig. 2D-E**).
198 Analysis of the cellularity of dissociated haemGx at the 216h timepoint using Giemsa-
199 Wright stained cytospins revealed differentiated cells of G, M, E and megakaryocytic
200 lineages (**Fig. 2S2A**), putatively reflecting the progeny of EMP-like cells.

201 We observed temporal progression of expression of key markers by real-time
202 quantitative PCR (qPCR), configuring a sustained hemato-endothelial program from
203 120h through 216h (*Dll4* and *Mecom*), and the generation of hematopoietic
204 populations based on the time-dependent patterns of *Myb, Runx1*, *Hlf* and *Hoxa9* (**Fig.**
205 **2F**). We note that *Myb* peaks at 144h, coincident with the emergence of CD41$^+$ cells,

206 which we postulate to have YS-like EMP characteristics. On the other hand, *Runx1*
207 has a bimodal pattern peaking at 144h, and later at 216h, suggesting the presence of
208 two distinct hematopoietic waves. The later timepoint of 216h has significantly higher
209 expression of both *Hoxa9* and *Hlf* raising the possibility that the second wave indeed
210 corresponds to the specification of AGM-like HSPC. In order to further test the capture
211 of YS-like and AGM-like hematopoietic waves, given the lack of markers that
212 specifically associate with AGM hematopoiesis, we took advantage of the distinct
213 regulatory requirements of polycomb-associated EZH2, which is necessary for EMP
214 differentiation in the YS, but not for AGM-derived hematopoiesis or primitive erythroid
215 progenitors (Neo *et al.*, 2018; Neo *et al.*, 2021). Upon EZH2 inhibition via treatment
216 with GSK216 from 120h onwards, we observed a reduction of CD41$^+$ cells selectively
217 at 144h, with no change in CD45$^+$ cells at the later timepoint of 216h (**Fig. 2S2B-C**).
218 This suggests that the cells present at the later timepoint may not be the progeny of
219 the former, thus supporting recapitulation of 2 independent hematopoietic waves, the
220 latter putatively AGM. Additionally, the vulnerability of 144h CD41$^+$ to EZH2 inhibition
221 suggests that they include YS-like EMP and do not predominantly reflect EZH2-
222 independent primitive YS hematopoiesis.

223

224

225

226 **HaemGx progenitors capture time-dependent signatures of YS-like and AGM-**
227 **like hematopoiesis**

228 To better characterize the extent and progression of developmental hematopoiesis in
229 haemGx, we performed single-cell RNA-sequencing (scRNA-seq) time-course
230 analysis of haemGx cell specification. We sorted cells from 2 independent haemGx
231 cultures at 120, 144, 168, 192 and 216h, and profiled a total of 846 cells using the
232 Smart-Seq2 protocol (Picelli *et al.,* 2014) (**Fig. 3S1A**). We analyzed unfractionated live
233 single haemGx cells at 120-192h and enriched the dataset with sorted CD41$^+$ cells at
234 their peak time of 144h, and sorted CD45$^+$ cells at 192 and 216h. At timepoints of
235 significant enrichment in CD41$^+$ (144h) and CD45$^+$ cells (192h) (**Fig. 1D**), we also
236 sorted C-Kit$^+$ cells. Library preparation and sequencing generated an average of
237 120000 reads/cell, which were mapped to an average of 4000 genes/cell, with minimal
238 signs of cell stress / death as measured by mitochondrial DNA fraction (**Fig. 3S1B**).

239 Read and gene counts were similar between biological replicates, cell types and at
240 different timepoints (**Fig. 3S1C**). The exception was 120h unfractionated haemGx
241 cells, which despite similar sequencing depth (average read counts), were mapped to
242 twice the number of genes (**Fig. 3S1C**), possibly reflecting multi-gene program priming
243 at the onset of hemogenic specification. We selected highly varying genes (HVG)
244 before principal component analysis (PCA) dimensionality reduction and retained the
245 most relevant dimensions. In the PCA reduced space, we constructed a KNN graph
246 and used uniform manifold approximation and projection (UMAP) to visualize the data
247 on 2 dimensions, looking for cell communities using Leiden clustering (**Fig. 3S1D**). We
248 identified 12 cell clusters of which 2 (clusters 0 and 5) almost exactly mapped to C-
249 Kit$^+$ cells (also positive for Sca-1), 2 (clusters 1 and 8) contained CD45$^+$ sorted cells,
250 and cluster 4 uniquely captured the CD41$^+$ cells observed at 144h (**Fig. 3S1D**).
251 Although some unfractionated cells overlapped with the hemogenic cell clusters (**Fig.**
252 **3S1D**, middle panel), most occupied different transcriptional spaces, reflecting the
253 relative frequency of hematopoietic cells and suggesting the presence of potential
254 hemogenic niches.

255 To explore tissue and lineage affiliation of the different clusters, we performed
256 differential gene expression against all other cells using Wilcoxon ranking test, and
257 established cluster classifier gene lists (**Supplemental File S1**). Top classifier genes
258 for clusters 0 and 5 had endothelial affiliation (**Supplemental File S1-S2**), and the
259 cellular composition of the clusters broadly reflected their time signature (**Fig. 3S1D**):
260 cluster 5 comprised 144h cells and cluster 0 included most C-Kit$^+$ cells at 192h. Cluster
261 5 occupied the vicinity of 144h CD41$^+$ cells (**Fig. 3S1D**), which carried an erythroid-
262 biased signature (**Supplemental File S2**). On the other hand, 192/216h-CD45$^+$ cells
263 in cluster 8 expressed myeloid and lymphoid genes (**Supplemental File S1-S2**). The
264 remaining clusters largely corresponded to time-restricted unfractionated cells (**Fig.**
265 **3S1D**, bottom panel), with clusters of cells present at later timepoints matching
266 mesenchymal stromal cell and autonomic neuron signatures (**Fig. 3S1E**), potentially
267 configuring incipient formation of niches relevant to production of HSC (Kapeni *et al.,*
268 2022; Fitch *et al.,* 2012).

269 To specifically explore the extent to which haemGx capture putative YS and AGM-like
270 hematopoiesis, we re-clustered cells sorted on hemato-endothelial markers CD41, C-
271 Kit, and CD45 at the 144h, 192h and 216h timepoints (**Fig. 3A**). We mapped the

272 expression pattern of endothelial (*Pecam* and *Cd34*), HE (*Mecom* and *Cd44*) and
273 hematopoietic (*Runx1*, *Mllt3*, *Hoxa9*, and *Hlf*) markers across the time-dependent
274 clusters (**Fig. 3B**). We observed a coincidence of *Pecam*, *Cd34*, and *Mecom*
275 expression with clusters 0 and 3, which capture C-Kit$^+$ cells at 144 and 192h,
276 respectively. Some positive cells were also seen within CD41$^+$ and CD45$^+$-enriched
277 clusters 2, and 1 and 4, respectively, supporting the presence of phenotypic
278 progenitors, as suggested by flow cytometry analysis (see **Fig. 2**). CD41$^+$ and CD45$^+$-
279 sorted clusters, on the other hand, show expression of *Runx1*, confirming their
280 hematopoietic affiliation. *Cd44*, which is upregulated at the endothelial-to-
281 hematopoietic transition (EHT) and in association with the AGM region (Oatley *et al.*,
282 2020; Fadlullah *et al.*, 2022), is detected predominantly in C-Kit$^+$-enriched cluster 3
283 and CD45$^+$-sorted clusters 1 and 4, corresponding to the later timepoints of 192 and
284 216h. This supports the notion that the later timepoints may capture AGM-like EHT,
285 which is also suggested by significant increases in expression of *VE-cadherin* (*Cdh5*),
286 *Sox17*, as well as *Cd34*, *Cd44* and *Procr* (Fadlullah *et al.*, 2022; Jung *et al.*, 2019;
287 Oatley *et al.*, 2020) in 192h, endothelial-enriched, C-Kit$^+$ cells (**Fig. 3C-D**). Important
288 in this context, these cells express arterial endothelial markers, including *Flt1*, *Efnb2*
289 and *Dll4* steadily at both timepoints (**Fig. 3C and 3S2B**). *Hoxa9* expression also
290 associates with arterial specification and definitive AGM-like hematopoiesis: albeit at
291 low frequency, it is detected exclusively in C-Kit and CD45-sorted cells in the 192 and
292 216h-timepoints of haemGx differentiation (**Fig. 3B, E**). These cells also express *Mllt3*
293 and, rarely, *Hlf*, compatible with incipient specification of an AGM-like definitive blood
294 program (**Fig. 3B, E**). *Mllt3* is additionally expressed in CD41$^+$-sorted cells of cluster
295 2, in line with its role in early erythroid and megakaryocytic lineage identity (Pina *et al.*,
296 2008); *Myb* is also more frequently expressed in this population (**Fig. 3E**). In contrast,
297 *Ikzf1*, which regulates lymphoid specification (Huang *et al.*, 2019), gradually increases
298 its expression, and indeed frequency, in 216h-CD45$^+$ sorted cells. A handover from
299 erythroid/megakaryocytic to myelo-lymphoid lineage affiliation is apparent between
300 144h and 216h haemGx blood cells (**Fig. 3S2C**). Of note, a small number of erythroid-
301 affiliated cells at 216h express adult globins, suggesting a definitive origin (**Fig. 3S2C**).

302 Altogether, scRNA-seq analysis support the suggestion from phenotypic and *in vitro*
303 functional data, that the haemGx model captures two waves of hematopoietic

specification from HE-like cells and may configure pre-definitive and definitive embryonic blood formation.

Late-stage haemGx cells transcriptionally resemble mouse AGM progenitors and have short-term engraftment potential

To investigate the putative YS-like pre-definitve and AGM-like definitive nature of haemGx-generated hematopoietic progenitors, we conducted projections of haemGx sc-RNAseq data onto *bona fide* embryonic populations. We considered 3 distinct scRNA-seq datasets, which (1) capture arterial and haemogenic specification in the para-splanchnopleura (pSP) and AGM region between E8.0 and E11 (Hou *et al.* 2020) (**Fig. 4S1A**); (2) uniquely capture YS, AGM and FL progenitors and the AGM EHT (Zhu *et al.*, 2020) (**Fig. 4S1B**); and (3) interrogate the mouse AGM at the point of HSC emergence from the dorsal aorta HE (Thambyrajah *et al.,* 2024) (**Fig. 4S1C**). We extracted highly-varying genes from each of the 3 datasets to construct dimensionality reduction maps of the data (**Fig. 4S1A-C**), and conducted k-Nearest Neighbor (KNN) analysis to classify and regress our haemGx scRNA-seq data onto those maps considering 3 neighbors (**Fig. 4A**). HaemGx cells with correlation metric >0.7 were projected onto the respective embryonic datasets (**Supplemental File S3)**. Projections onto Hou *et al.* (2020) positioned clusters 0 and 3, corresponding to endothelium, including arterial endothelium and HE-like cells, between days E8.5 and E10, putatively spanning YS-like and AGM-like blood formation. A subset of $CD45^+$-sorted cells, as well as a small numbers of $CD41^+$ cells from cluster 2 projected onto E10 HE and intra-aortic cluster (IAC) cells, suggesting the presence of definitive HSPC or their precursors in haemGx (**Fig. 4A**, **Supplemental File S3**). We sought to further check the transcriptional affiliation of haemGx with intra-embryonic, definitive hematopoiesis, by considering the dataset by Zhu *et al.* (2020), which uniquely captures different hematopoietic locations in the same study (**Fig. 4A**). We matched

338 the identity of a large proportion of CD41$^+$-sorted cells to YS EMP, with some of the
339 same cluster 2 cells mapping to erythroid progenitors, and some later CD45$^+$-sorted
340 cells resembling myeloid progenitors (**Fig. 4A**, **Supplemental File S3**). On the other
341 hand, a subset of CD45$^+$-sorted cells mapped to IAC and lymphoid-myeloid primed
342 progenitors (LMPP) in the FL (**Fig. 4A**, **Supplemental File S3**), compatible with the
343 presence of AGM-like definitive HSPC in haemGx. Interestingly, some CD41$^+$-sorted
344 cells also map to IAC, and indeed to HE, suggesting that this earlier population may
345 contain HSPC precursors, in addition to YS-like progenitors. IAC and HE-like cells
346 were also found within C-Kit$^+$ cells, mostly from the 192h timepoint, while the majority
347 of C-Kit$^+$ cells captured different endothelial identities, as expected from their
348 transcriptional profiles. Projection onto Thambyrajah *et al.* (2024) (**Fig. 4A**,
349 **Supplemental File S3**), conflates YS-like and AGM-like progenitors onto the
350 candidate HSC-enriched embryo AGM cluster 2, highlighting the similarity between
351 transcriptional programs at different sites of embryonic blood emergence. CD45$^+$ cells
352 are more distally-projected onto the same cluster, as shown for E11.5 cells in the
353 Thambyrajah *et al.* (2024) study, some of which correspond to the candidate FL-LMPP
354 in Zhu *et al.* (2020) (**Supplemental File S3**). HaemGx cell similarities to Thambyrajah
355 *et al.* (2024) HSC-enriched cluster 2 also include HE-like and IAC-like cells (as per
356 Zhu *et al.* 2020 projections), mostly E9.5-like as per Hou *et al.* (2020), overall
357 suggesting putative early specification of the AGM region within haemGx.

358

359 **Late-stage haemGx cells can engraft hematopoietic tissues upon maturation**

360 The presence of putative AGM-like hematopoietic cells within haemGx led us to
361 investigate whether haemGx-derived cells had the ability to engraft the hematopoietic
362 system of immunodeficient animals. In early experiments, we had transplanted
363 dissociated haemGx cells obtained at 192 and 216h into irradiated immunocompetent
364 C57Bl/6 recipients, sub-lethally irradiated immunodeficient NSG mice, or non-
365 conditioned NSGW41 hosts, but failed to detect signs of engraftment of peripheral
366 blood, spleen (Spl) or bone marrow (BM) at 12 days (including absence of spleen
367 colony formation), at 4-6 weeks (short-term engraftment), or beyond 8-10 weeks (long-
368 term engraftment) by either flow cytometry detection of CD45.2 or by genomic DNA
369 (gDNA) PCR analysis of the *Flk1-GFP* locus.

370 Failure to engraft was perhaps not surprising given that transcriptional programs
371 position haemGx no later than E9.5-E10, which precedes HSC specification in the
372 AGM. Additionally, specification of HSC-supporting stroma within haemGx, namely
373 mesenchymal cells (Kapeni *et al.*, 2022) and elements of the sympathetic nervous
374 system (Fitch *et al.*, 2012), is only detectable at the 192h timepoint, suggesting that
375 further 3D maturation *in vivo* may be required. We thus tested implantation of
376 undissociated haemGx in the adrenal gland of *Nude* mice (**Fig. 4B**), a topography
377 more commonly used to support tumor development, but that was recently shown to
378 be able to support HSC development, possibly with contributions from the sympathetic
379 adrenal medullary niche (Schyrr *et al.,* 2023). In order to facilitate the analysis of
380 engrafted animals, and given the coincidence of CD45 isoform between *Flk1-GFP*
381 mouse ES cells and the *Nude* mice in which the experiments were performed, we
382 engineered constitutive expression of BFP from the *Rosa26* locus in the *Flk1-GFP*
383 mES cell line (*Rosa26-BFP::Flk1-GFP*) (**Fig. 4S2A**). We implanted 3 haemGx each
384 unilaterally in the adrenal gland of unconditioned recipient mice (**Fig. 4B**). Control mice
385 were injected with an equivalent volume of PBS. Animals were sacrificed at 4 or 8
386 weeks after implantation, with 4-5 experimental animals and 1 control collected at each
387 timepoint. We checked Spl and BM BFP engraftment, by flow cytometry and by PCR
388 analysis of DNA. Sensitivity of PCR analysis was determined by serial dilution of
389 *Rosa26-BFP::Flk1-GFP* mouse ES cells in the human K562 cell line, with detection of
390 as low as 10 pg of mouse input in a total of 100 ng gDNA (**Fig. 4S2B**), corresponding
391 to approximately 3 cells (in 300,000). PCR analysis of BFP engraftment of the Spl and
392 BM 4 weeks after adrenal implantation of 3 haemGx / animal detected *BFP* amplicon
393 in both tissues in 3 of the 5 samples (#11, 12 and 15) (**Fig. 4C and 4S2C**). We sampled
394 the gDNA of each recipient 5 times and detected the presence of the amplicon in 3-5
395 of the replicate samples of the positive animals; replicate sampling of the control was
396 consistently negative (**Fig. 4C and 4S2C**). The same 3 animals showed BFP detection
397 by flow cytometry above control, but no higher than 0.1% (**Fig. 4D**). PCR engraftment
398 was nevertheless detectable in both erythroid Ter119^{+} and lympho-myeloid CD45^{+}
399 fractions of the recipient animals showed PCR engraftment in both lineage fractions
400 (**Fig. 4S2D-E**). Analysis of engraftment 8 weeks after implantation showed low-level
401 detection (<0.1%) of BFP^{+} cells with positivity for CD11b^{+} myeloid and B220^{+} lymphoid
402 compartments in BM of 1 recipient (**Fig. 4D and 4S2F-G**); a 4th recipient had to be
403 sacrificed early due to infection.

404 Overall, implantation in the adrenal niche and/or preservation of the 3D architecture
405 enabled low-level *in vivo* hematopoietic activity, although the mechanism remains
406 unclear. In some recipients, hematopoietic engraftment was accompanied by *in situ*
407 growth and multi-tissue differentiation of the implanted haemGx (**Fig. 4S2H**). However,
408 adrenal enlargement was not observed in all animals with hematopoietic engraftment
409 (e.g. 4 weeks' #11 had a normally sized adrenal, as did all animals analysed at 8
410 weeks), making it unlikely that hematopoietic activity results from teratoma formation
411 by residual pluripotent cells. Late-stage haemGx (216h) may thus contain
412 hematopoietic progenitors with some short-term engraftment potential but incomplete
413 functional maturation, which are present at low frequency. Engraftment is erythro-
414 myeloid at 4 weeks and lympho-myeloid at 8 weeks, reflecting different classes of
415 progenitors, putatively of YS-like and AGM-like affiliation.

416

417

418 **HaemGx resolve the early hemogenic ontogeny targeted by infant AML gene**
419 ***MNX1***

420 The ability of haemGx to reconstitute putative YS and AGM-like HE and hemopoietic
421 progenitors positions it as an attractive model to probe cellular origins of
422 developmental leukemia. In particular, forms of AML unique to the first 2 years of life
423 – infant (inf)AML – have been shown to transform fetal liver (FL) but not adult
424 hematopoietic cells (Waraky *et al.,* 2024; Mercher *et al.,* 2009; Chen, W. *et al.,* 2011;
425 Ragusa *et al.,* 2023), compatible with targeting of a transient embryonic cell. In the
426 case of t(7;12) AML (herein MNX1-*r*) (Ragusa *et al.,* 2023), in which the translocation
427 results in transformation-driving ectopic expression of *MNX1* (Waraky *et al.,* 2024),
428 analysis of patient transcriptional signatures (Ragusa *et al.,* 2022) for over-
429 representation of cell atlas-defined affiliations, indicates enrichment in
430 cardiomyogenic, endothelial, HSC/progenitor and mast cells (**Fig. 5A**). These
431 signature enrichments are compatible with an early hemogenic cell derivative,
432 capturable in the haemGx model. Other cell type-enrichments reflect MNX1 functions
433 in pancreatic and neuronal development (Ragusa *et al.,* 2022) (**Fig. 5S1A**).
434 We overexpressed *MNX1* (MNX1-OE) in *Flk1-GFP* mouse ES cells by lentiviral
435 transduction (**Fig. 5B**) and analyzed progression of haemGx generation (**Fig. 5C**). We
436 used human *MNX1* cDNA to distinguish from the endogenous gene, but the degree of

437 homology is nevertheless high (84%), supporting functional equivalence. We
438 confirmed that *MNX1* overexpression in haemGx is maintained to endpoint (**Fig.**
439 **5S1B**). MNX1-OE haemGx activated polarized *Flk1-GFP* expression and elongated
440 with similar kinetics to empty-vector (EV) control (**Fig. 5C**), albeit with formation of
441 larger haemGx (**Fig. 5D**) denoting increased cellularity (**Fig. 5S1C**). From 192h
442 onwards, MNX1-OE haemGx had a higher frequency of spontaneously contractile
443 structures (**Fig. 5S1D**), compatible with the cardiogenic cell association of patients'
444 transcriptomes (**Fig. 5A**).
445 We interrogated the time-dependent hemogenic cell composition of haemGx by flow
446 cytometry, quantifying C-Kit, CD41 and CD45 (**Fig. 5E-G**, **5S1E-F**), through which we
447 had captured endothelial/HE-like cells, YS-EMP-like, and later putative AGM-like
448 progenitors, respectively. We observed a small but significant expansion of the C-Kit^{+}
449 compartment specifically at 144h (**Fig. 5E**, **5S1E**). All cell markers were relatively
450 unchanged at later timepoints, although absolute cell numbers were higher in MNX1-
451 OE haemGx (**Fig. 5S1C**).
452 To better understand the consequences of MNX1-OE on hemogenic development, we
453 performed RNA-sequencing (RNA-seq) of haemGx at 144h and 216h, capturing
454 putative YS and AGM-like hemopoietic waves, respectively (**Fig. 5H**). We confirmed
455 expression of the human *MNX1* transgene in the sequenced reads (**Fig. 5S2A**); in
456 contrast, we could not detect endogenous mouse *Mnx1* in any of the samples (**Fig. 5**
457 **S2A**), confirming that it does not normally play a role in hemogenic development.
458 DEGs between MNX1-OE and EV (**Supplemental File S4**) configured distinct MNX1-
459 driven programs at 144h and 216h, with >80% genes timepoint-specific (**Fig. 5H**). Both
460 programs include up-regulation of hematopoietic-associated genes, capturing EHT
461 and progenitor regulators (e.g. *Runx1*, and *Epor* and *Zfpm1*, respectively) at the earlier
462 timepoint, and more differentiated effectors, e.g. *Hba-a1*, *Hbb-bh1* and *Il3ra* at 216h,
463 compatible with selective targeting and/or expansion of YS EMP-like cells and their
464 later differentiated progeny (**Fig. 5H**). An overview of transcriptional regulatory
465 programs at both timepoints denotes a clearer hemogenic enrichment at 144h, with
466 over-representation of GATA2, GATA1 and RUNX1 binding targets (**Fig. 5I**), putatively
467 capturing an expansion of a HE-to-EMP-like transition, also represented in enriched
468 GO categories related to VEGF signalling and epithelial-to-mesenchymal transition
469 (EMT) (**Fig. 5J**). In contrast, no enrichment of hemogenic transcription factor (TF)-
470 ChIP targets or related gene ontologies (GO) categories was observed in MNX1-OE

471 haemGx at 216h (**Fig. 5I-J**), suggesting specific targeting of YS-like hemogenic
472 programs. Instead, there was an enrichment in cardiogenic GO at 216h (**Fig. 5J**),
473 reflecting the increase in contractile foci in MNX1-OE late haemGx.
474 We attempted to further deconvolve the cellular heterogeneity underlying bulk RNA-
475 seq DEGs by interrogating the cluster signatures obtained from the single-cell analysis
476 of haemGx development (**Fig. 5S2B-C**; refer to **Fig. 3S1D**). In support of an
477 association with hemato-endothelial specification, MNX1-OE transcriptomes showed
478 enrichment of haemGx HE cluster 5 and 0 signatures; the YS-EMP-like cluster 4 was
479 also enriched (**Fig. 5S2B**). MNX1-OE DEGs at both timepoints also enriched for the
480 relatively less differentiated cluster 10, which predominantly captures unfractionated
481 haemGx cells at 120h, supporting an early ontogenic affiliation. Significantly, clusters
482 5 and 0 HE signatures were also enriched in *MNX1*-r infAML patient signatures
483 (Ragusa *et al.*, 2022) (**Fig. 5S2B-C**), putatively aligning MNX1-OE effects with HE and
484 the hemogenic transition. In contrast, MNX1-OE DEGs specific to 216h showed
485 enrichment in myeloid progenitor / MLP-enriched cluster 8 which captures 192/216h
486 haemGx cells, but diverges from *MNX1*-r infAML signatures (**Fig. 5S2B-C**), instead
487 approximating the distinct myelo-monocytic and/or mixed-lineage affiliation of *KMT2A*-
488 rearranged AML, which is also common in older pediatric and adult patients.
489
490
491
492 **Serial replating of MNX1-OE haemGx cells identifies a candidate C-Kit$^+$**
493 **leukemia-propagating cell more susceptible to transformation in the YS period**
494 We further investigated the potential of MNX1-OE haemGx to reflect *MNX1*-r AML
495 biology by performing serial replating of CFC assays, a classical *in vitro* assay of
496 leukemia transformation. We compared 144h and 216h haemGx to match functional
497 and transcriptome data, and plated unfractionated haemGx, given the transient nature
498 of C-Kit$^+$ enrichment. Both 144h (**Fig. 6A-B**) and 216h (**Fig. 6C-D**) MNX1-OE haemGx
499 replated for at least 5 platings, significantly different to control (EV) (**Fig. 6B,D**). EV-
500 haemGx replating could not be sustained beyond plate 3 in most experiments started
501 from 216h (**Fig. 6C**); 144h EV-haemGx exhibited some low-level replating until plate
502 5 (**Fig. 6A**), but the colonies were formed by few scattered cells clearly distinct from
503 the well-demarcated colonies obtained from MNX1-OE haemGx. We analyzed early-
504 stage (plate 1) and late-stage (plate 5) colonies by flow cytometry to identify the nature

of the replating cells. Cells selected through replating were C-Kit$^+$ (**Fig. 6E**), matching the cell phenotype transiently enriched at 144h. MNX1-OE C-Kit$^+$ cells co-expressed CD31 (**Fig. 6S1A-B**), underpinning the endothelial signatures of *MNX1-r* leukemia. Early replating of MNX1-OE haemGx also enriched for erythroid-affiliated Ter119$^+$ cells, particularly when initiated from the 144h timepoint (**Fig. 6S1C**), but this was not sustained through replating. Myelo-monocytic-affiliated markers, including CD11b or CD45, were not present (**Fig. 6E**, **6S1C**).

We performed RNA-seq analysis of MNX1-OE haemGx CFC cells obtained through re-plating and integrated the data with the 144h and 216h haemGx timepoints to identify genes enriched through the process of transformation. Hs*MNX1* reads were increased in MNX1-OE CFC samples, supporting selective expansion of transduced cells (**Fig. 7A**). We performed hierarchical clustering of all DEGs in pairwise comparisons (**Fig. 7B**), and used Gene Set Enrichment Analysis (GSEA) of *MNX1*-r patient signatures up-regulated in comparison to other AML in the infant (0-2) age group, to interrogate the different clusters. Cluster k14 had significant enrichment of MNX1-*r* signatures against all other AML subtypes (**Fig. 7C**), with considerable overlap between leading-edge (i.e. significantly-enriched) genes (**Fig. 7S1A**). Importantly, k14 captures the subgroup of DEGs up-regulated upon CFC replating which are also up-regulated in MNX1-OE haemGx at 144h (**Fig. 7B**), matching the time-dependent enrichment of the putative MNX1-OE propagating population. Interrogation of cell-type-specific signatures within the k14 MNX1-*r* leading edge genes against the single-cell PanglaoDB atlas captured similarities to hemogenic and non-hemogenic cell types, including interneuron, and other neural signatures (**Fig. 7D**), which reflect MNX1 biological functions (Harrison *et al.,* 1999; Thaler *et al.,* 1999; Ragusa *et al.,* 2022). The most highly significant signatures capture early developmental hemogenic components, such as ‘endothelial cells’ and ‘erythroid cell precursors’, reflecting the 144h, putative YS affiliation of MNX1-OE haemGx. The ‘mast cells’ affiliation of *MNX1*-r AML signatures within the haemGx k14 cluster is also likely to reflect a YS EMP-like affiliation (Chia *et al.*, 2023), and is accordingly aligned with MNX1-OE haemGx isolated at 144h (**Fig. 7S1B**). Giemsa-Wright-stained cytospins of serially-replated

539 CFC assays initiated with 144h MNX1-OE cells showed a shift from blast-like at plate
540 3 to precursor cell morphology with some differentiated mast cells at plate 5 (**Fig. 7E,**
541 **S1C**). We also observed mast cells in 216h-initiated colonies (**Fig.7E**), but these were
542 differentiated at an earlier plate, suggesting that MNX1-OE-responsive C-Kit^{+}
543 progenitors experience maturation between 144h and 216h.
544 Taken together, the data positioned the initiating cellular targets of MNX1-*r* AML at the
545 YS endothelial-to-EMP transition. More broadly, the analysis supports the potential of
546 haemGx to dissect the ontogeny of specific subtypes of hematological cancers
547 initiated *in utero*.
548
549
550
551
552

553 **DISCUSSION**

554 In this study, we developed and characterized a hemogenic gastruloid (haemGx)
555 model from mouse ES cells, which recapitulates YS-like and AGM-like waves of
556 definitive blood progenitor specification, and has incipient ability to engraft adult
557 hematopoietic tissues short-term upon *in vivo* maturation. By exploring the haemGx
558 model through enforced expression of the *MNX1* oncogene, which is a hallmark of
559 infant AML harbouring the t(7;12) translocation, we identified the YS endothelial-to-
560 hematopoietic transition as the transformation target of *MNX1* gene rearrangement,
561 shedding light on to its exclusive association with the infant age group.

562

563 Our data positions the haemGx model as an advanced platform to probe physiological
564 and pathological development of the hematopoietic system, that benefits from the
565 time-dependent and topographic organization associated with gastruloid models. We
566 constructed the haemGx model by introducing a pulse of Activin A and extending the
567 protocol under hematopoiesis-promoting cytokines, to reach the equivalent of mouse
568 embryonic day 10 (E10) and achieve sequential acquisition of YS-like and AGM-like
569 blood production from HE, benchmarked by surface markers CD41 and CD45,
570 respectively. In agreement with recent observations from Dias *et al.* (2025), which
571 highlight a critical balance between WNT-driven posterior and Nodal / TGFβ anterior
572 signals in patterning the midbody axis and endo-mesodermal derivatives, we observed
573 that the combination of Activin A and CHIR99021 was critical for robust *Flk1*/*Kdr*
574 activation and the efficient initiation of CD41 and CD45-marked hematopoietic
575 specification. This is in contrast with the hematopoietic output of cardiogenic
576 gastruloids reported by Rossi *et al.* (2022), which did not require Activin A. Although
577 our strict dependence on Activin A to achieve initial activation of *Flk1* programs and
578 specification of a vascular network and hematopoietic programs precluded a side-by-
579 side comparison with the Rossi protocol (Rossi *et al.,* 2022), we note their apparent
580 lower efficiency in blood formation, at least at the comparable timepoints of 144 and
581 168h, and their bias towards erythroid formation, reflecting a putative YS, and
582 potentially primitive, nature. This bias may also be captured in the reported spleen
583 colony formation; however, given the limited characterisation of the colonies in the
584 manuscript, which does not appraise origin, morphology or cellularity, the nature of the
585 spleen engraftment remains speculative. Importantly, standard gastruloid protocols

586 with only CHIR99021 pulse and no additional cytokines (Beccari *et al.,* 2018) can
587 capture a transcriptional erythrogenic wave.
588
589 Despite low frequency engraftment of adult hematopoietic tissues by haemGx cells,
590 we demonstrated donor origin of the engrafted cells by flow cytometry and PCR
591 detection of a BFP transgene up to 8 weeks. Engraftment progressed from erythroid-
592 biased at 4 weeks to myelo-lymphoid at 8 weeks, as observed in progenitor
593 engraftment from adult cells (Chun *et al.*, 2024). Critically, successful engraftment
594 required implantation of undissociated haemGx for further maturation in a supportive
595 microenvironment, in this case provided by the adrenal gland (Schyrr *et al.,* 2023). In
596 this study, we did not dissect the relative contribution of the continued maturation in
597 3D *vs* the influence of the niche for implantation, as we opted for performing the
598 transplantation as to maximise both. Future experiments will be necessary to
599 understand possible extrinsic niche contributions, including the sympathetic nervous
600 system we sought to include (Kapeni *et al.,* 2022; Fitch *et al.,* 2012), and how they can
601 be incorporated in the haemGx culture by extension and/or further modification of the
602 protocol. Although we observed continued growth and maturation of the haemGx *in*
603 *situ* after implantation, this was not the case in all animals with engraftment. On the
604 one hand, this dissociation excludes the likelihood of the engraftment resulting from
605 teratoma formation from the gastruloid itself. On the other hand, it raises the possibility
606 that the inefficiency of the engraftment is the consequence of a low frequency of cells
607 within each haemGx with HSPC properties, which we can seek to resolve by upscaling
608 of the haemGx culture or further expansion of dissociated HSPC (Ng *et al.*, 2024).
609
610 Although the system is shy of achieving HSC maturation *in situ*, it can mature, to
611 produce low-level engrafting multilineage progenitors, and potentially HSCs. This
612 suggests that access to the relevant hematopoietic molecular programs is preserved.
613 Incipient specification of non-hematopoietic tissues shown to support HSC
614 emergence, including PDGFRA^{+} mesenchymal (Chandrakanthan *et al.,* 2022) and
615 neuro-mesenchymal (Miladinovic *et al.,* 2024) cells, and neural-crest derivatives
616 (Kapeni *et al.,* 2022; Fitch *et al.,* 2012), occurs in sync with development of AGM-like
617 hematopoiesis, to indicate the presence of cues that with further maturation may
618 facilitate HSPC specification at the appropriate developmental time. Interestingly,
619 through extending the haemGx culture, and in addition to AGM-like hematopoiesis, we

also captured the potential to form renal tubular structures upon *in vivo* maturation, suggesting induction of lateral plate and intermediate mesoderm from which the AGM region derives, and putative reconstitution of the intra-embryonic hematopoietic niche.

Establishment of a tractable *in vitro* model of hematopoietic development which faithfully recapitulates time-dependent emergence of hematopoietic progenitors in their physiological niche would propel translational applications by enabling high-throughput studies of HSC emergence. Most ESC-based models of blood specification do not achieve concomitant recapitulation of the surrounding microenvironment. Protocols typically remove ESC-derived hemogenic cells from an early embryonic-like niche to direct differentiation through genetic manipulation (Sugimura *et al.,* 2017) and/or empirically-determined growth factor addition (Lis *et al.,* 2017; Vo and Daley, 2015; Fowler *et al.,* 2024; Nafria *et al.,* 2020; Sroczynska *et al.,* 2009; Piau *et al.,* 2023), with minimal probing of intermediate steps. When hematopoietic waves have been specifically inspected, emergence of YS and AGM-like stages were conflated in time (Pearson *et al.,* 2015; Murry and Keller, 2008), suggesting a bias towards intrinsic hematopoietic cell potential, at the expense of cell fate modulation by time-coherent extrinsic physiological cues, affecting the nature of cellular output. The present haemGx system, in contrast, successfully deconvolutes YS-like and AGM-like stages of blood production, including capturing differences in the HE generation, suggesting a level of organisation amenable to further development into an integrated hemogenic niche. Indeed, projections to scRNA-seq datasets of *in vivo* murine hematopoiesis place our endothelial and hematopoietic outputs at coherent temporal windows.

The haemGx protocol could be further developed to include more supportive structures of tissue organisation, as non-hematopoietic gastruloid models have benefitted from mechanical, or chemo-mechanical, support of artificial ECM embedding (Veenvliet *et al.,* 2020; van den Brink, Susanne C *et al.,* 2020; Dijkhuis *et al.,* 2023), to match temporal coherence and tissue polarisation, with detailed tissue architecture illustrated by somite formation. Prior to HSC emergence in the AGM, critical signals such as SCF and Shh, which we have delivered to the cultures, are secreted by different cells on opposite dorsal-ventral locations (Souilhol *et al.,* 2016), indicating the potential relevance of tissue organisation to improve and potentially extend hematopoietic formation. Moreover, HSPC formation in the AGM is facilitated

654 by circulatory shear stress (North *et al.,* 2009; Diaz *et al.,* 2015; Lundin *et al.,* 2020),
655 suggesting that introduction of flow may improve hemato-endothelial organization and
656 boost blood production.
657
658 We attested the utility of haemGx's to interrogate the biology of leukemias with
659 developmental origins, thanks to a nearly full recapitulation of embryonic hemogenic
660 specification, including critical early stages of HE production and EHT to EMP and
661 MLP/MPPs. We modelled *MNX1*-r AML, a poorly characterized form of leukemia
662 exclusive to the infant age group (Ragusa *et al.,* 2023). Overexpression of *MNX1* in
663 the haemGx model propagated cells bearing transcriptional resemblance to patient
664 *MNX1*-r AML, overall positioning the haemGx as an attractive tractable model in which
665 to study developmental leukemias. We showed that MNX1-OE transiently expanded
666 a C-Kit$^+$ cell with hemato-endothelial characteristics at the YS-like stage, and could
667 selectively sustain the phenotype through serial replating of CFC assays. Replating
668 CFCs could be initiated beyond the YS-like stage, albeit at the relative expense of an
669 early differentiation state and of the similarity to patient transcriptome, putatively
670 defining a window of susceptibility to MNX1-driven transformation.
671
672 Importantly, the haemGx model was able to resolve the cellularity and transcriptional
673 features inferred from patient transcriptomes, with the advantage of a higher resolution
674 in dissecting the contribution of cell-of-origin in the appropriate developmental context.
675 A recent paper using MNX1-OE has shown that MNX1 can initiate a transplantable
676 leukemia in immunodeficient mice by targeting FL, but not BM, cells (Waraky *et al.,*
677 2024). This suggests that ectopic expression of *MNX1* transforms an embryonic cell,
678 rather than depends on an embryonic niche. It also suggests that MNX1-OE can cause
679 leukemia as a single genetic hit, although it cannot be excluded that other mutations
680 are required for, and may indeed have contributed to, full transformation *in vivo*, as
681 observed in a recent report confirming the *in utero* origin of the leukemia (Bousquets-
682 Muñoz *et al.,* 2024). However, the mouse model does not address the nature of the
683 embryonic cell targeted by MNX1-OE. Like the MNX1-OE haemGx cells selected
684 through serial replating, MNX1-OE mouse leukemia are also C-Kit$^+$ (Waraky *et al.,*
685 2024). Interestingly, clusters of genes up-regulated in the mouse MNX1-OE leukemia
686 (**Fig. 7S1D** – k1, 2, 5, 6, and 9) do not robustly separate *MNX1*-r from other infAML by
687 GSEA analysis (**Fig. 7S1E**), and have diverged more from the patient data in their cell-

type signature analysis than the original MNX1-OE cells in the FL (**Fig. 7S1F**), putatively selecting a more generic program of leukemia transformation. With the resolution provided by the haemGx model, we propose that the divergence from *MNX1*-r infAML signatures increases from MNX1-OE haemGx 144h , to 216h to FL, positioning the origin of the AML at the YS stage, likely at the HE-to-EMP transition. This cell may persist in the haemGx culture system to the 216h timepoint, and migrate to the FL in the embryo, eventually extinguishing through differentiation, as suggested by the MNX1-OE haemGx CFC assays. Persistence of the original MNX1-OE target cell will determine the sensitivity to MNX1-OE, and can be addressed in the haemGx system through lineage tracing. Whether additional mutations, namely a reported association between *MNX1*-r and tri(19) (Espersen *et al.,* 2018; Ragusa *et al.,* 2023), contribute to prolonging the undifferentiated state of the candidate YS target cell with increased probability of transformation, or otherwise expand the targeted cell in a more classical second-hit fashion, can potentially also be screened in the haemGx system, in combination with transplantation experiments.

Over the last decade, the effort to achieve robust hematopoietic production from PSC, mouse and human, has been centred on the generation of HSC, a critical translational need (Chang *et al.,* 2023; Dijkhuis *et al.,* 2023; Fidanza and Forrester, 2021). There are good indications that this is achievable, both with (Vo and Daley, 2015; Dang *et al.,* 2002) and without (Piau *et al.,* 2023) genetic manipulation. However, success remains episodic, mainly because mechanistic understanding of processes underpinning blood generation during development is limiting, exacerbated by limiting access to human embryonic and fetal material. On-going optimization of gastruloid models and the success at integration of organoid cultures in microfluidic systems (Saorin, Caligiuri and Rizzolio, 2023; Quintard *et al.,* 2024) will allow us to further optimize haemGx culture. In conclusion, the haemGx represents a faithful *in vitro* reference to replace and complement embryo studies, holding promise to understand the biology of healthy and malignant developmental hematopoiesis and advance blood production for clinical use.

722 **EXPERIMENTAL PROCEDURES**

723 **MATERIALS**

	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
Anti-mouse Ter119 - PC7	Biolegend	Cat. #116223
Anti-mouse CD31 - biotin	BDPharmingen	Cat. #553371
Anti-mouse CD34 - APC	Biolegend	Cat. #128611
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
DMEM/F-12	Gibco	Cat. #11320033
Neurobasal medium	Gibco	Cat. #21103049
N-2 Supplement	Gibco	Cat. #17502048
B-27 Supplement	Gibco	Cat. #17504044
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 144h and 216 h	This paper	ArrayExpress: E-MTAB-12173
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

Mus musculus: Flk-1-GFP::Rosa26-BFP mouse embryonic stem cell line	This paper	n/a
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 *et al.*, 2018	n/a
BFP (forward 5-GCACCGTGGACAACCATCACTT-3; reverse 5-CAGTTTGCTAGGGAGGTCGC-3)	This paper	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
pcdna3-rosa26-rosa26-bfp	This paper	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 *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
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

725

726 **METHODS**

727 **Cell culture**

728 Flk-1-GFP (Jakobsson *et al.,* 2010), Flk-1-GFP::Rosa26-BFP, Sox17-GFP (Niakan *et*
729 *al.,* 2010), T/Bra::GFP (Fehling *et al.,* 2003), and E14Tg2A (Hooper *et al.,* 1987)
730 mouse embryonic stem cells (mES) lines were cultured in ES+LIF medium in
731 gelatinized flasks (0.1% gelatin) with daily medium change, as previously described
732 (Turner *et al.,* 2017). The ES+LIF medium contained 500 ml Glasgow MEM BHK-21
733 (Gibco), 50 ml of fetal bovine serum (FBS, Embryonic Stem Cells tested, biosera,
734 Nuaillé, France), 5 ml GlutaMAX supplement (Gibco), 5 ml MEM Non-Essential Amino
735 Acids (Gibco), 5 ml Sodium pyruvate solution (100 mM, Gibco), and 1 ml 2-
736 Mercaptoethanol (50mM, Gibco). Murine LIF (Peprotech) was added at 1000 U/ml.
737 HEK293T cells were grown in Dulbecco's modified Eagle medium (Gibco)
738 supplemented with 10% FBS (Gibco) and 1% penicillin/streptomycin (Gibco). All
739 cultures were kept at 37°C and 5% CO_2.

740

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

742 mES were maintained in ES+LIF medium and transferred to 2i+LIF (containing Chiron
743 and MEK inhibitor PD03) for 24 hours prior to the assembly into gastruloids. 250
744 untransduced mES cells (or 400 for GX-MNX1 and GX-EV) were seeded in each well
745 of a U-bottom, cell-repellent 96-well plate (Greiner Bio-One, Stonehouse, UK) in 40 µl
746 of N2B27 medium [Takara Bio or home-made as Cold Spring Harbour Protocols
747 'N2B27 Medium' (2017)]. The plate was centrifuged at 750 rpm for 2 minutes to
748 promote deposition and aggregation of the cells and was then incubated at 37°C, 5%
749 CO2 for 48 hours. After 48 hours, 150 µl of N2B27 medium supplemented with 100
750 ng/ml Activin A (QKine, Cambridge, UK) and 3 µM chiron (Peprotech) was added to
751 each well. At 72 hours, 150 µl of medium were removed, without disrupting the
752 gastruloids in the wells. 100 µl of N2B27 with 5 ng/ml of Vegf and Fgf2 each
753 (Peprotech) were added to each well. From 72 h to 144 h, each day 100 µl of medium
754 were removed and replaced with N2B27 + Vegf + Fgf2. At 144 h, the medium was
755 further supplemented with Shh at 20 ng/ml. From 168 h to 216 h, the medium was

N2B27 + 5 ng/ml Vegf, plus 20 ng/ml mTpo, 100 ng/ml mFlt3l, and 100 ng/ml mScf (Peprotech). To dissociate cells from the gastruloid structures, medium was removed and individual gastruloids were collected using a pipette and precipitated at the bottom of a microcentrifuge tube. The remaining medium was aspirated and the bulk of gastruloids was washed in PBS. 50 µl of TrypLE express was added to pelleted gastruloids to be incubated at 37°C for 2 minutes to dissociate cells.

Animals and implantation

Female athymic nude-Foxn1^{nu} (*nu/nu*) mice (Envigo, Bresso, Italy) were housed under pathogen-free conditions. In accordance with the “3Rs policy”, experiments were reviewed and approved by the Animal Welfare Body (OPBA) of IRCCS Ospedale Policlinico San Martino and by the Italian Ministry of Health (n. 883/2020-PR).

Intact haemGx (3 per mouse in 10 µL of PBS) were inoculated in the adrenal gland of five-week-old mice. Briefly, mice were anesthetized with a mixture of xylazine (ROMPUN, Bayer) and ketamine (LOBOTOR, Acme S.r.l.) and injected with haemGx, after laparotomy, in the capsule of the left adrenal gland, as previously described (Brignole *et al.,* 2023; Pastorino *et al.,* 2003). All mice survived surgery. Control mice (CTR) received vehicle only. Mice body weight and general physical conditions were recorded daily. One month after haemGx inoculation, mice were sacrificed and spleen and bone marrow (BM) from femurs collected. BM cells and spleens, mechanically reduced to cell suspension, were stored at -80°C until use. Adrenal glands were paraffin-embedded for subsequent IHC studies.

Methylcellulose colony forming assays (CFC)

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

787 described above. Dissociated gastruloid cells from CFC were collected as cell
788 suspensions in PBS and centrifuged onto slides (Shandon Cytospin 2 Cytocentrifuge)
789 at 350rpm for 5-7min, with a PBS-only pre-spin at 1000rpm, 2min. Slides were stained
790 in Giemsa-Wright stain for 30 seconds followed by addition of phosphate buffer pH 6.5
791 for 5 min.

792

793 **Immunofluorescence staining**

794 Immunostaining of whole gastruloids was performed as described before (Baillie-
795 Johnson *et al.*, 2015). Briefly, gastruloids were fixed in 4% paraformaldehyde (PFA)
796 dissolved in PBS for 4 hours at 4°C on orbital shaking and permeabilised in PBSFT
797 (10% FBS and 0.2% Triton X-100), followed by one hour blocking in PBSFT at 4°C on
798 orbital shaking. Antibody dilutions were made in PBSFT at 1:200 for primary and 1:500
799 for secondary antibodies. Antibody incubations were performed overnight at 4°C on
800 orbital shaking, and subjected to optical clearing overnight in ScaleS4 clearing
801 solution. Individual gastruloids were then mounted on glass coverslips by pipetting as
802 droplets in ScaleS4 and DAPI nuclear stain.

803

804 **Histology and H&E staining**

805 Samples were fixed with 10% formalin at room temperature for 48 hours followed by
806 an ethanol series of ethanol and paraffin; ethanol 70% (90 minutes), ethanol 80% (120
807 minutes), ethanol 95% (180 minutes), 3 times ethanol 100% of 120, 120 and 180
808 minutes, 50% ethanol/paraffin (3x120 minutes) and 3 times paraffin (60, 90 and 120
809 minutes) before embedding in histological cassettes. Blocks were sectioned with a
810 HistoCore BIOCUT Microtome (Leica) at 10 mm and slides were stained with
811 Haematoxylin/Eosin.

812

813 **Imaging**

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

817 microscopy was performed on LSM700 on a Zeiss Axiovert 200 M with Zeiss EC Plan-
818 Neofluar 10x/0.30 M27 and Zeiss LD Plan-Neofluar 20x/0.4 M27 objective lens. Tissue
819 sections were imaged with a THUNDER Imager 3D microscope system (Leica).

820

821 **Lentiviral vector packaging and transduction**

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

829

830 **Flow cytometry**

831 Surface cell marker analysis was performed by staining using the antibodies listed in
832 Key Resources Table. Disassembled gastruloids cells were resuspended in PBS, 2%
833 FBS and 0.5 mM EDTA and stained at a dilution of 1:100 for primary antibodies for 20
834 minutes at 4°C. When indicated, streptavidin was added at a dilution of 1:200. Analysis
835 was performed on ACEA Novocyte (Agilent) or AttuneNxT (Thermo) analyzers, using
836 the respective software packages. Cell sorting was performed using a CS&T calibrated
837 BDFACS Aria III system (488nm 40 mW, 633nm 20 mW, 405nm 30 mW, and 561nm
838 50 mW), set with the 100µm nozzle at 20PSI and a 4-way purity mask or, in the case
839 of haemGx adrenal implants, a Beckman Coulter CytoFlex SRT (488nm 50 mW,
840 633nm 100 mW, 405nm 90 mW, and 561nm 30 mW), set with 100µm nozzle at 15PSI,
841 also using 4-way purity. Single-cell deposition in 96-well plates was performed using
842 single-cell sorting mode. Intact cells were gated on FSC-A vs SSC-A plot, followed by
843 doublet exclusion on FSC-A vs FSC-H and SSC-A vs SSC-H, prior to gating on
844 fluorescent parameters for the markers described in the results.

845

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

847 Gastruloids were collected at different timepoints of the protocol, disassembled and
848 FACS deposited into 96-well plates, either as unsorted (global) or sorted by CD45,
849 CD41 and C-Kit/ScaI markers (sorted). RNA from the cells were extracted from single
850 cells using Smart-seq2 technology at 500Kb and depth of 151 bases. Sequencing
851 reads were quality-checked using FastQC (v0.11.9). We trimmed the samples using
852 trimGalore! (0.6.6) with a cutoff of 30, clipping 15 base pairs and retaining reads of
853 more than 100 bases. Alignment was performed on STAR (2.7.8a) and Mus musculus
854 annotations from GENCODE vM26. Aligned BAM files were annotated using
855 featureCounts (v2.0.1) and count matrices were computed in python by directly
856 accessing the BAM files with gtfparse packages and collapsing lanes.

857 For quality control, based on the histogram of counts and multimodality distributions,
858 we set a minimum count threshold of 200000 counts, a minimum threshold of 1000
859 expressed genes, and a maximum threshold of 20% mitochondrial fraction per cell.
860 We performed the same procedure of the second dataset with a more restrictive
861 threshold of 400000 counts and similar expressed genes and mitochondrial
862 thresholds. Cells that did not pass the quality control metrics were omitted from
863 analysis. We normalized the cells to the mean count number per dataset and applied
864 a plus-one-log transformation of the data before proceeding to the downstream
865 analysis.

866 Dimensionality reduction was performed on feature selection of the gene space using
867 the function scanpy.highly_varying_genes with default parameters. Selection of
868 principal components in principal component analysis (PCA) was performed by
869 heuristic elbow method. Nearest neighbor analysis was constructed by KNN graphs
870 using a correlation metric and 10 nearest neighbors. Data was projected for low
871 dimensional visualization using the UMAP algorithm with default parameters as
872 implemented in scanpy.tl.umap. We used the leiden algorithm as implemented in
873 scanpy.tl.leiden to partition the data into clusters. In order to assess the election of the
874 resolution parameter we used Newman-Girvan modularity as a metric of clustering
875 quality. Differential expressed genes were computed comparing each cluster against
876 the rest using the Wilcoxon test with Benjamini-Hochberg correction.

877

878 **Annotation and projection to additional datasets**

879 Raw counts matrices were downloaded and processed following the same pipeline as
880 described above for scRNA-seq of gastruloids to generate UMAP and clustering. We
881 implemented the scmap algorithm to compare our scRNA sequencing of gastruloids
882 with the available datasets. We reduced the dimensionality of the space by selecting
883 highly varying genes from the annotated dataset. Then, we constructed a KNN
884 classifier with correlation metric and computed the nearest neighbors of the target
885 data. If neighbors with correlation metrics below 0.7 default standards, the projected
886 cells were not projected onto any cell from the annotated dataset. To visualize the cells
887 over the UMAP plots of the other datasets, we constructed a KNN regressor with three
888 neighbors and a correlation metric. Confusion matrices were used to visualize the
889 overlap between gastruloid clusters and the embryo dataset assigned clusters using
890 scmap.

891

892 **Bulk RNA sequencing**

893 Total RNA was extracted from disassembled gastruloid cells at 216 hours. Sequencing
894 was performed on NovaSeq PE150 platform, at 20M paired-end reads per sample.
895 Tophat2 with Bowtie2 were used to map paired-end reads to the reference Mus
896 musculus genome build GRCm39. GENCODE Release M30 (Frankish *et al.,* 2019)
897 was used as the reference mouse genome annotation. Aligned reads were filtered by
898 quality using samtools (Li, H. *et al.,* 2009) with a minimum threshold set at 30 (q30).
899 Transcript assembly and quantification was achieved using htseq (Putri *et al.*, 2022).
900 Differential expression between sample and control was performed by collapsing
901 technical replicates for each condition using Deseq2 (Love, Huber and Anders, 2014)
902 in R environment (Deseq2 library v 1.32.0). The differential expression was expressed
903 in the form of log2 fold change and filtered by false discovery rate (FDR) of 0.05.

904

905 **Gene list enrichment analyses**

906 Gene ontology (GO) analysis was performed in ExpressAnalyst (available at
907 www.expressanalyst.ca) using the PANTHER or GO Biological Process (BP)
908 repository. GO terms and pathways were filtered by false discovery rate (FDR) with a
909 cut-off of ≤ 0.1 for meaningful association. EnrichR (Xie *et al.,* 2021; Kuleshov *et al.,*
910 2016; Chen, E. Y. *et al.,* 2013) was used for cell type analysis using the Panglao DB

911 (Franzén, Gan and Björkegren, 2019) databases using a FDR threshold of ≤ 0.1 or p
912 value ≤ 0.01, where specified, as well as transcription factor (TF) binding site
913 enrichment using the ENCODE and ChEA Consensus TFs from ChIP database.

914

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

916 Custom gene signatures were used as gene sets for GSEA analysis (Subramanian *et*
917 *al.,* 2005) on the GSEA software v4.2.3 on RNA sequencing expression values in
918 counts units. GSEA was ran in 10000 permutations on gene set using the weighted
919 Signal2Noise metric. Enrichment metrics are shown as normalized enrichment score
920 (NES) and filtered by FDR ≤ 0.05. Leading edge genes (LEGs) are genes with a "Yes"
921 values for core enrichment. For AML patient analysis, clinical phenotype and
922 expression data (in counts units) were extracted from the GDC TARGET-AML cohorts
923 in the Therapeutically Applicable Research to Generate Effective Treatments project
924 (TARGET, https://ocg.cancer.gov/programs/target), downloaded from the University of
925 California Santa Cruz (UCSC) Xena public repository (last accessed 31st August
926 2022). Patient samples were selected according to the reported karyotype to include
927 t(7;12), inv(16), MLL, normal karyotype, and t(8;21). GSEA was performed comparing
928 RNA sequencing counts of t(7;12) samples against pooled AML subtypes (inv(16),
929 MLL, normal karyotype and t(8;21)) as "other AML".

930

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

932 Extracted RNA was reverse-transcribed into complementary DNA (cDNA) using High-
933 Capacity RNA-to-cDNA Kit (Applied Biosystems). QPCR was performed using
934 FastGene 2x IC Green Universal qPCR Mix (Nippon Genetics, Duren, Germany) using
935 primers for human *MNX1* (forward 5-GTTCAAGCTCAACAAGTACC-3; reverse 5-
936 GGTTCTGGAACCAAATCTTC-3) (Gulino *et al.*, 2021) and *Ppia* for endogenous
937 control (forward 5-TTACCCATCAAACCATTCCTTCTG-3; reverse 5-
938 AACCCAAAGAACTTCAGTGAGAGC-3) (Moris *et al.,* 2018), or Taqman gene
939 expression assays (Mm00501741_m1 *Myb*; Mm00439364_m1 *Hoxa9*;
940 Mm00723157_m1 *Hlf*; Mm01213404_m1 *Runx1*; Mm00444619_m1 *Dll4*;
941 Mm00491303_m1 *Mecom*; Mm02342430_g1 *Ppia*) using Taqman Gene Expression

942 Mastermix (Applied Biosystems). Differential gene expression was calculated using
943 the delta delta Ct (ΔΔCt) method.

944

945 **Polymerase chain reaction (PCR)**

946 Genomic DNA was extracted using Monarch Genomic DNA Purification Kit (New
947 England Biolabs, Hitchin, UK) using manufacturer's instructions. PCR on genomic
948 DNA was performed using Phire PCR Master Mix (Thermo Scientific) to amplify the
949 BFP locus using forward primer 5'-GCACCGTGGACAACCATCACTT-3' and reverse
950 primer 5'-CAGTTTGCTAGGGAGGTCGC-3'.

951

952 **Statistical analysis**

953 Experiments were performed at least in triplicates, unless specified otherwise. Data
954 are plotted to include standard deviation (+/-SD) between replicates. Statistical
955 significance was set at a threshold of p value < 0.05. Statistical analysis was performed
956 in R environment (version 4.1.3) or using GraphPad Prism 8.0 software and is detailed
957 in respective figure legends.

958

959 **Data availability**

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

971

ACKNOWLEDGEMENTS

This project was funded by a start-up grant and a BRIEF award from Brunel University London, a Little Princess Trust (LPT) Project Grant (CCLGA 2023 22 Pina) and a National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) project grant (NC/Z500677/1) to CP, by an Agencia Estatal de Investigación grant (PLEC2021-007518) to AB and AMA, and by an ERC Advanced Grant (MiniEmbryoBlueprint 834580) to AMA. LPT research is funded in partnership with CCLG through the CCLG Charity Research Network. DR received funding from the Royal Society of Biology (MRSB Travel Grant). GTC was funded by grant FPU18/05091 from the Spanish Ministry of Universities. AJ is funded by a Lady Tata Memorial Trust International Scholarship (2022). SvdB was funded by an EMBO Postdoctoral Fellowship (ALTF 195-2021) and is in receipt of a HFSP Postdoctoral Fellowship (LT0047/2022-L). Work in the CP lab was also funded by a KKLF Intermediate Fellowship (KL888), a Leuka John Goldman Fellowship for Future Science (2017-2019), and a Wellcome Trust / ISSF Bridge Funding award at the University of Cambridge (2019). JGO acknowledges financial support from the Spanish Ministry of Science and Innovation and FEDER (grant PGC2018-101251-B-I00), by the Maria de Maeztu Programme for Units of Excellence in R&D (grant CEX2018-000792-M), and by the Generalitat de Catalunya (ICREA Academia programme). MP acknowledges funding from the Italian Ministry of Health (Ricerca Finalizzata 5 per mille and Ricerca Corrente to MP). 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 acknowledge Tina Balayo, Ana Filipa Domingues, Shikha Gupta, 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, FP, SvdB, MC,MP, KRK, VH-H, AB, JGO, AMA, CP; Software: GTC, JGO; Validation: CWS, DR, AJ, YC, SvdB, CP; Investigation: CWS, DR, GTC, FP, AJ, YC, LD, CB, G-AI, CP; Formal Analysis: GTC, DR, JGO, CP; Resources: GTC, JC, MP, 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, AB.

DECLARATION OF INTERESTS

Authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental Figures S1-S7

Supplemental Files S1-S3 (Excel files):

Supplemental File S1: Cluster classifier gene lists for each cluster by differential gene expression analysis of scRNA-seq of haemGx (all conditions), obtained by Wilcoxon rank test of each cluster against all other clusters.

Supplemental File S2: Cell type enrichment analysis of cluster classifier genes, showing the inferred identity of clusters of scRNA-seq of haemGx (all conditions). Classifier genes (in Supplemental File S1) were obtained by differential expression in comparison to all other clusters and subjected to enrichment analysis using the PanglaoDB repository (Franzén, Gan and Björkegren, 2019).

Supplemental File S3: Summary of projections of scRNA-seq of haemGx cells onto available datasets by Hou et al. (2020), Zhu et al. (2020), and Thambyrajah et al. (2024), showing the identity of each haemGx cell projecting to annotated clusters of the respective studies.

Supplemental File S4: List of differentially expressed genes between haemGx-MNX1 and haemGx-MNX1 at 144h and 216h compared to empty vector (EV) control.

1033 **FIGURE LEGENDS**

1034

1035 **Figure 1 – Hemogenic gastruloids (haemGx) produced from mES cells promote hemato-**
1036 **endothelial specification with spatio-temporally accurate ontogeny. (A)** Timeline of mES
1037 cells assembly and culture into gastruloids over a 216h period with the addition of specified
1038 factors for the promotion of hemato-endothelial specification. The 24-hour pre-treatment in
1039 2i+LIF was omitted when mES cultures showed no signs of differentiation. **(B)** Imaging of
1040 haemGx over time from 72h to 216h at 10x magnification, showing the assembly and growth
1041 of the 3D structures and the polarization of the Flk-1-GFP marker from 96h; scale bar: 100mm.
1042 **(C)** Flow cytometry timecourse analysis of haemGx for the presence of CD144 (VE-cadherin),
1043 C-Kit, CD41, and CD45 markers; representative plots. **(D)** Quantification of analysis in (C).
1044 Violin plots with median and interquartile range of up to 26 replicates; Kruskal-Wallis CD144
1045 ($p=0.0755$), C-Kit ($p<0.0001$), CD41 ($p<0.0342$), CD45 ($p<0.0001$). Shown are significant
1046 Dunn's multiple comparisons tests; adjusted q-value. **(E)** Immunostaining of whole individual
1047 haemGx at 216h showing the localized expression of Flk1-GFP (green), C-Kit (yellow), and
1048 CD45 (red) markers; nuclear staining is in DAPI (blue); scale bar: 100μm.

1049

1050 **Figure 1 Supplement 1 – Optimization and reproducibility of the haemGx protocol. (A)**
1051 Imaging of a 120h-haemGx with and without the addition of Activin A in conjunction with Chiron
1052 99021 (CHIR) to achieve polarized expression of *Flk-1-GFP*; scale bar: 100mm. **(B)** Titration
1053 of Activin A dosage (from 25 ng/ml to 150 ng/ml) required for *Flk-1-GFP* expression activation
1054 (green). **(C)** Representative flow cytometry plots of %CD45$^+$ cells at 216h without and without
1055 addition of SCF, Flt3L and TPO (SFT) from 168h; representative plots. **(D)** Quantification of
1056 %CD45$^+$ cells by flow cytometry analysis of haemGx cultured with or without SFT, between
1057 192-216h and 168-216h. Mean ± SD of n=7-8 independent experiments; Tukey's multiple
1058 testing with significant q-value<0.05. **(E)** Quantification of %CD45$^+$ cells at 216h by flow
1059 cytometry analysis of haemGx cultured in VEGF+FGF or VEGF-only during the 168-216h
1060 period. Mean ± SD of n=7 independent experiments; unpaired t-test; ns, not significant. **(F)**
1061 Flow cytometry monitoring of C-Kit, CD41, and CD45 markers in haemGx established from
1062 E14 and *Flk-1-GFP* mES lines; n=3-4 replicate experiments (n=1, CD45$^+$ 120h). Mixed effects
1063 model with Sidak's multiple comparison test; significant adjusted p-value <0.05. **(G)**
1064 Representative flow cytometry plots of CD41$^+$ and C-Kit$^+$ cells in E14 and *Flk-1-GFP* mES cell-
1065 initiated haemGx. **(H)** Representative flow cytometry plots of CD45$^+$ cells in E14 and *Flk-1-*
1066 *GFP* mES cell-initiated haemGx. **(I)** Quantification of %CD45$^+$ cells in *Flk-1-GFP*, *Sox17-GFP*,
1067 and *TBra-GFP* mES cell- initiated haemGx; n=5, mean ± SD. Welch's t-test Sox17-GFP vs.
1068 Flk1-GFP, $p=0.5254$; Tbra-GFP vs. Flk1-GFP, $p=0.3782$.

1069

1070 **Figure 2 – HaemGx produce hematopoietic output following the emergence of**
1071 **specialised endothelium. (A)** Immunostaining of whole individual haemGx at 144h showing
1072 the establishment of a vascular network by co-expression of CD31 and Flk1-GFP; nuclear
1073 staining is in DAPI (blue); scale bar: 100μm. **(B)** Flow cytometry analysis of haemGx cells at
1074 144h and 216h stained for CD41, CD45, CD31, and CD34. **(C)** Quantification of flow cytometry
1075 analysis in (B) of CD41+ and CD45+ fractions co-expressing CD34 and CD31. Mixed effect
1076 analysis with a Ŝidàk test for multiple comparisons. **(D)** Colony-forming cell (CFC) assay of
1077 216h-haemGx in multipotential methylcellulose-based medium. GEM: granulocyte-erythroid-
1078 monocyte; GM: granulocyte-monocyte; M: monocyte; E: erythroid. Please note that the
1079 medium does not include TPO and does not assess the presence of megakaryocytic
1080 progenitors. CFC frequency in 3 haemGx, n=3, mean±SD. **(E)** Representative photographs of
1081 colonies quantified in C; scale bar=2mm. **(F)** Real-time quantitative (qPCR) analysis of
1082 expression of hemato-endothelial genes by timepoint. Relative gene expression fold change
1083 calculated by normalization to Ppia. Bars represent mean of 3 replicates and show individual
1084 data points; p values by ANOVA across datapoints; post-hoc statistical significance between
1085 specific variables by Tukey's test and shown by brackets.

1086

1087 **Figure 2 Supplement 1 – Multi-color flow cytometry detection of surface markers in**
1088 **haemGx. (A)** Representative flow cytometry plots of CD31 and CD34 staining of haemGx at
1089 144h and 216h gated on Flk-1-GFP$^+$. **(B)** Quantification of Flk-1-GFP$^+$ haemGx populations at
1090 144h and 216h expressing CD34 and CD31. Crossbar shows mean of 4 replicate experiments
1091 with individual data points shown; Welch's t-test; p value significant < 0.05. **(C-E)** Unstained
1092 controls (C-D) and fluorescence minus one (FMO) controls (E) relative to flow cytometry
1093 detection of multi-color of surface markers in Figure 2. **(F)** Representative flow cytometry plot
1094 of haemGx of CD45$^+$CD41lo fractions co-expressing C-Kit and Cd144 (Ve-Cadherin).

1095

1096 **Figure 2 Supplement 2 – Characterization of hematopoietic output from haemGx. (A)**
1097 Representative image of cytospins of dissociated haemGx at 216h stained with Giemsa-
1098 Wright's stain. Annotated are cells in the monocytic (dashed open arrow), granulocytic (solid
1099 open arrow), megakaryocytic (solid arrow) and erythroid (asterisk) lineages; arrowheads
1100 indicate cells with a non-specific blast-like morphology. 40x magnification, scale bar = 100 μm.
1101 **(B)** Representative flow cytometry plots of CD41 and CD45 expressing populations in haemGx
1102 treated with 0.5 μM EZH2 inhibitor GSK126 or with 0.05% DMSO (control) at 144h and 216h.
1103 **(C)** Quantification of flow cytometry analysis in (B) showing the proportion of CD41$^+$ and
1104 CD45$^+$CD41^{lo} populations in GSK126-treated haemGx. Welch's t-test; p value significant <
1105 0.05.

1106
1107
1108 **Figure 3 – Time-resolved analysis of haemGx by scRNA-seq captures successive**
1109 **waves of hematopoietic specification. (A)** UMAP clustering of subsets of sequenced cells
1110 expressing CD41, C-Kit, and CD45 at 144h, 192h and 216h, annotated by timepoint (left
1111 panel), sorting condition (middle panel), and Leiden clustering (right panel). **(B)** UMAP panels
1112 highlighting the expression of specific hemato-endothelial markers in the transcriptional
1113 spaces defined in (A). Color scale indicates expression levels in counts. **(C)** Heatmaps of
1114 differentiation of an arterial endothelial programme in haemGx from scRNA-seq data of sorted
1115 on expression of C-Kit at 144h and 192h. Color scale indicates expression in counts. **(D)**
1116 Violin plots quantifying the expression of hemato-endothelial markers in C-Kit+ fraction from
1117 clusters in (A) comparing 144h and 192h timepoints. Wilcoxon test; significant p value<0.05.
1118 **(E)** Expression of definitive hematopoietic genes in in C-Kit$^+$, CD41$^+$ and CD45$^+$ cells at 144h
1119 and 192/216h timepoints; expression levels as normalised counts per million reads. Mann-
1120 Whitney or Kruskal-Wallis with Dunn's multiple comparisons; significant p or q-value<0.05.
1121
1122 **Figure 3 Supplement 1 – Single-cell-RNAseq (scRNAseq) analysis of haemGx identifies**
1123 **time-dependent signatures of endothelial, hemogenic and stromal cells**. **(A)** Summary
1124 of plating strategy for scRNA-seq analysis of gastruloids at 120h, 144h, 168h, 192h, and 216h
1125 without selection of surface markers ('all') or sorted as C-Kit/Sca1+ (green shading), CD41+
1126 (blue), or CD45+ (orange) cells; P = plate. **(B)** ScRNA-seq quality control measures
1127 (quantification of mapped reads, individual genes and mapping of reads on to mitochondrial
1128 (mt)DNA. **(C)** Mapped read and gene counts for individual Smart-seq2 libraries (plate, P),
1129 biological replicate (repl) populations of single cells, and read and map distributions across
1130 unsorted and cell surface phenotype-sorted cells. **(D)** UMAP projection of all sequenced cells
1131 colored by annotated clusters (left), by sorting markers C-Kit/ScaI, CD41, and CD45, or
1132 unfractionated cells ('whole' corresponding to 'all' in panel A) (center), and by haemGx culture
1133 time (right). **(E)** Cell type enrichment analysis of cluster classifier genes, using the PanglaoDB
1134 repository (Franzén, Gan and Björkegren, 2019); classifier genes are obtained by differential
1135 expression in comparison to all other clusters. Represented are clusters 1, 3 and 9, which are
1136 characteristic of timepoints 168-216h, and capture stromal cells putatively relevant for
1137 hemogenic support, including autonomic neurons (cluster 9). The statistical power of
1138 representation of individual cell types is expressed as the EnrichR combined score with a p-
1139 value threshold of < 0.05.
1140
1141 **Figure 3 Supplement 2 – Single-cell-RNAseq (scRNAseq) analysis of haemGx. (A)**
1142 Heatmap representation of the expression of endothelial and hematopoietic marker genes in

individual haemGx CD41$^+$ and CD45$^+$ cells sorted at 192 and 216h. Cells and genes ordered by unsupervised hierarchical clustering. **(B)** Violin plots quantifying the expression of arterial markers in C-Kit+ fraction from clusters in Fig. 3A comparing 144h and 192h timepoints. Wilcoxon test; significant p value<0.05. **(C)** Summary representation of the proportion of hGx cells expressing endothelial (endo), erythroid/myeloid (Ery+My), myeloid/lymphoid (My+Ly) and HSPC gene signatures. Endo: *Kdr*; Ery+My: *Epor*±*Gata1*±*Klf1*±*Hbb*±*Hba*±*Hbg* + *Spi1*±*Mpo*±*Anpep*±*Csf1/2/3r*); My+Ly: *Spi1*±*Csf1/2a/3r* + *Ikzf1*±*Ighm/d*±*Igk*); HSPC: *Ptprc* + *Myb*).

Figure 4 – Late-stage haemGx contain hematopoietic output with transcriptional alignment to mouse embryonic populations and in vivo engraftment potential. (A) Projection of clustered haemGx cells (CD41, C-Kit, and CD45 at 144h, 192h and 216h; Fig. 3A) onto mouse single-cell RNA-seq datasets capturing: arterial and haemogenic specification in the para-splanchnopleura (pSP) and AGM region between E8.0 and E11 (Hou et al. 2020); YS, AGM and FL progenitors and the AGM EHT (Zhu et al., 2020); HSC emergence from the dorsal aorta HE (Thambyrajah et al., 2024). Gastruloid cell projection identified as per their cluster of origin (Fig. 3A). **(B)** Schematic representation of the experimental workflow for execution and analysis of haemGx implantation in the adrenal gland of immunodeficient mice. **(C)** PCR detection of haemGx genomic (g)DNA in the bone marrow (BM) and spleen (Spl) of immunodeficient mice 4 weeks after unilateral adrenal implantation of 3 gastruloids/gland. Analysis of 5 replicates of 100ng gDNA/recipient tissue; control animal was injected unilaterally with PBS in an adrenal gland in parallel with experimental implantation. Reaction positive control used 100ng of gDNA from *Rosa26-BFP*::*Flkj1-GFP* mES cells (mES) used to generate haemGxs. NTC: no template control. **(D)** Flow cytometry plots of engraftment detection by BFP expression in bone marrow (BM) of recipient mice 4 weeks following implantation of 216h haemGx. **(E)** Flow cytometry plots of lineage affiliation of BFP$^+$ engraftment in bone marrow (BM) of recipient mice 8 weeks following implantation of 216h haemGx.

Figure 4 Supplement 1 – Analysis of mouse embryonic scRNA-seq datasets for transcriptional similarities of haemGx outputs. (A-C) UMAPs of mouse single-cell RNA-seq datasets annotated as in respective publications: (A) arterial and haemogenic specification in the para-splanchnopleura (pSP) and AGM region between E8.0 and E11 (Hou et al. 2020); (B) YS, AGM and FL progenitors and the AGM EHT (Zhu et al., 2020); (C) HSC emergence from the dorsal aorta HE (Thambyrajah et al., 2024).

1179 **Figure 4 Supplement 2 – Detection of engraftment potential of late-stage haemGx in**
1180 **adrenal glands of Nude mice. (A)** Map of the *Rosa26-BFP* construct used to target *Flk1-*
1181 *GFP* mES cells employed in implant assays. Primers used for gDNA PCR analysis are
1182 indicated. **(B)** PCR detection limit of the BFP amplicon using the primers in (A), in a serial
1183 dilution of *Rosa26-BFP*::*Flk1-GFP* mES gDNA into human leukemia cell line K562 gDNA; total
1184 100ng DNA per lane. **(C)** Uncropped image of agarose gel electrophoresis of BFP detection
1185 shown in Fig. 4C. **(D)** Representative flow cytometry plot of the spleen of immunodeficient
1186 mouse recipients, showing myelo-lymphoid ($CD45^{+}Ter119^{-}$) and erythroid ($Ter119^{+}CD45^{+/-}$)
1187 cell fractions at 4 weeks post implantation. **(E)** PCR analysis of *Rosa26-BFP*::*Flk1-GFP* gDNA
1188 in the cell fractions in (D) of implanted animal #15 at 4 weeks post-implantation. PCR run in
1189 triplicate material from each sample; positive and NTC controls as per Fig. 4C. **(F-G)**
1190 Representative flow cytometry interrogation of BFP detection within lineage compartments in
1191 the bone marrow (BM) of control and implanted recipient animal #2 (F) and animal #1 (G) at
1192 8 weeks after hGx implantation. **(H)** Histology of implanted and control adrenal glands; paraffin
1193 sections stained with hematoxylin & eosin (H&E). Inserts in implanted animal highlight neural
1194 (orange) and hematopoietic and renal tubular epithelium (green); inserts in control animal
1195 highlight normal cortical (blue) and medullar (red) adrenal architecture, absent in implanted
1196 adrenal glands.
1197
1198 **Figure 5 – HaemGx with *MNX1* overexpression have increased cellularity and enhanced**
1199 **hemogenic endothelial potential. (A)** Cell type enrichment analysis in MNX1-r AML
1200 transcriptomes, compared to normal paediatric bone marrow (BM) or other paediatric AML
1201 from the TARGET database. GSEA used representative cell type gene sets from the 2021 DB
1202 database; bubble plot shows NES scores (color gradient) and statistical significance (–
1203 log10(FDR), bubble size). **(B)** Schematic representation of generation of *MNX1*-
1204 overexpressing and empty vector (EV) mES cells by lentiviral transduction, with subsequent
1205 assembly into haemGxs. **(C)** Imaging of haemGxs with *MNX1* overexpression and EV controls
1206 at 10x magnification, showing appropriate assembly and polarization of the Flk1-GFP marker.
1207 scale bar: 300mm. **(D)** Size of haemGx at each timepoint, determined by the distance between
1208 the furthest extreme points in µm. Mean ± SD of 3 replicate experiments; 2-tailed t-test, $p<$
1209 0.05 (*), 0.001 (**), 0.0001 (***), and 0.00001 (****). **(E-G)** Flow cytometry timecourse analysis
1210 of **(E)** C-Kit^{+}, **(F)** $CD41^{+}$, and **(G)** $CD45^{+}$ cell abundance in MNX1 and EV haemGxs (120-
1211 216h). Mean ± standard deviation of 3-7 independent experiments; 2-way ANOVA and Sidak's
1212 multiple comparison test significant at $p<0.05$ for C-Kit^{+} cells only (construct contribution to
1213 variance $p=0.0191$; 144h comparison *$p=0.0190$). **(H)** Volcano plots of differentially expressed
1214 genes (DEGs) by bulk RNA-seq between MNX1 and EV haemGxs at 144h and 216h.
1215 Intersection of statistically significant DEGs between 144 h and 216 h (shown in Venn diagram

on top right) identified unique DEGs for each condition (labelled in red for 144 h and blue for 216 h). **(J)** Gene Ontology (GO) term enrichment for differentially upregulated genes between MNX1 and EV haemGx at 144h (red) and 216h (blue), computed in EnrichR using the GO Biological Process repository. **(I)** Transcription factor (TF) binding site enrichment on differentially upregulated genes between MNX1 and EV haemGx at 144h (red) and 216 hr (blue) using the ENCODE and ChEA Consensus TFs from ChIP database on EnrichR.

Figure 5 Supplement 1 – MNX1 overexpression promotes hemogenic specification in haemGx. **(A)** Cell type enrichment analysis in MNX1-r AML transcriptomes, compared to normal paediatric bone marrow (BM) or other paediatric AML from the TARGET database. GSEA used representative cell type gene sets from the 2021 DB database; bubble plot shows NES scores (color gradient) and statistical significance (–log10(FDR), bubble size). **(B)** Quantitative (q)RT-PCR analysis of *MNX1* overexpression in 216h-haemGx. Gene expression fold change calculated by normalization to HPRT1. Mean ± SD of 3 replicates; 2-tailed t-test, $p< 0.001$ (**). **(C)** Cell counts of disassembled haemGx at 216h. Mean ± SD of 3 replicate experiments; 2-tailed t-test, $p<0.001$ (**). **(D)** Proportion of EV and MNX1 haemGx exhibiting spontaneous unifocal contractility at 192h, observed in commercial N2B27 medium (see Experimental Procedures); mean ± SD of 3 replicate experiments; 2-tailed t-test, $p<0.001$ (**). **(E-F)** Flow cytometry analysis of specification of **(E)** hemato-endothelial – C-Kit$^+$ and VE-Cadherin$^+$ – and **(F)** hematopoietic progenitor – CD41$^+$ and CD45$^+$ – cells over a 120-216h timecourse of EV and MNX1 haemGx cultures; representative plots.

Figure 5 Supplement 2 – Transcriptional analysis of haemGx with MNX1 overexpression. (A) Expression of human *MNX1* and murine *Mnx1* genes in FPKM units from RNA-seq of MNX1 and EV haemGx. **(B)** UMAP of time-resolved global clustering of scRNA-seq of haemGx showing the enriched clusters in GX-MNX1 at 144 hr (orange boxes) and 216 hr (blue boxes), and MNX1-*r* AML (black boxes) or MLL-AML (green boxes) determined by GSEA. **(C)** 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 at 144 hr and 216 hr compared to respective GX-EV, and MNX1-*r* or MLL AML samples compared to other pediatric AML.

Figure 6 – MNX1 selects C-Kit$^+$ clonogenic cells and selectively transforms end-stage haemGx. (A) Representative photographs of serial replating of colony-forming cell assays initiated from EV or MNX1 cells obtained at 144h of the haemGx protocol. **(B)** Quantification of colony-replating efficiency of EV and MNX1 144h-haemGx (GX) cells. Mean+SD of n>3 replicates; Kruskal-Wallis with Dunn's multiple comparison testing at significant $q<0.05$. **(C)**

1253 Representative photographs of serial replating of colony-forming cell assays initiated from EV
1254 or MNX1 cells obtained at 216h of the haemGx protocol. **(D)** Quantification of colony-replating
1255 efficiency of EV and MNX1 216h-haemGx cells. Mean+SD of n=5-8 replicates; Kruskal-Wallis
1256 with Dunn's multiple comparison testing at significant q<0.05. **(E)** Flow cytometry analysis of
1257 hemogenic progenitor C-Kit^{+} cells and myeloid-affiliated CD11b^{+} cells at first round of plating
1258 of 144h and 216h-haemGx initiated from EV and MNX1 cells; representative plots.
1259
1260 **Figure 6 Supplement 1 – Characterization of MNX1-overexpressing replating cells from**
1261 **haemGx. (A)** Representative flow cytometry plot of CD31 and C-Kit staining of CFC plating of
1262 GX-MNX1 (left) and quantification of sustained expression of CD31^{+}CKit^{+} cells through
1263 replatings. **(B)** Representative flow cytometry plots of serially re-plating MNX1 cells from 144h
1264 and 216h-gastruloids (plate 5 cells) stained with Ter119 and CD45.
1265
1266 **Figure 7 – HaemGx with *MNX1* overexpression recapitulate MNX1-Acute Myeloid**
1267 **Leukemia patient signatures. (A)** Quantification of human *MNX1* transcripts in FPKM units
1268 from RNA-seq of GX-EV, GX-MNX1, and CFC at 144 hr and 216 hr. Bars represent mean of
1269 replicates. **(B)** Heatmap comparing the expression of all differentially expressed genes (DEGs)
1270 between all conditions (GX-MNX1 144 hr vs GX-EV 144 hr; GX-MNX1 216 hr vs GX-EV 216
1271 hr; CFC MNX1 144 hr vs GX MNX1 144 hr; CFC MNX1 216 hr vs GX MNX1 216 hr) as Z
1272 score. Hierarchical clustering by Ward D method on Euclidean distances identifies 14 clusters
1273 (k). **(C)** GSEA plots for gene set extracted from k14 from **(B)** against MNX1-r patients RNA-
1274 seq counts vs MLL, core-binding factors (CBF), or other pediatric AML. **(D)** Cell type analysis
1275 of the intersect of leading edge genes LEGs (n=476) from Fig. S5D in MNX1-r patients (k14
1276 LEGs) **(B)** using the Panglao DB 2021 database. **(E)** Representative Giemsa-Wright stained
1277 dissociated CFC replating (4th plating) MNX1 haemGx cells from 144h and 216h. Solid black
1278 arrows, mast cell precursors; dashed black arrows, mast cells.
1279
1280 **Figure 7 Supplement 1 – Transcriptional analyses of haemGx with *MNX1***
1281 **overexpression compared to patient signatures and current *in vivo* model. (A)** Venn
1282 diagram showing the intersection of leading-edge genes enriched in *MNX1*-r patient vs MLL,
1283 core-binding factors (CBF), or other pediatric AML from GSEA analysis using the K14
1284 signature. **(B)** Cell type enrichment analysis in *MNX1*-r AML transcriptomes, compared to
1285 normal paediatric bone marrow (BM) or other paediatric AML from the TARGET database, with
1286 comparison with transcriptomes of haemGx and CFCs at from MNX1 vs EV conditions at 144h
1287 and 216h. GSEA used representative cell type gene sets from the Panglao DB 2021database;
1288 bubble plot shows NES scores (color gradient) and statistical significance (–log10(FDR),
1289 bubble size). **(C)** Representative Giemsa-Wright stained dissociated CFC replating MNX1

haemGx cells from 144h; purple arrows, undifferentiated blasts; solid black arrow, mast cell precursors; dashed black arrow, mast cells. **(D)** Transcriptional re-analysis of Waraky et al (2024) in vivo model of MNX1-OE leukemia from transplanted fetal liver (FL) cells. Heatmap comparing the expression levels of all DEG between all conditions, i.e. FL control (FL Ctrl) vs FL with MNX1-OE (FL MNX1) in liquid culture, and bone marrow (BM) from mice transplanted with FL Ctrl (BM TX Ctrl) vs leukemic animals transplanted with FL MNX1 cells (BM TX MNX1). Hierarchical clustering by Ward D method on Euclidean distances identifies 12 clusters (k). Clusters k1, k2, k5, k6 and k9 are specifically up-regulated in BM TX MNX1 and putatively represent MNX1-OE leukemia programs. Analysis matches Fig. 7B for MNX1-OE haemGx. **(E)** GSEA plots of significant results (by p value < 0.05) for gene set extracted from k1, k2, k5, k6 and k9, individually or combined, from (D) against MNX1-r patients RNA-seq counts vs MLL, core-binding factors (CBF), or other pediatric AML. FDR values are indicated for each enrichment point. Analysis matches Fig. 7C for MNX1-OE haemGx. **(F)** Cell type enrichment analysis in MNX1-r AML transcriptomes, compared to normal pediatric bone marrow (BM) or other pediatric AML from the TARGET database. GSEA used representative cell type gene sets from the Panglao DB 2021 database; bubble plot shows NES scores (color gradient) and statistical significance (–log10(FDR), bubble size). Analysis matches Fig. S6E for MNX1-OE haemGx.

REFERENCES

Alexander, T.B. and Mullighan, C.G. (2021) 'Molecular biology of childhood leukemia', *Annual Review of Cancer Biology,* 5, pp. 95-117.

Ballabio, E., Cantarella, C.D., Federico, C., Di Mare, P., Hall, G., Harbott, J., Hughes, J., Saccone, S. and Tosi, S. (2009) 'Ectopic expression of the HLXB9 gene is associated with an altered nuclear position in t (7; 12) leukaemias', *Leukemia,* 23(6), pp. 1179-1183.

Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A., Lutolf, M.P., Duboule, D. and Arias, A.M. (2018) 'Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids', *Nature,* 562(7726), pp. 272-276.

Boisset, J., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E. and Robin, C. (2010) 'In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium', *Nature,* 464(7285), pp. 116-120.

Bolouri, H., Farrar, J.E., Triche Jr, T., Ries, R.E., Lim, E.L., Alonzo, T.A., Ma, Y., Moore, R., Mungall, A.J. and Marra, M.A. (2018) 'The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions', *Nature medicine,* 24(1), pp. 103.

Bousquets-Muñoz, P., Molina, O., Varela, I., Álvarez-Eguiluz, Á, Fernández-Mateos, J., Gómez, A., Sánchez, E.G., Balbín, M., Ruano, D. and Ramírez-Orellana, M. (2024) 'Backtracking NOM1:: ETV6 fusion to neonatal pathogenesis of t (7; 12)(q36; p13) infant AML', *Leukemia,* , pp. 1-5.

Bowie, M.B., McKnight, K.D., Kent, D.G., McCaffrey, L., Hoodless, P.A. and Eaves, C.J. (2006) 'Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect', *The Journal of clinical investigation,* 116(10), pp. 2808-2816.

Brignole, C., Calarco, E., Bensa, V., Giusto, E., Perri, P., Ciampi, E., Corrias, M.V., Astigiano, S., Cilli, M. and Loo, D. (2023) 'Antitumor activity of the investigational B7-H3 antibody-drug conjugate, vobramitamab duocarmazine, in preclinical models of neuroblastoma', *Journal for Immunotherapy of Cancer,* 11(9).

Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M. and Crooks, G.M. (2019) 'MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment', *Nature,* 576(7786), pp. 281-286.

Cazzola, A., Cazzaniga, G., Biondi, A., Meneveri, R., Brunelli, S. and Azzoni, E. (2021) 'Prenatal origin of pediatric leukemia: lessons from hematopoietic development', *Frontiers in cell and developmental biology,* 8, pp. 618164.

Chandrakanthan, V., Rorimpandey, P., Zanini, F., Chacon, D., Olivier, J., Joshi, S., Kang, Y.C., Knezevic, K., Huang, Y. and Qiao, Q. (2022) 'Mesoderm-derived PDGFRA cells regulate the emergence of hematopoietic stem cells in the dorsal aorta', *Nature cell biology,* 24(8), pp. 1211-1225.

Chang, Y., Hummel, S.N., Jung, J., Jin, G., Deng, Q. and Bao, X. (2023) 'Engineered hematopoietic and immune cells derived from human pluripotent stem cells', *Experimental hematology,* .

Chaudhury, S., O'Connor, C., Cañete, A., Bittencourt-Silvestre, J., Sarrou, E., Prendergast, Á, Choi, J., Johnston, P., Wells, C.A. and Gibson, B. (2018) 'Age-specific biological and molecular profiling distinguishes paediatric from adult acute myeloid leukaemias', *Nature communications,* 9(1), pp. 5280.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V. and Clark, N.R. (2013) 'A. 1126 Ma'ayan, Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC 1127', *Bioinformatics,* .

Chen, W., O'Sullivan, M.G., Hudson, W. and Kersey, J. (2011) 'Modeling human infant MLL leukemia in mice: leukemia from fetal liver differs from that originating in postnatal marrow', *Blood,* 117(12), pp. 3474.

Chia, S.L., Kapoor, S., Carvalho, C., Bajénoff, M. and Gentek, R. (2023) 'Mast cell ontogeny: From fetal development to life‐long health and disease', *Immunological reviews,* 315(1), pp. 31-53.

Chun, I.Y., Maser, R., Marches, F., Banchereau, J. and Palucka, K. (2024) 'Engraftment of adult hematopoietic stem and progenitor cells in a novel model of humanized mice' *Iscience*, *27*(3).

Ciau-Uitz, A., Pinheiro, P., Gupta, R., Enver, T. and Patient, R. (2010) 'Tel1/ETV6 specifies blood stem cells through the agency of VEGF signaling', *Developmental cell,* 18(4), pp. 569-578.

Ciriza, J., Thompson, H., Petrosian, R., Manilay, J.O. and García-Ojeda, M.E. (2013) 'The migration of hematopoietic progenitors from the fetal liver to the fetal bone marrow: lessons learned and possible clinical applications', *Experimental hematology,* 41(5), pp. 411-423.

Costa, G., Kouskoff, V. and Lacaud, G. (2012) 'Origin of blood cells and HSC production in the embryo', *Trends in immunology,* 33(5), pp. 215-223.

Dang, S.M., Kyba, M., Perlingeiro, R., Daley, G.Q. and Zandstra, P.W. (2002) 'Efficiency of embryoid body formation and hematopoietic development from embryonic stem cells in different culture systems', *Biotechnology and bioengineering,* 78(4), pp. 442-453.

Deguchi, Y. and Kehrl, J.H. (1991) 'Selective expression of two homeobox genes in CD34-positive cells from human bone marrow', *Blood,* 78(2), pp. 323-328.

Dias, A., Pascual-Mas, P., Torregrosa-Cortés, G., McNamara, H.M., Wehmeyer, A.E., Arnold, S.J., Martinez Arias A. (2025) 'A temporal coordination between Nodal and Wnt signalling governs the emergence of the mammalian body plan' *bioRxiv* 2025.01.11.632562; doi: https://doi.org/10.1101/2025.01.11.632562

Diaz, M.F., Li, N., Lee, H.J., Adamo, L., Evans, S.M., Willey, H.E., Arora, N., Torisawa, Y., Vickers, D.A. and Morris, S.A. (2015) 'Biomechanical forces promote blood development through prostaglandin E2 and the cAMP–PKA signaling axis', *Journal of Experimental Medicine,* 212(5), pp. 665-680.

Dijkhuis, L., Johns, A., Ragusa, D., van den Brink, Susanne C and Pina, C. (2023) 'Haematopoietic development and HSC formation in vitro: promise and limitations of gastruloid models', *Emerging Topics in Life Sciences,* 7(4), pp. 439-454.

Dong, Y., Shi, O., Zeng, Q., Lu, X., Wang, W., Li, Y. and Wang, Q. (2020) 'Leukemia incidence trends at the global, regional, and national level between 1990 and 2017', *Experimental hematology & oncology,* 9, pp. 1-11.

Dzierzak, E. and Bigas, A. (2018) 'Blood development: hematopoietic stem cell dependence and independence', *Cell stem cell,* 22(5), pp. 639-651.

Espersen, A.D.L., Noren‐Nyström, U., Abrahamsson, J., Ha, S., Pronk, C.J., Jahnukainen, K., Jónsson, ÓG., Lausen, B., Palle, J. and Zeller, B. (2018) 'Acute myeloid leukemia (AML) with t (7; 12)(q36; p13) is associated with infancy and trisomy 19: Data from Nordic Society for Pediatric Hematology and Oncology (NOPHO‐AML) and review of the literature', *Genes, Chromosomes and Cancer,* 57(7), pp. 359-365.

Fadlullah, M.Z.H., Neo, W.H., Lie-A-Ling, M., Thambyrajah, R., Patel, R., Mevel, R., Aksoy, I., Do Khoa, N., Savatier, P., Fontenille, L. and Baker, S.M. (2022) 'Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation marked by ACE' *Blood, 139*(3), pp.343-356.

Fehling, H.J., Lacaud, G., Kubo, A., Kennedy, M., Robertson, S., Keller, G. and Kouskoff, V. (2003) 'Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation', .

Fidanza, A. and Forrester, L.M. (2021) 'Progress in the production of haematopoietic stem and progenitor cells from human pluripotent stem cells', *Journal of Immunology and Regenerative Medicine,* 13, pp. 100050.

Fitch, S.R., Kimber, G.M., Wilson, N.K., Parker, A., Mirshekar-Syahkal, B., Göttgens, B., Medvinsky, A., Dzierzak, E. and Ottersbach, K. (2012) 'Signaling from the sympathetic nervous system regulates hematopoietic stem cell emergence during embryogenesis', *Cell stem cell,* 11(4), pp. 554-566.

Fowler, J.L., Zheng, S.L., Nguyen, A., Chen, A., Xiong, X., Chai, T., Chen, J.Y., Karigane, D., Banuelos, A.M. and Niizuma, K. (2024) 'Lineage-tracing hematopoietic stem cell origins in vivo to efficiently make human HLF HOXA hematopoietic progenitors from pluripotent stem cells', *Developmental Cell,* 59(9), pp. 1110-1131. e22.

Frankish, A., Diekhans, M., Ferreira, A., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J. and Armstrong, J. (2019) 'GENCODE reference annotation for the human and mouse genomes', *Nucleic acids research,* 47(D1), pp. D766-D773.

Franzén, O., Gan, L. and Björkegren, J.L. (2019) 'PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data', *Database,* 2019, pp. baz046.

Fujimoto, T., Ogawa, M., Minegishi, N., Yoshida, H., Yokomizo, T., Yamamoto, M. and Nishikawa, S. (2001) 'Step‐wise divergence of primitive and definitive haematopoietic and endothelial cell lineages during embryonic stem cell differentiation', *Genes to Cells,* 6(12), pp. 1113-1127.

Ganuza, M., Hall, T., Myers, J., Nevitt, C., Sánchez-Lanzas, R., Chabot, A., Ding, J., Kooienga, E., Caprio, C. and Finkelstein, D. (2022) 'Murine foetal liver supports limited detectable expansion of life-long haematopoietic progenitors', *Nature cell biology,* 24(10), pp. 1475-1486.

Ghiaur, G., Ferkowicz, M.J., Milsom, M.D., Bailey, J., Witte, D., Cancelas, J.A., Yoder, M.C. and Williams, D.A. (2008) 'Rac1 is essential for intraembryonic hematopoiesis and for the initial seeding of fetal liver with definitive hematopoietic progenitor cells', *Blood, The Journal of the American Society of Hematology,* 111(7), pp. 3313-3321.

Harrison, K.A., Thaler, J., Pfaff, S.L., Gu, H. and Kehrl, J.H. (1999) 'Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice', *Nature genetics,* 23(1), pp. 71-75.

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987) 'HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells', *Nature,* 326(6110), pp. 292-295.

Hou, S., Li, Z., Zheng, X., Gao, Y., Dong, J., Ni, Y., Wang, X., Li, Y., Ding, X., Chang, Z. and Li, S. (2020) 'Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses' *Cell research*, 30(5), pp.376-392.

Huang, Y., Lu, Y., He, Y., Feng, Z., Zhan, Y., Huang, X., Liu, Q., Zhang, J., Li, H., Huang, H. and Ma, M. (2019) 'Ikzf1 regulates embryonic T lymphopoiesis via Ccr9 and Irf4 in zebrafish' *Journal of Biological Chemistry*, *294*(44), pp.16152-16163.

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R.A., Turner, M.L. and Medvinsky, A. (2011) 'Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region', *Journal of Experimental Medicine,* 208(12), pp. 2417-2427.

Jakobsson, L., Franco, C.A., Bentley, K., Collins, R.T., Ponsioen, B., Aspalter, I.M., Rosewell, I., Busse, M., Thurston, G. and Medvinsky, A. (2010) 'Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting', *Nature cell biology,* 12(10), pp. 943-953.

Jung, H.S., Uenishi, G., Park, M.A., Liu, P., Suknuntha, K., Raymond, M., Choi, Y.J., Thomson, J.A., Ong, I.M. and Slukvin, I.I. (2021) 'SOX17 integrates HOXA and arterial programs in hemogenic endothelium to drive definitive lympho-myeloid hematopoiesis' *Cell reports*, *34*(7).

Kapeni, C., Nitsche, L., Kilpatrick, A.M., Wilson, N.K., Xia, K., Mirshekar-Syahkal, B., Chandrakanthan, V., Malouf, C., Pimanda, J.E. and Göttgens, B. (2022) 'p57Kip2 regulates embryonic blood stem cells by controlling sympathoadrenal progenitor expansion', *Blood, The Journal of the American Society of Hematology,* 140(5), pp. 464-477.

Kikuchi, K. and Kondo, M. (2006) 'Developmental switch of mouse hematopoietic stem cells from fetal to adult type occurs in bone marrow after birth', *Proceedings of the National Academy of Sciences,* 103(47), pp. 17852-17857.

Kiselev, V.Y., Yiu, A. and Hemberg, M. (2018) 'scmap: projection of single-cell RNA-seq data across data sets', *Nature methods,* 15(5), pp. 359-362.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M. and Lachmann, A. (2016) 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic acids research,* 44(W1), pp. W90-W97.

Lacaud, G. and Kouskoff, V. (2017) 'Hemangioblast, hemogenic endothelium, and primitive versus definitive hematopoiesis', *Experimental hematology,* 49, pp. 19-24.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) 'The sequence alignment/map format and SAMtools', *bioinformatics,* 25(16), pp. 2078-2079.

Li, W., Ferkowicz, M.J., Johnson, S.A., Shelley, W.C. and Yoder, M.C. (2005) 'Endothelial cells in the early murine yolk sac give rise to CD41-expressing hematopoietic cells', *Stem cells and development,* 14(1), pp. 44-54.

Lis, R., Karrasch, C.C., Poulos, M.G., Kunar, B., Redmond, D., Duran, J.G.B., Badwe, C.R., Schachterle, W., Ginsberg, M. and Xiang, J. (2017) 'Conversion of adult endothelium to immunocompetent haematopoietic stem cells', *Nature,* 545(7655), pp. 439-445.

Love, M., Huber, W. and Anders, S. (2014) 'No title', *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 Genome Biology 15 (12), 550,* .

Lundin, V., Sugden, W.W., Theodore, L.N., Sousa, P.M., Han, A., Chou, S., Wrighton, P.J., Cox, A.G., Ingber, D.E. and Goessling, W. (2020) 'YAP regulates hematopoietic stem cell formation in response to the biomechanical forces of blood flow', *Developmental cell,* 52(4), pp. 446-460. e5.

Masetti, R., Vendemini, F., Zama, D., Biagi, C., Pession, A. and Locatelli, F. (2015) 'Acute myeloid leukemia in infants: biology and treatment', *Frontiers in pediatrics,* 3, pp. 37.

McGrath, K.E. and Palis, J. (2005) 'Hematopoiesis in the yolk sac: more than meets the eye', *Experimental hematology,* 33(9), pp. 1021-1028.

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 Palis, J. (2015) 'Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo', *Cell reports,* 11(12), pp. 1892-1904.

Medvinsky, A. and Dzierzak, E. (1996) 'Definitive hematopoiesis is autonomously initiated by the AGM region', *Cell,* 86(6), pp. 897-906.

Medvinsky, A., Rybtsov, S. and Taoudi, S. (2011) 'Embryonic origin of the adult hematopoietic system: advances and questions', *Development,* 138(6), pp. 1017-1031.

Mercher, T., Raffel, G.D., Moore, S.A., Cornejo, M.G., Baudry-Bluteau, D., Cagnard, N., Jesneck, J.L., Pikman, Y., Cullen, D. and Williams, I.R. (2009) 'The OTT-MAL fusion oncogene activates RBPJ-mediated transcription and induces acute megakaryoblastic leukemia in a knockin mouse model', *The Journal of clinical investigation,* 119(4), pp. 852-864.

Miladinovic, O., Canto, P., Pouget, C., Piau, O., Radic, N., Freschu, P., Megherbi, A., Brujas Prats, C., Jacques, S. and Hirsinger, E. (2024) 'A multistep computational approach reveals a neuro-mesenchymal cell population in the embryonic hematopoietic stem cell niche', *Development,* 151(7), pp. dev202614.

Moore, M.A. and Metcalf, D. (1970) 'Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo', *British journal of haematology,* 18(3), pp. 279-296.

Moris, N., Edri, S., Seyres, D., Kulkarni, R., Domingues, A.F., Balayo, T., Frontini, M. and Pina, C. (2018) 'Histone acetyltransferase KAT2A stabilizes pluripotency with control of transcriptional heterogeneity', *Stem cells,* 36(12), pp. 1828-1838.

Murry, C.E. and Keller, G. (2008) 'Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development', *Cell,* 132(4), pp. 661-680.

Nafria, M., Bonifer, C., Stanley, E.G., Ng, E.S. and Elefanty, A.G. (2020) 'Protocol for the generation of definitive hematopoietic progenitors from human pluripotent stem cells', *STAR protocols,* 1(3), pp. 100130.

Nagel, S., Scherr, M., Quentmeier, H., Kaufmann, M., Zaborski, M., Drexler, H.G. and MacLeod, R. (2005) 'HLXB9 activates IL6 in Hodgkin lymphoma cell lines and is regulated by PI3K signalling involving E2F3', *Leukemia,* 19(5), pp. 841.

Neo, W.H., Booth, C.A., Azzoni, E., Chi, L., Delgado-Olguín, P., de Bruijn, M.F., Jacobsen, S.E.W. and Mead, A.J. (2018) 'Cell-extrinsic hematopoietic impact of Ezh2 inactivation in fetal liver endothelial cells' *Blood*, 131(20), pp.2223-2234.

Neo, W.H., Meng, Y., Rodriguez-Meira, A., Fadlullah, M.Z., Booth, C.A., Azzoni, E., Thongjuea, S., de Bruijn, M.F., Jacobsen, S.E.W., Mead, A.J. and Lacaud, G. (2021) 'Ezh2 is essential for the generation of functional yolk sac derived erythro-myeloid progenitors' *Nature Communications*, *12*(1), p.7019.

Ng, E.S., Sarila, G., Li, J.Y., Edirisinghe, H.S., Saxena, R., Sun, S., Bruveris, F.F., Labonne, T., Sleebs, N., Maytum, A. and Yow, R.Y. (2024) 'Long-term engrafting multilineage hematopoietic cells differentiated from human induced pluripotent stem cells' *Nature biotechnology*, pp.1-14.

Niakan, K.K., Ji, H., Maehr, R., Vokes, S.A., Rodolfa, K.T., Sherwood, R.I., Yamaki, M., Dimos, J.T., Chen, A.E. and Melton, D.A. (2010) 'Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal', *Genes & development,* 24(3), pp. 312-326.

North, T.E., Goessling, W., Peeters, M., Li, P., Ceol, C., Lord, A.M., Weber, G.J., Harris, J., Cutting, C.C. and Huang, P. (2009) 'Hematopoietic stem cell development is dependent on blood flow', *Cell,* 137(4), pp. 736-748.

Oatley, M., Bölükbası, Ö.V., Svensson, V., Shvartsman, M., Ganter, K., Zirngibl, K., Pavlovich, P.V., Milchevskaya, V., Foteva, V., Natarajan, K.N. and Baying, B. (2020) 'Single-cell transcriptomics identifies CD44 as a marker and regulator of endothelial to haematopoietic transition' *Nature communications*, *11*(1), p.586.

Palis, J., Robertson, S., Kennedy, M., Wall, C. and Keller, G. (1999) 'Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse', *Development,* 126(22), pp. 5073-5084.

Pastorino, F., Brignole, C., Marimpietri, D., Cilli, M., Gambini, C., Ribatti, D., Longhi, R., Allen, T.M., Corti, A. and Ponzoni, M. (2003) 'Vascular damage and anti-angiogenic effects of tumor vessel-targeted liposomal chemotherapy', *Cancer research,* 63(21), pp. 7400-7409.

Pearson, S., Cuvertino, S., Fleury, M., Lacaud, G. and Kouskoff, V. (2015) 'In vivo repopulating activity emerges at the onset of hematopoietic specification during embryonic stem cell differentiation', *Stem Cell Reports,* 4(3), pp. 431-444.

Piau, O., Brunet-Manquat, M., L'homme, B., Petit, L., Birebent, B., Linard, C., Moeckes, L., Zuliani, T., Lapillonne, H. and Benderitter, M. (2023) 'Generation of transgene-free hematopoietic stem cells from human induced pluripotent stem cells', *Cell Stem Cell,* 30(12), pp. 1610-1623. e7.

Picelli, S., Faridani, O.R., Björklund, ÅK., Winberg, G., Sagasser, S. and Sandberg, R. (2014) 'Full-length RNA-seq from single cells using Smart-seq2', *Nature protocols,* 9(1), pp. 171-181.

Pina, C., May, G., Soneji, S., Hong, D. and Enver, T. (2008) 'MLLT3 regulates early human erythroid and megakaryocytic cell fate', *Cell stem cell,* 2(3), pp. 264-273.

Quintard, C., Tubbs, E., Jonsson, G., Jiao, J., Wang, J., Werschler, N., Laporte, C., Pitaval, A., Bah, T. and Pomeranz, G. (2024) 'A microfluidic platform integrating functional vascularized organoids-on-chip', *Nature Communications,* 15(1), pp. 1452.

Ragusa, D., Cicirò, Y., Federico, C., Saccone, S., Bruno, F., Saeedi, R., Sisu, C., Pina, C., Sala, A. and Tosi, S. (2022) 'Engineered model of t (7; 12)(q36; p13) AML recapitulates patient-specific features and gene expression profiles', *Oncogenesis,* 11(1), pp. 50.

Ragusa, D., Dijkhuis, L., Pina, C. and Tosi, S. (2023) 'Mechanisms associated with t (7; 12) acute myeloid leukaemia: from genetics to potential treatment targets', *Bioscience reports,* 43(1), pp. BSR20220489.

Rossi, G., Broguiere, N., Miyamoto, M., Boni, A., Guiet, R., Girgin, M., Kelly, R.G., Kwon, C. and Lutolf, M.P. (2021) 'Capturing cardiogenesis in gastruloids', *Cell stem cell,* 28(2), pp. 230-240. e6.

Rossi, G., Giger, S., Hübscher, T. and Lutolf, M.P. (2022) 'Gastruloids as in vitro models of embryonic blood development with spatial and temporal resolution', *Scientific Reports,* 12(1), pp. 13380.

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O. and Medvinsky, A. (2014) 'Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43− embryonic precursor', *Stem cell reports,* 3(3), pp. 489-501.

Saorin, G., Caligiuri, I. and Rizzolio, F. (2023) *Microfluidic organoids-on-a-chip: The future of human models.* Elsevier, pp. 41.

Schyrr, F., Alonso-Calleja, A., Vijaykumar, A., Gebhard, S., Sarkis, R., Lopes, S.F., Oggier, A., De Leval, L., Nombela-Arrieta, C. and Naveiras, O. (2023) 'Adrenal extramedullary hematopoiesis as an inducible model of the adult hematopoietic niche', *bioRxiv,* , pp. 2023.03. 15.531679.

Souilhol, C., Gonneau, C., Lendinez, J.G., Batsivari, A., Rybtsov, S., Wilson, H., Morgado-Palacin, L., Hills, D., Taoudi, S. and Antonchuk, J. (2016) 'Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells', *Nature communications,* 7(1), pp. 10784.

Sroczynska, P., Lancrin, C., Pearson, S., Kouskoff, V. and Lacaud, G. (2009) 'In vitro differentiation of embryonic stem cells as a model of early hematopoietic development', *Leukemia: Methods and Protocols,* , pp. 317-334.

Steliarova-Foucher, E., Colombet, M., Ries, L.A., Moreno, F., Dolya, A., Bray, F., Hesseling, P., Shin, H.Y. and Stiller, C.A. (2017) 'International incidence of childhood cancer, 2001–10: a population-based registry study', *The Lancet Oncology,* 18(6), pp. 719-731.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R. and Lander, E.S. (2005) 'Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles', *Proceedings of the National Academy of Sciences,* 102(43), pp. 15545-15550.

Sugimura, R., Jha, D.K., Han, A., Soria-Valles, C., Da Rocha, E.L., Lu, Y., Goettel, J.A., Serrao, E., Rowe, R.G. and Malleshaiah, M. (2017) 'Haematopoietic stem and progenitor cells from human pluripotent stem cells', *Nature,* 545(7655), pp. 432-438.

Taketani, T., Taki, T., Sako, M., Ishii, T., Yamaguchi, S. and Hayashi, Y. (2008) 'MNX1–ETV6 fusion gene in an acute megakaryoblastic leukemia and expression of the MNX1 gene in leukemia and normal B cell lines', *Cancer genetics and cytogenetics,* 186(2), pp. 115-119.

Thaler, J., Harrison, K., Sharma, K., Lettieri, K., Kehrl, J. and Pfaff, S.L. (1999) 'Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9', *Neuron,* 23(4), pp. 675-687.

Thambyrajah, R., Maqueda, M., Neo, W.H., Imbach, K., Guillén, Y., Grases, D., Fadlullah, Z., Gambera, S., Matteini, F. and Wang, X. (2024) 'Cis inhibition of NOTCH1 through JAGGED1 sustains embryonic hematopoietic stem cell fate', *Nature Communications,* 15(1), pp. 1604.

Tosi, S., Kamel, Y.M., Owoka, T., Federico, C., Truong, T.H. and Saccone, S. (2015) 'Paediatric acute myeloid leukaemia with the t (7; 12)(q36; p13) rearrangement: a review of the biological and clinical management aspects', *Biomarker research,* 3(1), pp. 1.

Turner, D.A., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., Glodowski, C.R., Hayward, P.C., Collignon, J., Gustavsen, C. and Serup, P. (2017) 'Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids', *Development,* 144(21), pp. 3894-3906.

van den Brink, Susanne C and van Oudenaarden, A. (2021) '3D gastruloids: A novel frontier in stem cell-based in vitro modeling of mammalian gastrulation', *Trends in cell biology,* 31(9), pp. 747-759.

van den Brink, Susanne C, Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivie, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F. and Martinez Arias, A. (2020) 'Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids', *Nature,* 582(7812), pp. 405-409.

Van den Brink, Susanne C, Baillie-Johnson, P., Balayo, T., Hadjantonakis, A., Nowotschin, S., Turner, D.A. and Martinez Arias, A. (2014) 'Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells', *Development,* 141(22), pp. 4231-4242.

Veenvliet, J.V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Kumar, A.S. and Pustet, M. (2020) 'Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites', *Science,* 370(6522), pp. eaba4937.

Vo, L.T. and Daley, G.Q. (2015) 'De novo generation of HSCs from somatic and pluripotent stem cell sources', *Blood, The Journal of the American Society of Hematology,* 125(17), pp. 2641-2648.

Waraky, A., Östlund, A., Nilsson, T., Weichenhan, D., Lutsik, P., Bähr, M., Hey, J., Tunali, G., Adamsson, J. and Jacobsson, S. (2024) 'Aberrant MNX1 expression associated with t (7; 12)(q36; p13) pediatric acute myeloid leukemia induces the disease through altering histone methylation', *Haematologica,* 109(3), pp. 725.

Weichenhan, D., Riedel, A., Sollier, E., Toprak, U.H., Hey, J., Breuer, K., Wierzbinska, J.A., Touzart, A., Lutsik, P. and Bähr, M. (2023) 'Altered enhancer-promoter interaction leads to MNX1 expression in pediatric acute myeloid leukemia with t (7; 12)(q36; p13)', *bioRxiv,* , pp. 2023.09. 13.557546.

Wildenhain, S., Ingenhag, D., Ruckert, C., Degistirici, O., Dugas, M., Meisel, R., Hauer, J. and Borkhardt, A. (2012) 'Homeobox protein HB9 binds to the prostaglandin E receptor 2 promoter and inhibits intracellular cAMP mobilization in leukemic cells', *The Journal of biological chemistry,* 287(48), pp. 40703-40712 Available at: 10.1074/jbc.M111.308965 [doi].

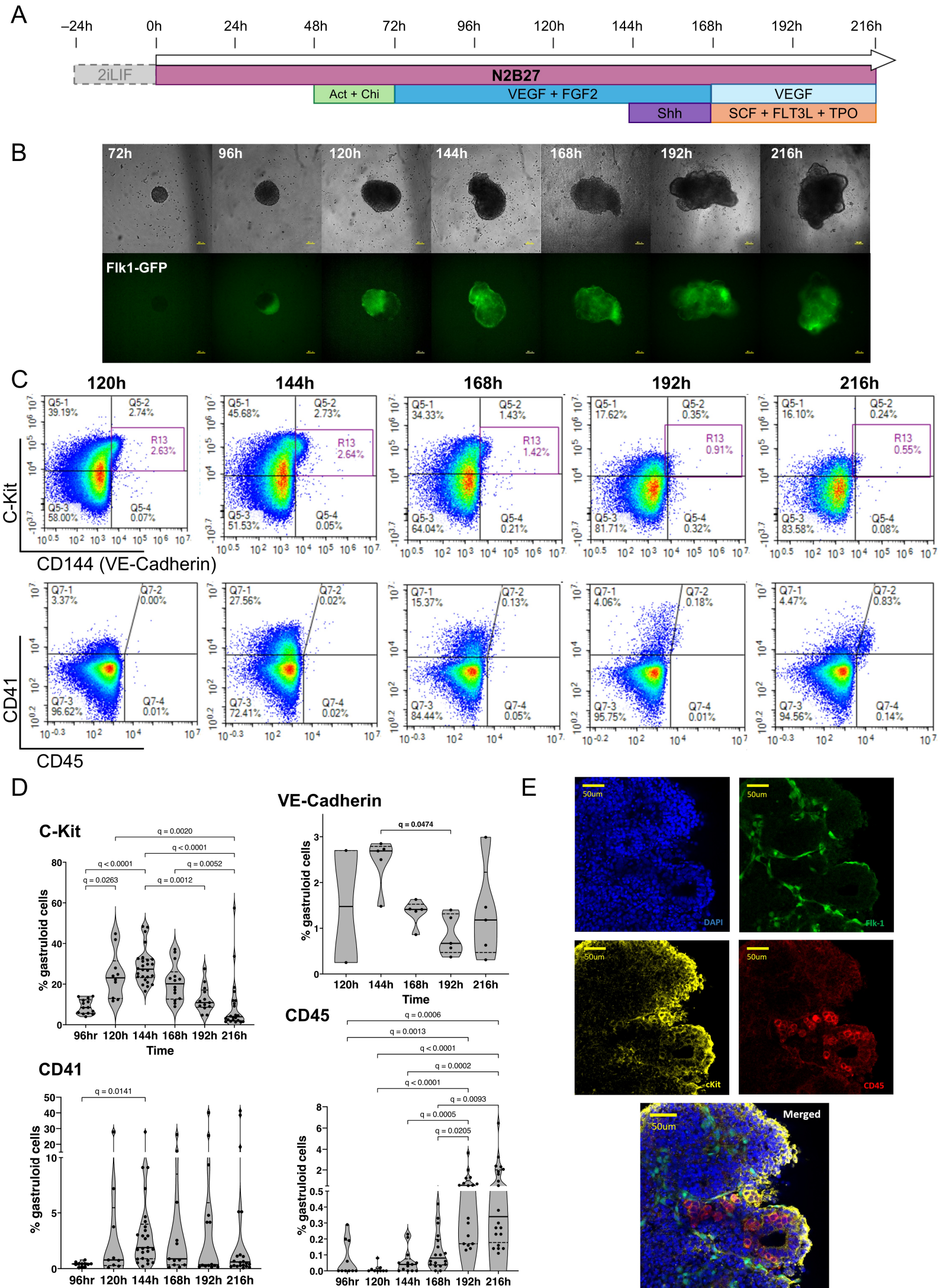
Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E. and Jagodnik, K.M. (2021) 'Gene set knowledge discovery with Enrichr', *Current protocols,* 1(3), pp. e90.

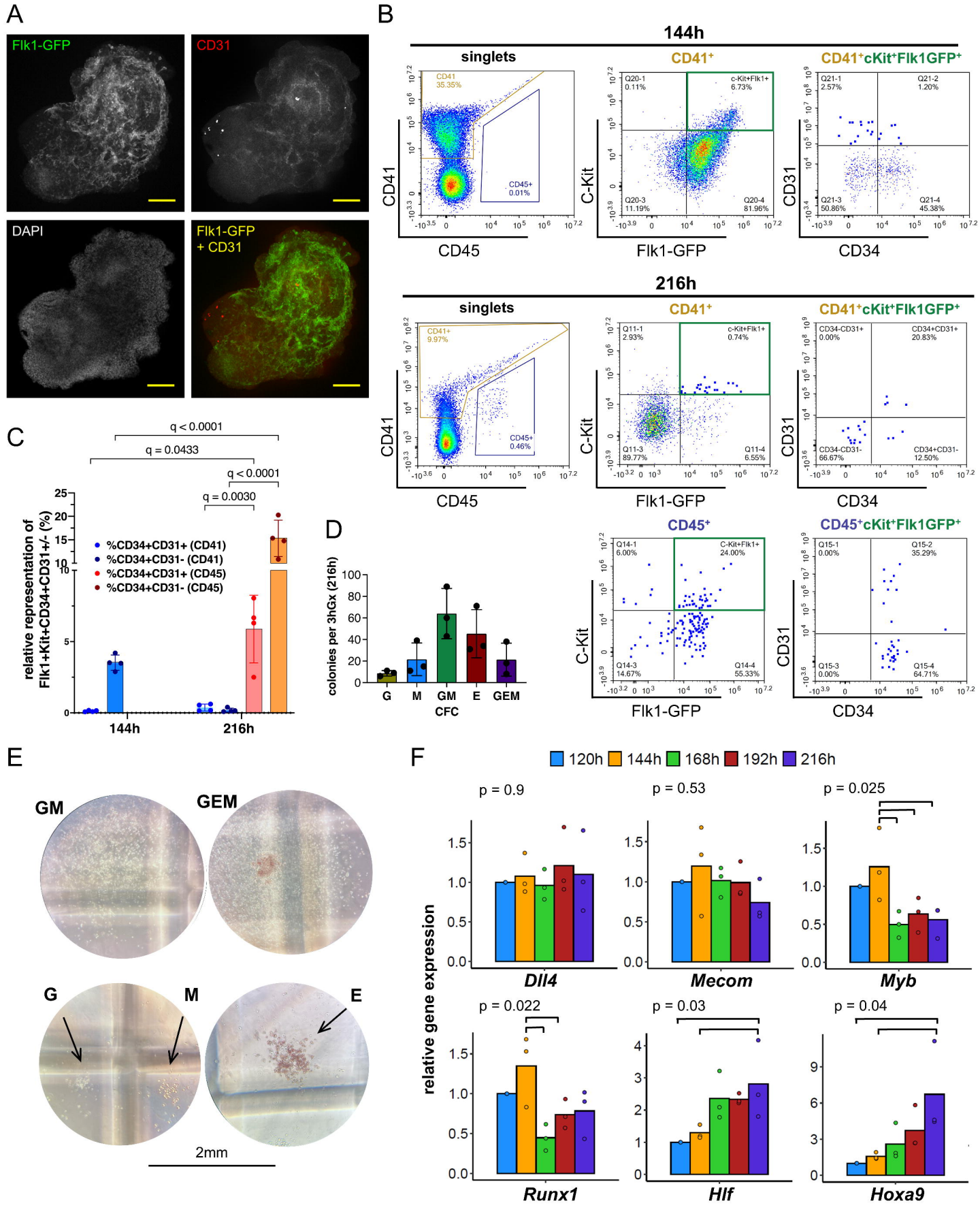
Yokomizo, T., Ideue, T., Morino-Koga, S., Tham, C.Y., Sato, T., Takeda, N., Kubota, Y., Kurokawa, M., Komatsu, N. and Ogawa, M. (2022) 'Independent origins of fetal liver haematopoietic stem and progenitor cells', *Nature,* 609(7928), pp. 779-784.

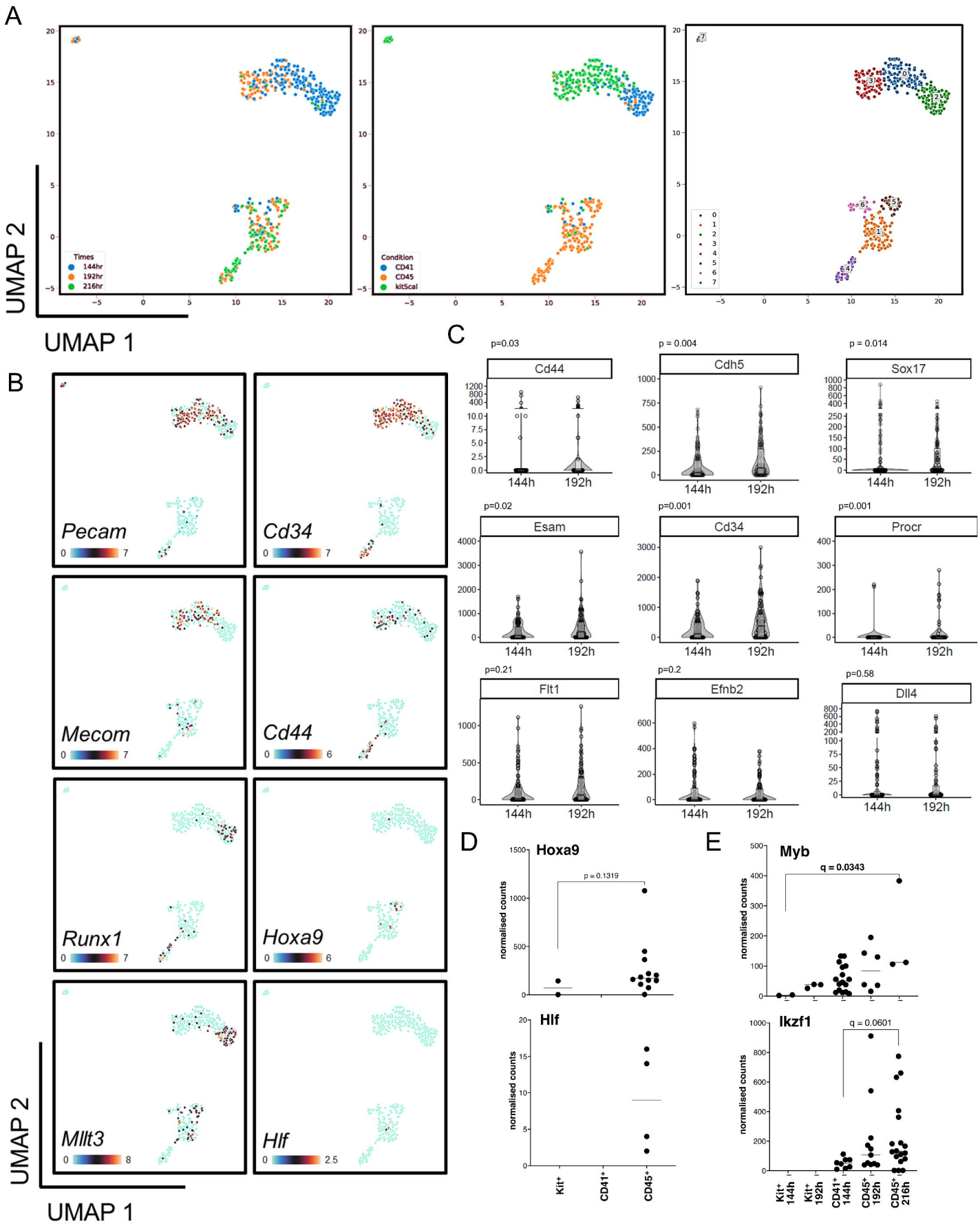
Yoshimoto, M., Kosters, A., Cornelius, S., Valiente, N., Cheng, H., Latorre, A., Nishida, C., Ghosn, E.E. and Kobayashi, M. (2022) 'Mast cell repopulating ability is lost during the transition from pre-HSC to FL HSC', *Frontiers in Immunology,* 13, pp. 896396.

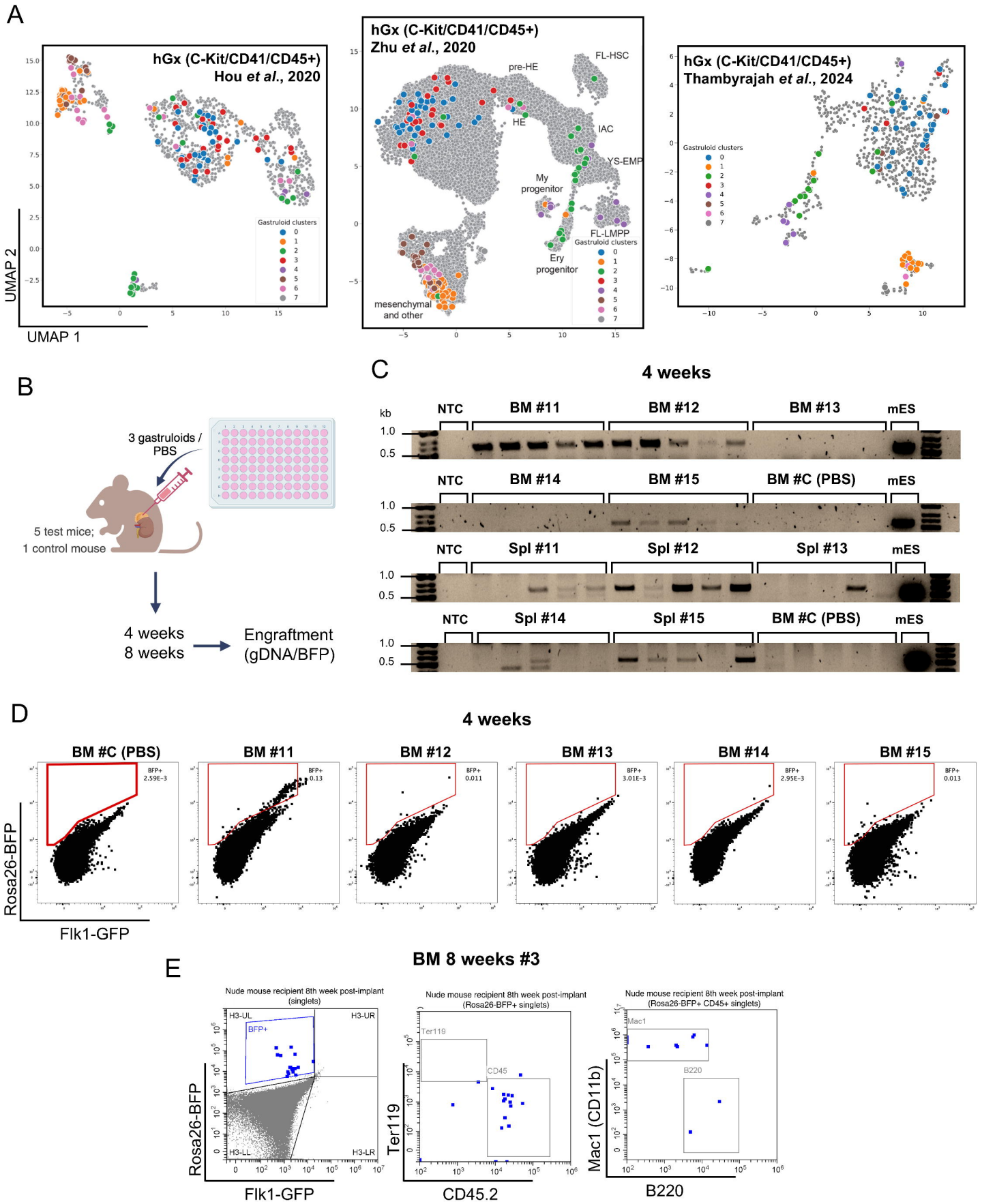
Zhu, Q., Gao, P., Tober, J., Bennett, L., Chen, C., Uzun, Y., Li, Y., Howell, E.D., Mumau, M., Yu, W. and He, B. (2020) 'Developmental trajectory of prehematopoietic stem cell formation from endothelium' *Blood*, 136(7), pp.845-856.

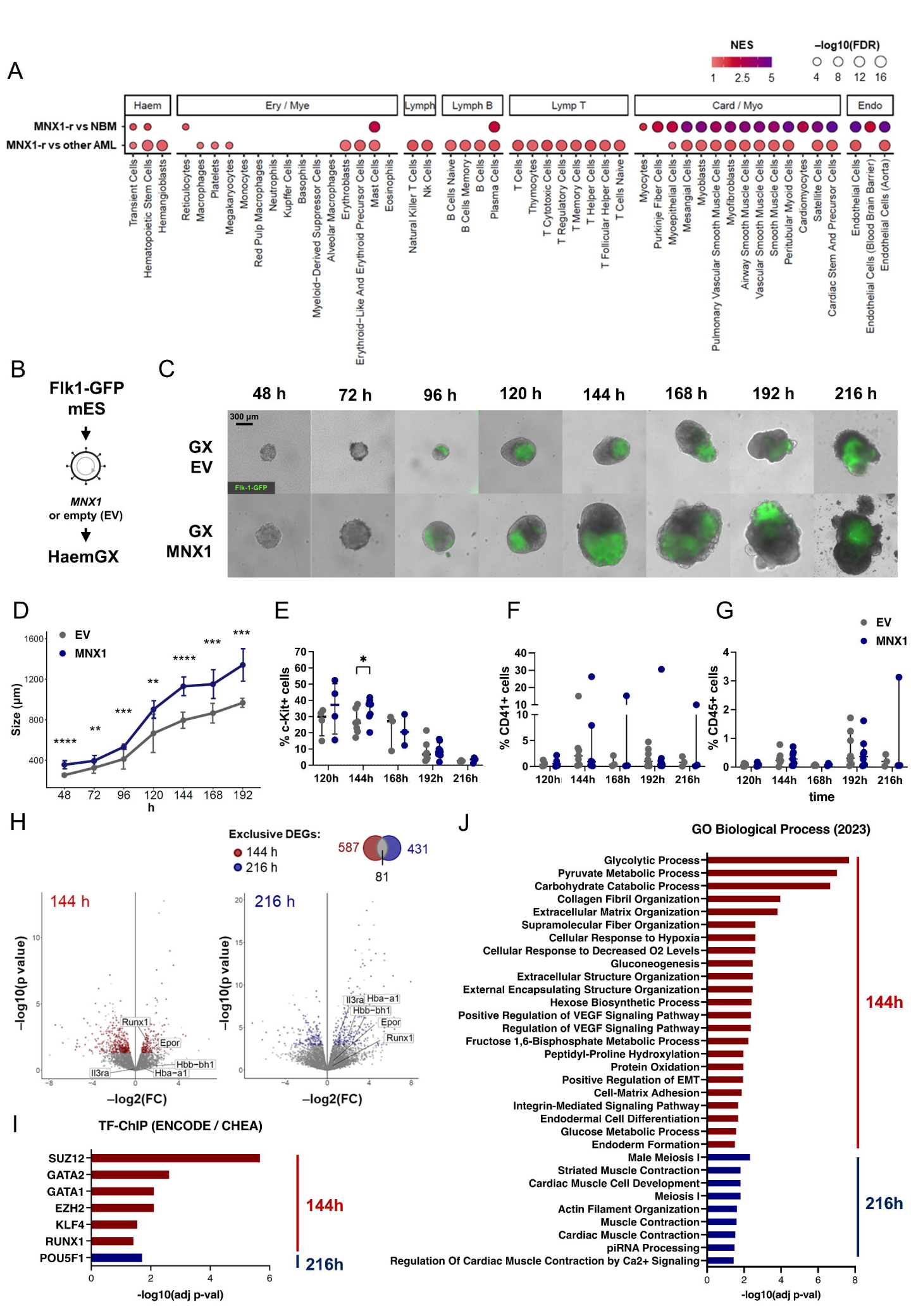
Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K.A., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E. and Gasson, J.C. (2008) 'Fate tracing reveals the endothelial origin of hematopoietic stem cells', *Cell stem cell,* 3(6), pp. 625-636.

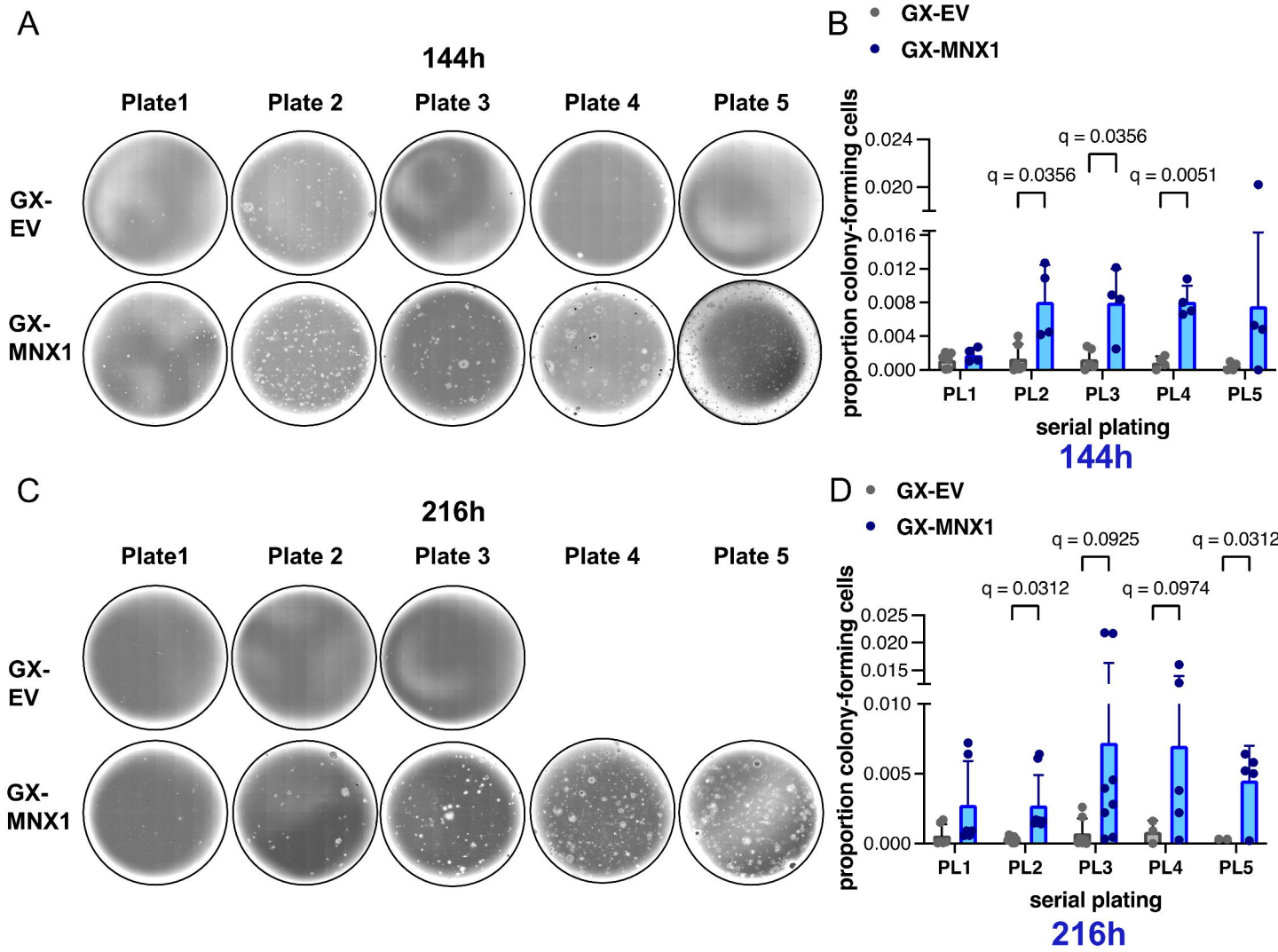

A
–24h
0h
24h
48h
72h
96h
120h
144h
168h
192h
216h
2iLIF
N2B27
Act + Chi
VEGF + FGF2
VEGF
Shh
SCF + FLT3L + TPO
B
72h
96h
120h
144h
168h
192h
216h
Flk1-GFP
C
120h
144h
168h
192h
216h
C-Kit
CD144 (VE-Cadherin)
CD41
CD45
D
C-Kit
VE-Cadherin
CD41
CD45
% gastruloid cells
Time
96hr 120h 144h 168h 192h 216h
E
50um
DAPI
Flk-1
cKit
CD45
Merged

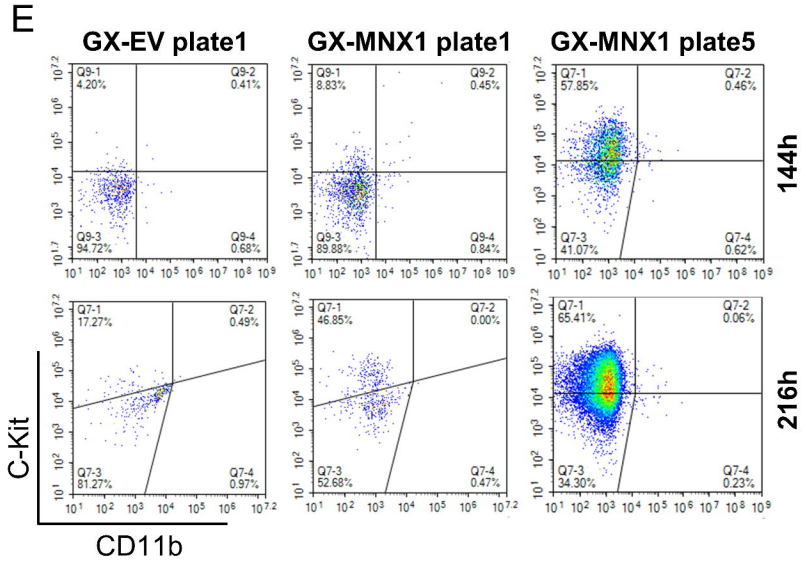

A
Flk1-GFP
CD31
DAPI
Flk1-GFP + CD31
B
144h
singlets
CD41+
CD41+cKit+Flk1GFP+
CD41
CD45
C-Kit
Flk1-GFP
CD31
CD34
216h
singlets
CD41+
CD41+cKit+Flk1GFP+
CD45+
C-Kit
Flk1-GFP
CD31
CD34
C
q < 0.0001
q = 0.0433
q < 0.0001
q = 0.0030
relative representation of Flk1+Kit+CD34+CD31+/- (%)
%CD34+CD31+ (CD41)
%CD34+CD31- (CD41)
%CD34+CD31+ (CD45)
%CD34+CD31- (CD45)
144h
216h
D
colonies per 3hGx (216h)
G
M
GM
E
GEM
CFC
E
GM
GEM
G
M
E
2mm
F
120h
144h
168h
192h
216h
p = 0.9
p = 0.53
p = 0.025
relative gene expression
Dll4
Mecom
Myb
p = 0.022
p = 0.03
p = 0.04
Runx1
Hlf
Hoxa9

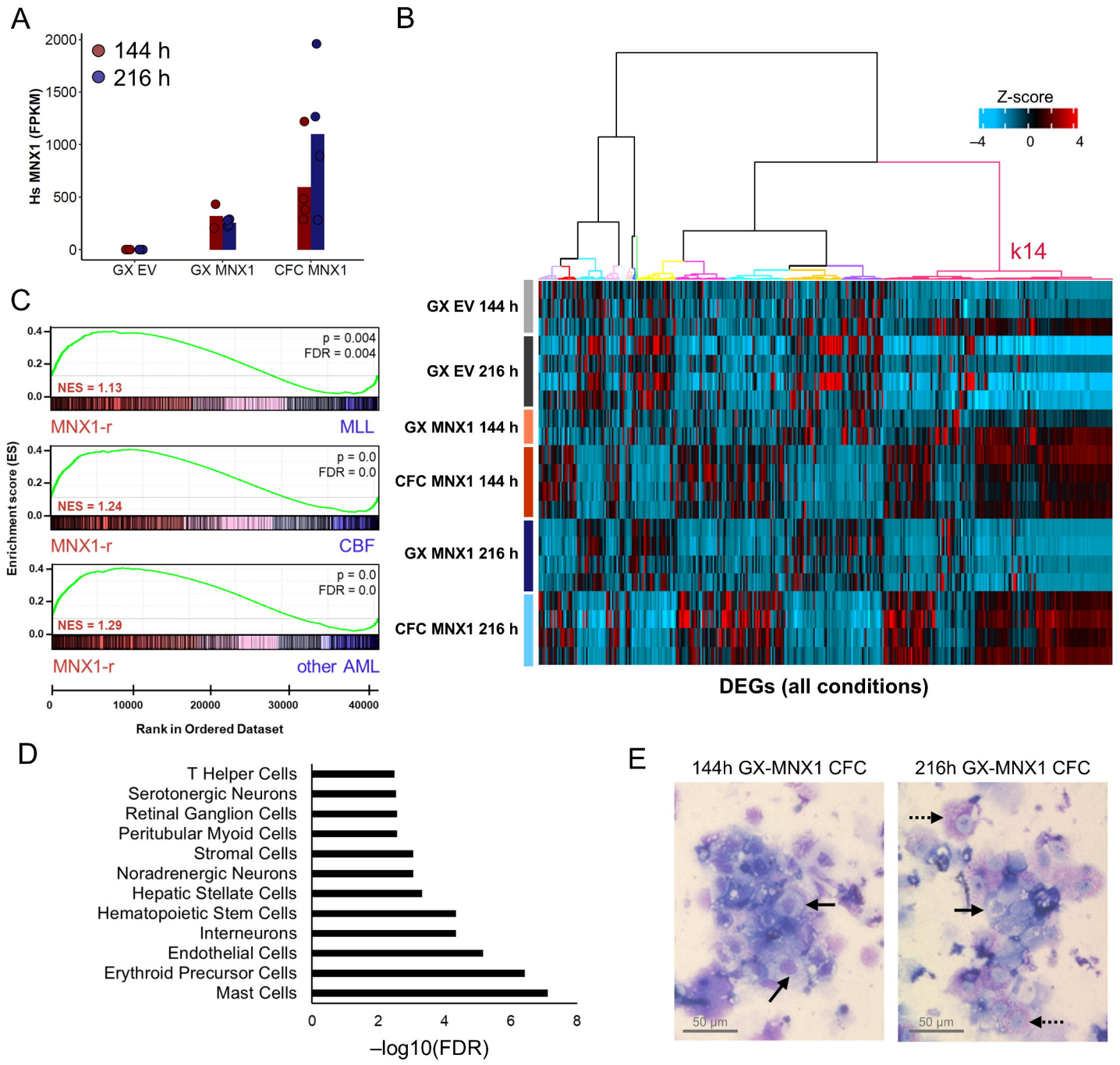

A
UMAP 2
UMAP 1
Times
144hr
192hr
216hr
Condition
CD41
CD45
kitScal
B
Pecam
Cd34
Mecom
Cd44
Runx1
Hoxa9
Mllt3
Hlf
C
p=0.03
Cd44
p = 0.004
Cdh5
p = 0.014
Sox17
p=0.02
Esam
p=0.001
Cd34
p=0.001
Procr
p=0.21
Flt1
p=0.2
Efnb2
p=0.58
Dll4
144h
192h
D
Hoxa9
p = 0.1319
Hlf
normalised counts
Kit+
CD41+
CD45+
E
Myb
q = 0.0343
Ikzf1
q = 0.0601
Kit+ 144h
Kit+ 192h
CD41+ 144h
CD45+ 192h
CD45+ 216h


A
hGx (C-Kit/CD41/CD45+)
Hou et al., 2020
Gastruloid clusters
UMAP 2
UMAP 1
hGx (C-Kit/CD41/CD45+)
Zhu et al., 2020
pre-HE
FL-HSC
HE
IAC
YS-EMP
My progenitor
FL-LMPP
Ery progenitor
mesenchymal and other
hGx (C-Kit/CD41/CD45+)
Thambyrajah et al., 2024
B
3 gastruloids / PBS
5 test mice;
1 control mouse
4 weeks
8 weeks
Engraftment (gDNA/BFP)
C
4 weeks
kb
NTC
BM #11
BM #12
BM #13
mES
BM #14
BM #15
BM #C (PBS)
Spl #11
Spl #12
Spl #13
Spl #14
Spl #15
1.0
0.5
D
4 weeks
BM #C (PBS)
BM #11
BM #12
BM #13
BM #14
BM #15
BFP+ 2.59E-3
BFP+ 0.13
BFP+ 0.011
BFP+ 3.01E-3
BFP+ 2.95E-3
BFP+ 0.013
Rosa26-BFP
Flk1-GFP
E
BM 8 weeks #3
Nude mouse recipient 8th week post-implant (singlets)
H3-UL
H3-UR
H3-LL
H3-LR
BFP+
Rosa26-BFP
Flk1-GFP
Nude mouse recipient 8th week post-implant (Rosa26-BFP+ singlets)
Ter119
CD45
CD45.2
Nude mouse recipient 8th week post-implant (Rosa26-BFP+ CD45+ singlets)
Mac1
B220
Mac1 (CD11b)


A
NES
1 2.5 5
−log10(FDR)
4 8 12 16
Haem
Ery / Mye
Lymph
Lymph B
Lymp T
Card / Myo
Endo
MNX1-r vs NBM
MNX1-r vs other AML
Transient Cells
Hematopoietic Stem Cells
Hemangioblasts
Reticulocytes
Macrophages
Platelets
Megakaryocytes
Monocytes
Red Pulp Macrophages
Neutrophils
Kupffer Cells
Basophils
Myeloid−Derived Suppressor Cells
Alveolar Macrophages
Erythroblasts
Erythroid−Like And Erythroid Precursor Cells
Mast Cells
Eosinophils
Natural Killer T Cells
Nk Cells
B Cells Naive
B Cells Memory
B Cells
Plasma Cells
T Cells
Thymocytes
T Cytotoxic Cells
T Regulatory Cells
T Memory Cells
T Helper Cells
T Follicular Helper Cells
T Cells Naive
Myocytes
Purkinje Fiber Cells
Myoepithelial Cells
Mesangial Cells
Myoblasts
Pulmonary Vascular Smooth Muscle Cells
Myofibroblasts
Airway Smooth Muscle Cells
Vascular Smooth Muscle Cells
Smooth Muscle Cells
Peritubular Myoid Cells
Cardiomyocytes
Satellite Cells
Cardiac Stem And Precursor Cells
Endothelial Cells
Endothelial Cells (Blood Brain Barrier)
Endothelial Cells (Aorta)
B
Flk1-GFP mES
MNX1 or empty (EV)
HaemGX
C
48 h
72 h
96 h
120 h
144 h
168 h
192 h
216 h
GX EV
GX MNX1
300 µm
Flk-1-GFP
D
EV
MNX1
Size (µm)
h
E
% c-Kit+ cells
120h
144h
168h
192h
216h
F
% CD41+ cells
G
% CD45+ cells
time
H
Exclusive DEGs:
144 h
216 h
587
431
81
144 h
216 h
−log10(p value)
−log2(FC)
Runx1
Epor
Hbb−bh1
Il3ra
Hba−a1
J
GO Biological Process (2023)
Glycolytic Process
Pyruvate Metabolic Process
Carbohydrate Catabolic Process
Collagen Fibril Organization
Extracellular Matrix Organization
Supramolecular Fiber Organization
Cellular Response to Hypoxia
Cellular Response to Decreased O2 Levels
Gluconeogenesis
Extracellular Structure Organization
External Encapsulating Structure Organization
Hexose Biosynthetic Process
Positive Regulation of VEGF Signaling Pathway
Regulation of VEGF Signaling Pathway
Fructose 1,6-Bisphosphate Metabolic Process
Peptidyl-Proline Hydroxylation
Protein Oxidation
Positive Regulation of EMT
Cell-Matrix Adhesion
Integrin-Mediated Signaling Pathway
Endodermal Cell Differentiation
Glucose Metabolic Process
Endoderm Formation
Male Meiosis I
Striated Muscle Contraction
Cardiac Muscle Cell Development
Meiosis I
Actin Filament Organization
Muscle Contraction
Cardiac Muscle Contraction
piRNA Processing
Regulation Of Cardiac Muscle Contraction by Ca2+ Signaling
144h
216h
-log10(adj p-val)
I
TF-ChIP (ENCODE / CHEA)
SUZ12
GATA2
GATA1
EZH2
KLF4
RUNX1
POU5F1
144h
216h
-log10(adj p-val)


A
144h
Plate1
Plate 2
Plate 3
Plate 4
Plate 5
GX-EV
GX-MNX1
B
GX-EV
GX-MNX1
proportion colony-forming cells
q = 0.0356
q = 0.0356
q = 0.0051
PL1
PL2
PL3
PL4
PL5
serial plating
144h
C
216h
Plate1
Plate 2
Plate 3
Plate 4
Plate 5
GX-EV
GX-MNX1
D
GX-EV
GX-MNX1
proportion colony-forming cells
q = 0.0925
q = 0.0312
q = 0.0312
q = 0.0974
PL1
PL2
PL3
PL4
PL5
serial plating
216h


E
GX-EV plate1
GX-MNX1 plate1
GX-MNX1 plate5
144h
216h
C-Kit
CD11b


A
144 h
216 h
Hs MNX1 (FPKM)
2000
1500
1000
500
0
GX EV
GX MNX1
CFC MNX1
B
Z-score
–4
0
4
k14
GX EV 144 h
GX EV 216 h
GX MNX1 144 h
CFC MNX1 144 h
GX MNX1 216 h
CFC MNX1 216 h
DEGs (all conditions)
C
p = 0.004
FDR = 0.004
NES = 1.13
MNX1-r
MLL
p = 0.0
FDR = 0.0
NES = 1.24
MNX1-r
CBF
p = 0.0
FDR = 0.0
NES = 1.29
MNX1-r
other AML
Enrichment score (ES)
0.4
0.2
0.0
0
10000
20000
30000
40000
Rank in Ordered Dataset
D
T Helper Cells
Serotonergic Neurons
Retinal Ganglion Cells
Peritubular Myoid Cells
Stromal Cells
Noradrenergic Neurons
Hepatic Stellate Cells
Hematopoietic Stem Cells
Interneurons
Endothelial Cells
Erythroid Precursor Cells
Mast Cells
0
2
4
6
8
–log10(FDR)
E
144h GX-MNX1 CFC
216h GX-MNX1 CFC
50 µm
50 µm