Developmental Biology
Cancer Biology

Dissecting infant leukemia developmental origins with a hemogenic gastruloid model

Reviewed Preprint
v1 • February 3, 2025
Not revised

Denise Ragusa, Chun-Wai Suen, Gabriel Torregrosa-Cortés, Fabio Pastorino, Ayona Johns, Ylenia Cicirò, Liza Dijkhuis, Susanne van den Brink, Michele Cilli, Connor Byrne, Giulia-Andreea Ionescu, Joana Cerveira, Kamil R Kranc, Victor Hernandez-Hernandez, Mirco Ponzoni, Anna Bigas, Jordi Garcia-Ojalvo, Alfonso Martinez Arias, Cristina Pina

College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom • Centre for Genome Engineering and Maintenance, Brunel University London, London, United Kingdom • Department of Genetics, University of Cambridge, Cambridge, United Kingdom • Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain • Laboratory of Experimental Therapies in Oncology, IRCCS Istituto Giannina Gaslini, Genoa, Italy • Program in Cancer Research, Hospital de Mar Research Institute, Barcelona, Spain • Josep Carreras Leukemia Research Institute, Barcelona, Spain • Animal Facility, IRCCS Policlinico San Martino, Genova, Italy • Department of Pathology, University of Cambridge, Cambridge, United Kingdom • Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom • Centre for In Vivo Modelling, Institute of Cancer Research, London, United Kingdom

eLife Assessment

This **valuable** study presents a mouse gastruloid model that can be used to generate hematopoietic progenitors as well as leukemic cells. However, in its current form, the manuscript is **inadequate** because the primary claims are not supported. Overall, the hematopoietic progenitor cells generated in this system need to be better defined.

https://doi.org/10.7554/eLife.102324.1.sa4

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 (hGx) that captures multi-wave blood formation, progenitor specification from hemogenic endothelium (HE), and generates hematopoietic SC precursors capable of short-term engraftment of immunodeficient mice upon maturation in an adrenal niche. We took advantage of the hGx 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 hGx promotes the expansion and in vitro transformation of yolk sac-like erythroid-myeloid progenitors (EMP) 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 hGx 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 paediatric leukemias originate *in utero*, and a subset of these are driven by mutations exclusive to the paediatric group (Masetti *et al*., 2015; Cazzola *et al*., 2021). Such pediatric-exclusive mutations target blood cell types restricted to development or depend on signals that are specific to embryonic hematopoietic niches, effectively defining distinct biological entities to adult malignancies (Cazzola *et al*., 2021; Bolouri *et al*., 2018; Chaudhury *et al*., 2018). Dissection of these forms of developmental leukemia is confounded by the inability to robustly identify and isolate their embryonic cells of origin, and/or to fully recapitulate the disease in post-natal cells, which leads to failure in generating relevant targeted therapies (Alexander and Mullighan, 2021). One such example is t(7;12)(q36;p13) acute myeloid leukemia (AML) driven by rearrangement and overexpression of the *MNX1* gene locus: *MNX1*-rearranged (MNX1-*r*) AML is restricted to infants under the age of 2 and carries a high risk of relapse and poor survival (Ragusa *et al*., 2023). *MNX1* is a developmental gene required for specification of motor neurons and pancreatic acinar cells (Thaler *et al*., 1999; Harrison *et al*., 1999). MNX1 is not known to participate in embryonic haematopoiesis, however it has been discordantly described to be expressed in some subsets of adult haematopoietic cells (Deguchi and Kehrl, 1991; Nagel *et al*., 2005; Wildenhain *et al*., 2012; Taketani *et al*., 2008). It is ectopically expressed in AML cells carrying one of a variable set of t(7;12) translocations which place the normally silent *MNX1* gene under the control of *ETV6* regulatory elements (Weichenhan

et al., 2023; Bousquets-Muñoz *et al*., 2024; Ragusa *et al*., 2023; Ballabio *et al*., 2009). *ETV6* is required for hematopoietic cell specification from HE via VEGF signalling (Ciau-Uitz *et al*., 2010), potentially positioning *MNX1* within a critical embryonic hemogenic regulatory network.

Faithful *in vitro* recapitulation of developmental blood cell types and their embryonic niches has the potential to dissect the cellular and molecular mechanisms presiding at initiation of developmental leukemia, reveal prognostic biomarkers, and indicate therapeutic vulnerabilities which are otherwise challenging to test in the embryo. Despite widespread use of embryonic stem (ES) cell and induced pluripotent stem (iPS) cell-based models of blood specification, a system which recapitulates the spatial and temporal complexity of embryonic blood formation within the respective time-dependent niches is still lacking (Dijkhuis *et al*., 2023). In the last few years, gastruloid models have emerged as robust avatars of early development, self-organising processes of symmetry breaking, elongation, multi-axis formation, somitogenesis and early organogenesis, with remarkable similarity to embryo processes in space and time (van den Brink, Susanne C and van Oudenaarden, 2021; Beccari *et al*., 2018; van den Brink, Susanne C *et al*., 2020). Gastruloids are grown individually in a multi-well format that facilitates tracking and screening of developmental processes. Relevant to blood formation, a variant gastruloid culture system, which successful organises a heart primordium (Rossi *et al*., 2021), was also able to recapitulate YS blood formation with erythroid and myeloid cellularity (Rossi *et al*., 2021; Rossi *et al*., 2022).

Herein, we describe a new protocol of gastruloid formation, i.e. hemogenic gastruloids (hGx), that captures YS and AGM-like waves of definitive blood progenitor specification from HE. The AGM-like wave includes cells capable of short-term hematopoietic engraftment of the spleen and BM of immunodeficient mice, overall suggesting nearly full EMP-to-MPP/HSC recapitulation of embryonic blood development. We harnessed the embryonic haematopoietic context provided by hGx for modelling MNX1-*r* infant leukemia by proxy of *MNX1* overexpression. MNX1-overexpressing hGx showed expansion of cells at the HE-to-EMP transition, which became susceptible to transformation *in vitro* and recapitulated MNX1-*r* AML patient signatures, placing *MNX1*'s leukemogenic effects at a very early stage of developmental blood formation.

Results

HGxs capture cellularity and topography of developmental blood formation

With the aim of recapitulating embryonic haematopoiesis with spatial and temporal accuracy, we modified existing gastruloid formation protocols to promote the specification of hemogenic mesoderm derivatives and extend developmental time beyond the E8.0-equivalent of early developmental gastruloids (Van den Brink, Susanne C *et al*., 2014) and the E9.0 timepoint of cardiogenic gastruloids (Rossi *et al*., 2022). We established a 216h hGx protocol (**Fig. 1A**) using a *Flk1-GFP* mouse ES cell line (Jakobsson *et al*., 2010), which allows the tracking of early hemato-endothelial specification by expression of the *Flk1* (i.e. *Kdr*) *locus*. Activation of the *Flk1* expression, visible by fluorescence of the GFP tag, necessitated induction of the TGF-β signalling pathway by Activin A, in addition to the characteristic gastruloid patterning via WNT activation by CHI99021 supplementation at 48h (Fig. S1A). We sought to promote hemato-endothelial programmes through addition of VEGF and FGF2 at 72h (Sroczynska *et al*., 2009) (**Fig. 1A**), and observed extension of a polarised *Flk1-GFP*-expressing cellular network (**Fig. 1B**), which included VE-cadherin$^+$ C-Kit$^+$ cells suggestive of EHT (**Fig. 1C-D**). This process of HE specification and EHT peaked at 120-144h; it was followed by a surge in CD41$^+$ candidate hematopoietic cells at 144h (**Fig. 1D**). In our experience, observation of early polarised activation of the *Flk1 locus* at 96h (**Fig. 1B**) was predictive of CD41$^+$ detection 2 days later, suggesting that TGF-β activation in early patterning is required for later hematopoietic development, putatively by modifying mesoderm specification. We extended the protocol by incorporation of a 24-hour pulse of sonic

hedgehog (Shh) between 144-168h to mimic the aortic patterning that precedes cluster formation (**Fig. 1A**) (Rybtsov *et al*., 2014). Continued exposure to VEGF after the Shh pulse resulted in specification of $CD45^+$ hematopoietic cells progenitors (**Fig. 1D-E**), which increased after 168h. The proportion of $CD45^+$ cells was enhanced by the addition of SCF, FLT3L and TPO (**Fig. S1B-C**), a combination of cytokines normally used for HSC maintenance; conversely, it did not require continued addition of FGF2 after 168h (**Fig. S1D**), which was omitted. $CD45^+$ cells appeared in a small number of 'cluster'-like structures adjacent to $CD31^+$ endothelium (**Fig. 1E**), which were reminiscent of HSC and progenitor cell budding from the aortic HE, and suggested recapitulation of hemogenic tissue architecture. Time-dependent specification of $C\text{-}Kit^+$, $CD41^+$ and $CD45^+$ cells in hGxs could be obtained with other mouse ES cell lines, including commonly used E14Tg2a (E14) cells (**Fig. S1E-G**), with successful specification of comparable frequencies of $CD45^+$ progenitors (**Fig. S1H**), attesting to the reproducibility of the protocol.

Overall, our extended hGx model could capture successive specification of HE, $CD41^+$ and $CD45^+$ cells, with temporal and spatial congruence relative to the embryo, thus showing initial promise as a tool for the exploration of normal and perturbed developmental hematopoiesis.

To better characterize the extent and progression of developmental hematopoiesis in hGx, we performed single-cell RNA-sequencing (scRNA-seq) time-course analysis of gastruloid cell specification. We sorted cells from 2 independent hGx cultures at 120, 144, 168, 192 and 216h, and profiled a total of 846 cells using the Smart-Seq2 protocol (Picelli *et al*., 2014) (Fig. S2A). We analysed unfractionated live single hGx cells at 120-192h and enriched the dataset with sorted $CD41^+$ cells at their peak time of 144h, and sorted $CD45^+$ cells at 192 and 216h. At timepoints of significant enrichment in $CD41^+$ (144h) and $CD45^+$ cells (192h) (**Fig. 1D**), we also sorted $C\text{-}Kit^+$ cells. Library preparation and sequencing generated an average of 120000 reads/cell, which were mapped to an average of 4000 genes/cell, with minimal signs of cell stress / death as measured by mitochondrial DNA fraction (**Fig. S2B**). Read and gene counts were similar between biological replicates, cell types and at different time points (**Fig. S2C**). The exception was 120h unfractionated hGx cells, which despite similar sequencing depth (average read counts), were mapped to twice the number of genes (**Fig. S2C**), possibly reflecting multi-gene program priming at the onset of hemogenic specification. We selected highly varying genes (HVG) before principal component analysis (PCA) dimensionality reduction and retained the most relevant dimensions. In the PCA reduced space, we constructed a KNN graph and used uniform manifold approximation and projection (UMAP) to visualize the data on 2 dimensions, looking for cell communities using Leiden clustering (**Fig. 2A**). We identified 12 cell clusters of which 2 (clusters 0 and 5) almost exactly mapped to $C\text{-}Kit^+$ cells (also $Sca\text{-}1^+$), 2 (clusters 1 and 8) contained $CD45^+$ sorted cells, and cluster 4 uniquely captured the $CD41^+$ cells observed at 144h (**Fig. 2A**). Although some unfractionated cells overlapped with the hemogenic cell clusters (**Fig. 2A** - center), most occupied different transcriptional spaces, reflecting the frequency of hematopoietic cells and suggesting the presence of potential hemogenic niches.

To explore tissue and lineage affiliation of the different clusters, we performed differential gene expression against all other cells using Wilcoxon ranking test, and established cluster classifier gene lists (**Supplemental File S1**). Top classifier genes for clusters 0 and 5 had endothelial affiliation (**Supplemental File S1**), and the cellular composition of the clusters broadly reflected their time signature (**Fig. 2A**): cluster 5 comprised 144h cells and cluster 0 included most $C\text{-}Kit^+$ cells at 192h. Cluster 5 occupied the vicinity of 144h-$CD41^+$ cells (**Fig. 2A**), which carried an erythroid-biased signature (**Supplemental File S1**). On the other hand, 192/216h-$CD45^+$ cells in cluster 8 expressed myeloid and lymphoid genes (**Supplemental File S1**). The remaining clusters largely corresponded to time-restricted unfractionated cells (**Fig. 2A** - right), with clusters of cells present at later time-points matching mesenchymal stromal cell and autonomic neuron signatures (**Fig. S2D**), potentially configuring incipient formation of niches relevant to production of HSC (Kapeni *et al*., 2022; Fitch *et al*., 2012).

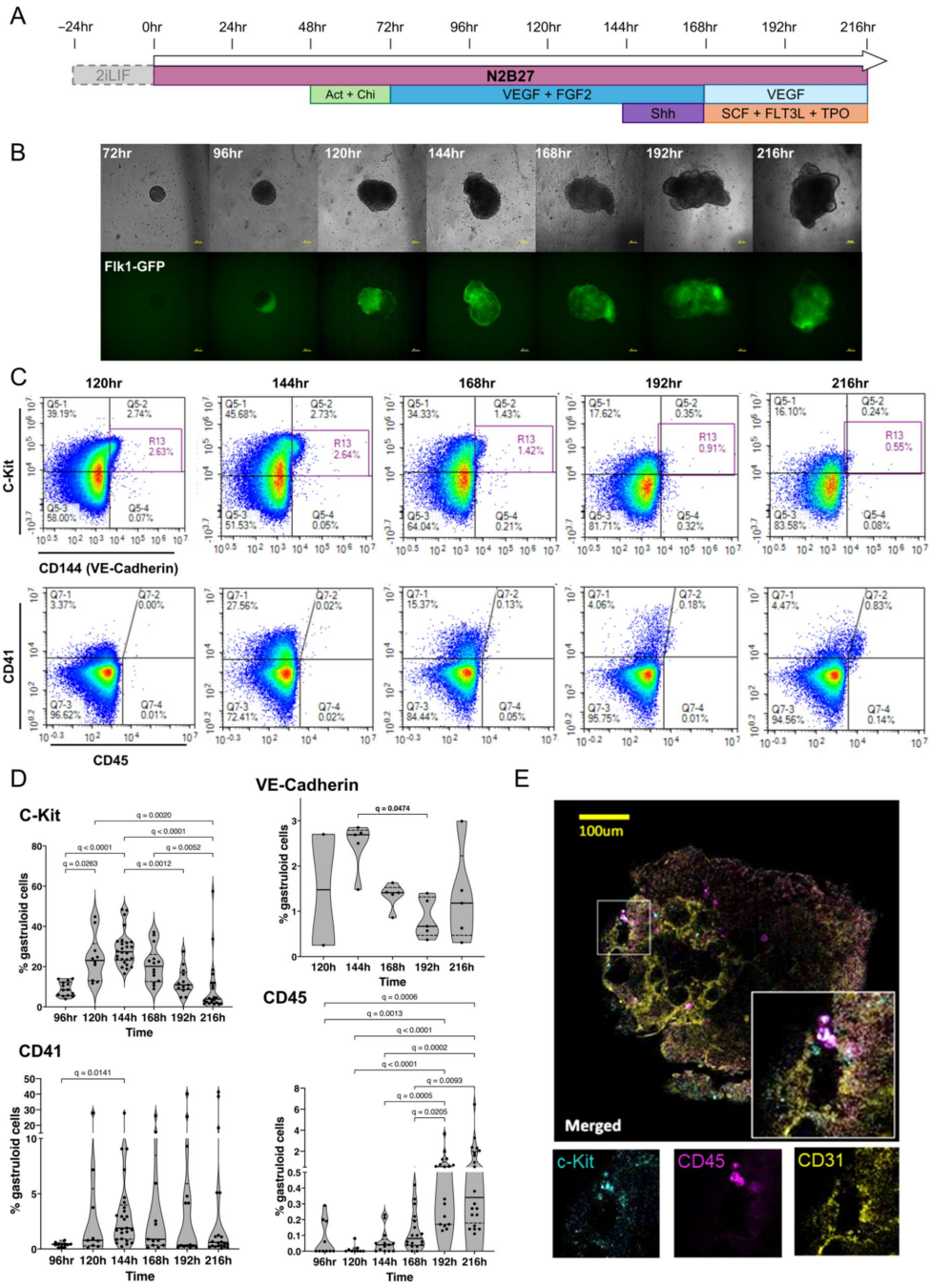


Figure 1

Hemogenic gastruloids (hGx) produced from mES cells promote hemato-endothelial specification with spatio-temporally accurate ontogeny.

(A) Timeline of mES cells assembly and culture into gastruloids over a 216h period with the addition of specified factors for the promotion of hemato-endothelial specification. The 24-hour pre-treatment in 2i+LIF was omitted when mES cultures showed no signs of differentiation. (B) Imaging of hGx over time from 72h to 216h at 10x magnification, showing the assembly and growth of the 3D structures and the polarization of the Flk-1-GFP marker from 96h; scale bar: 100mm. (C) Flow cytometry timecourse analysis of hGx for the presence of CD144 (VE-cadherin), C-Kit, CD41, and CD45 markers; representative plots. (D) Quantification of analysis in (D). Violin plots with median and interquartile range of up to 26 replicates; Kruskal-Wallis CD144 (p=0.0755), C-Kit (p<0.0001), CD41 (p<0.0342), CD45 (p<0.0001). Shown are significant Dunn's multiple comparisons tests; adjusted q-value. (E) Immunostaining of whole individual gastruloids at 216h showing the localized expression of CD31 (yellow), C-Kit (turquoise), and CD45 (pink) markers; scale bar: 100µm.

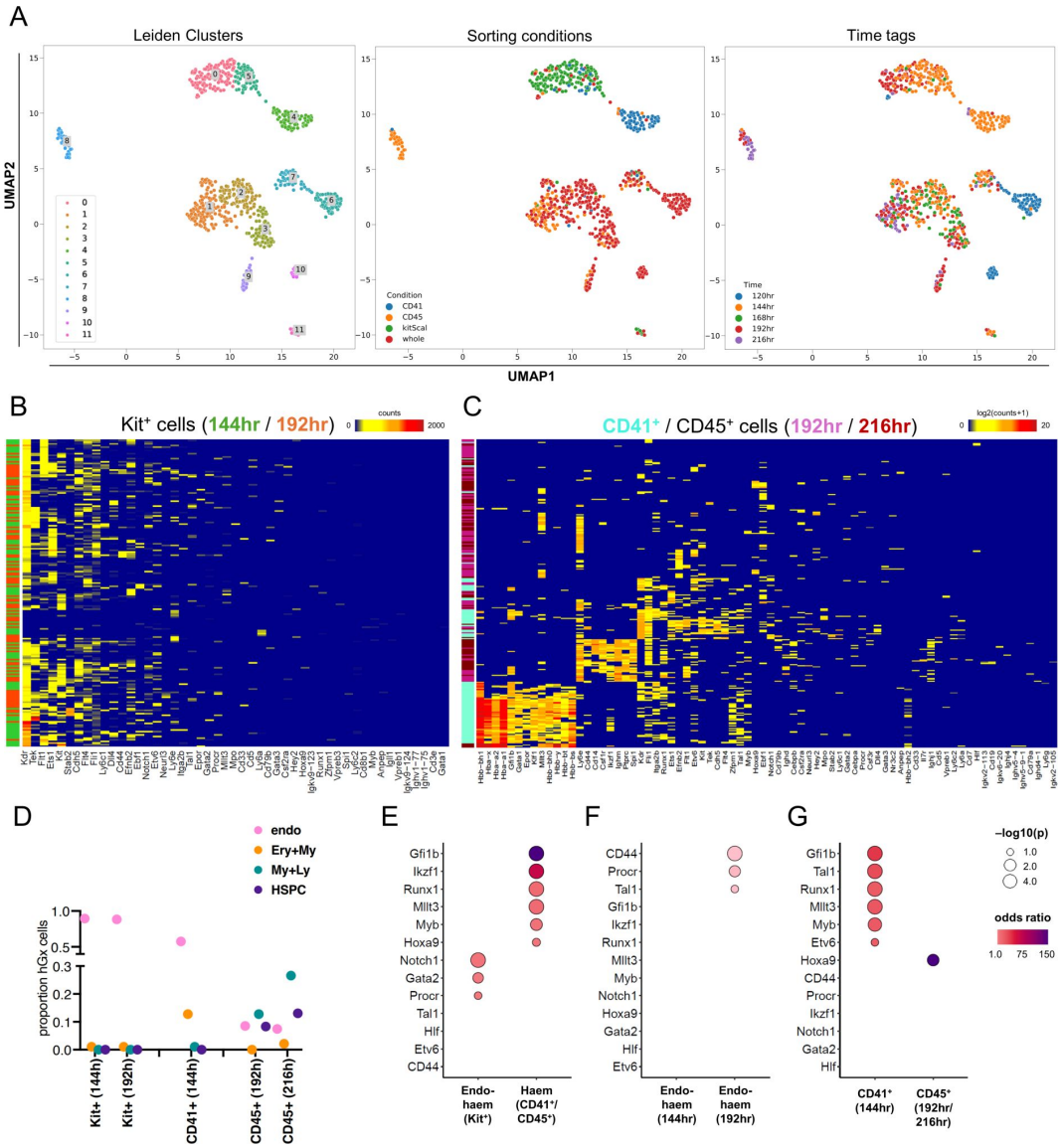


Figure 2

Time-resolved analysis of scRNA-seq of hGxs captures successive waves of hematopoietic specification.

(A) UMAP projection of all sequenced cells colored by annotated clusters (left), by sorting markers C-Kit/ScaI, CD41, and CD45, or unfractionated cells ('whole' corresponding to 'all' in panel A) (center), and by hGx culture time (right). **(B)** Heatmap representation of the expression of endothelial and hematopoietic marker genes in individual hGx C-Kit$^+$ cells sorted at 144 and 192h. Cells and genes ordered by unsupervised hierarchical clustering; cell cluster dendrogram in Fig. S2E. **(C)** Heatmap representation of the expression of endothelial and hematopoietic marker genes in individual hGx CD41$^+$ and CD45$^+$ cells sorted at 144 and 192/216h, respectively. Cells and genes ordered by unsupervised hierarchical clustering; cell cluster dendrogram in Fig. S2F. **(D)** 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*). **(E-G)** Cell type-associations of selected surface markers and transcription factors affiliated with the HE-to-hematopoietic transition (*CD44*, *Procr*, *Etv6*, *Gata2*, *Notch1*, *Tal1*) and with hematopoietic (*Gfi1b*, *Mllt3*, *Runx1*) and definitive hematopoietic (*Ikzf1*, *Hlf*, *Hoxa9*, *Myb*) gene expression programs with **(E)** C-Kit$^+$ (endo-haem) and CD41$^+$/CD45$^+$ (haem), **(F)** endo-haem cells obtained at 144h or 192h, and **(G)** CD41$^+$ or CD45$^+$ cells. Gene expression analysed as present (non-zero expression) or absent (zero-expression). Odds ratio with significance (-log$_{10}$p<1) estimated by Fisher's test.

To further characterise the identity of hGx cells, we interrogated lineage priming in individual hemato-endothelial cells (**Fig. 2B-C**), to probe the presence of putative multilineage YS (EMP) and AGM (MLP/MPP) progenitors, and HE cells. The heatmaps (**Fig. 2B-C**) show expression of lineage markers and regulators, with unsupervised clustering of cells (**Fig. S2E-F**) and genes. The gene list used was identical in both analyses, with only expressed genes shown. Concordance between sorting antigens and expression of the respective genes (*Kit*, *Itga2b* [CD41] and *Ptprc* [CD45]) was partial potentially reflecting low levels of transcript expression in the hGx system. Summary of endothelial, co-expressing erythro-myeloid (EM) and co-expressing myelo-lymphoid (ML) programs (**Fig. 2D**) clearly associates the first with C-Kit$^+$ and the latter with CD45$^+$ cells; CD41$^+$ cells capture a heterogeneous population, which includes a proportion of cells co-expressing erythroid and myeloid genes that constitute candidate EMPs.

Clustering of cells sorted as C-Kit$^+$ did not separate the 144h and 192h time-points (**Figs. 2B** and **S2E**), indicating that the relative separation on the global UMAP was not based on endothelial identity genes. We observed widespread expression of arterial genes (e.g. *Flt1*, *Efnb2*) with relative reciprocity with venous endothelial genes (e.g. *Flt4*) supporting specification of both types of endothelia (**Fig. 2B**). There was limited co-expression of hematopoietic genes, but genes associated with endothelial-to-hematopoietic transition (EHT), e.g. *Gata2*, *Notch1*, and *Procr*, were significantly more frequent within the C-Kit$^+$ population (**Fig. 2E**). Specific associations of *Cd44* and *Procr* with the 192h time-point (**Fig. 2F**) putatively represent aorta-like EHT.

Clustering of CD41$^+$ and CD45$^+$ cells (**Fig. 2C** and **S2F**) more clearly separates unique hematopoietic signatures. A subset of cells, predominantly CD41$^+$, expresses high levels of globin genes. While embryonic β-globin chains are highly expressed, there is a gain in fetal *Hbb-y* and adult *Hbb-bt* and *Hbb-bs* (C57Bl/6 strain-specific β major and β minor, respectively) in CD45$^+$ cells at 216h, indicative of AGM-like progenitor identity (**Fig. 2C**). 192 and 216h cells with high expression of *Ptprc* have an MLP profile (**Fig. 2C-D**), co-expressing myeloid and lymphoid genes including immunoglobulin chains. There is progression of the lymphoid signature between 192 and 216h (**Fig. 2D** and **S2G**). Other CD45$^+$ cells have low-frequency expression of HSC-associated transcription factors, including *Myb*, *Hlf* and *Hoxa9* (**Fig. 2G** and **S2I-J**), suggesting incipient specification of AGM-like MPP and pre-HSC, albeit with incomplete programming.

Altogether, single-cell RNA-seq data suggest that hGx specifies YS-like and AGM-like progenitors in a time-dependent manner, with concomitant emergence of hematopoietic-supporting niche cells. The data suggested putative incipient specification of rare candidate MPP or pre-HSC, which we investigated further in comparison with mouse AGM datasets and at a functional level.

We made use of a dataset recently published by one of us (Thambyrajah *et al*., 2024) that interrogates the mouse AGM at the point of HSC emergence from the dorsal aorta HE. In this study, a combination of surface markers and *Gfi1* locus reporting of EHT, tested functionally through transplantation, identified HE and HSC-enriched clusters (Fig. S3A - clusters 1 and 2, respectively, with cluster 4 a putatively heterogenous cluster at the transition between the 2 states). We projected cell transcription profiles from hGx to AGM using a version of scmap algorithm (Kiselev, Yiu and Hemberg, 2018) with 1-cell and 3-cell neighbourhoods for lower and higher levels of confidence on similarity projections (respectively gray and black circles on Fig. S3A - left). The numbers of cluster-specific hGx cells projected onto AGM clusters is quantified in Fig. S3B. Top differentially expressed genes (DEG) between AGM HE and HSC-projected hGx cells are represented in Fig. S3C and Supplemental File S2.

With respect to HE specification, we observed that a subset of hGx C-Kit$^+$ cells present in endothelial clusters 5 and 0 (**Fig. 2A**) projected onto the AGM HE-identity cluster 1 (**Fig. S3A**), with more frequent similarity amongst hGx 192h cluster 0 (41%) than 144h cluster 5 (21%). Overall, this supports the notion that specification of endothelial and HE cells in hGx follows with time-dependent developmental progression into putative AGM-like HE. Lineage identity genes

supported the nature of HE-projected hGx cells (**Fig. S3D**), which expressed *Gata2* but not *Runx1*, *Myb*, or *Gfi1b*, indicating that EHT occurs downstream of these cells. We then asked if any of the hGx cells in hemopoietic clusters bore resemblance to cells from the HSC-enriched AGM cluster 2. We observed that most (85%) hGx CD45^{+} cells in cluster 8 (**Fig. 2A**) projected onto the HSC-enriched cluster 2, which also showed similarity with 76% of CD41^{+}-enriched hGx cluster 4 cells. The relative positioning of hGx cells cluster 4 and cluster 8 progressively further away from the AGM HE cluster was in line with the time-dependent specification in the hGx. However, the lineage marker programs (Fig. **S3D**) expressed by the hGx cells, although definitive in nature (e.g. there is expression of immunoglobulin chains and adult β-globins), lacked detection of HSC genes like *Hlf*, *Gfi1*, or *Hoxa9*. On the other hand, *Mllt3* was present (Calvanese *et al*., 2019), even in the absence of an erythroid signature (Pina *et al*., 2008), suggesting a developing pre-HSC program compatible with the co-expression of multi-lineage markers (Fig. S3D).

Late-stage hGx cells can engraft hematopoietic tissues upon maturation

We inspected the cellularity of hGx at the 216h time-point more closely through a combination of multi-parameter flow cytometry (**Fig. 3A-B**), colony-forming progenitor assays (CFC) (**Fig. 3C-D**) and *in vivo* engraftment of haematopoietic tissues (**Fig. 3E-G**). In order to facilitate the analysis of engrafted animals, we engineered constitutive expression of BFP from the *Rosa26* locus in the *Flk1-GFP* mES cell line (**Fig. S4A**), which we also used in some of the *in vitro* analyses. Phenotypic characterisation of 216h-hGx (4 replicates, 1 with *Rosa26-BFP::Flk1-GFP* line) (**Fig. 3A**, **S4B**) showed the presence of pre-HSC-like CD45^{+}CD41^{lo}c-Kit^{+}VE-Cadherin^{+} cells (1.8% of all CD45^{+}CD41^{lo} cells) (**Fig. 3B**), corresponding to an estimated 8 cells per hGx. Analysis of the hGx progenitor content using multipotential CFC assays (**Fig. 3C-D**) revealed the presence of >50 CFC/hGx, of which an average of 7 have mixed-lineage granulocytic-monocytic-erythroid (GEM) potential, numerically matching the estimated number of pre-HSC. Other progenitors were predominantly granulo-monocytic (GM) and erythroid (E) progenitors (**Fig. 3C-D**). Differentiated cells of G, M and E lineages (**Fig. S4C-D**) were also detectable in 216h hGx, putatively reflecting the progeny of EMP-like cells specified earlier at the 144h timepoint.

As an ultimate test of the presence of pre-HSC, we tested whether late-stage 216h hGx contained cells capable of hematopoietic engraftment of immunodeficient mice. In early experiments, we transplanted dissociated hGx cells into irradiated immunocompetent C57Bl/6 recipients, sub-lethally irradiated immunodeficient NSG mice, or non-conditioned NSGW41 hosts, but failed to detect signs of engraftment of peripheral blood, spleen (Spl) or bone marrow (BM) at 12 days (including absence of spleen colony formation), at 4-6 weeks (short-term engraftment), or beyond 8-10 weeks (long-term engraftment) by either flow cytometry detection of CD45.2 or by genomic DNA (gDNA) PCR analysis of the *Flk1-GFP* locus. However, given the absence of clear HSC-like transcriptional programs in 216h CD45^{+} cells, as analysed by scRNA-seq (**Fig. 2** and **S2**), and in light of the late specification of mesenchymal or sympathetic neuronal cells (**Fig. 2A** and **S2D**) putatively supportive of HSC formation, we hypothesized that hGx pre-HSC required additional maturation in a relevant supportive niche. We thus tested implantation of undissociated hGx in the adrenal gland of *Nude* mice (**Fig. 3E**), a topography commonly used to support tumor development, and recently shown to support adult haematopoietic development (Schyrr *et al*., 2023). Given the coincidence of CD45 isoform between *Flk1-GFP* mouse ES cells and *Nude* mice, we made use of the *Rosa26-BFP::Flk1-GFP* mES cell line (Fig. S3A). We implanted 3 hGx each unilaterally in the adrenal gland of unconditioned recipient mice. Control mice were injected with an equivalent volume of PBS. Animals were sacrificed at 24 hours, 12 days, and 4 weeks after implantation, with 5 experimental animals and 1 control collected at each timepoint. We checked Spl and BM BFP engraftment at the different timepoints and analyzed the morphology of adrenal glands receiving implants (**Fig. 3F**). We did not observe BFP engraftment, spleen enlargement or colony development, or macroscopic changes in the adrenal gland at the earlier timepoints. However, 3 of 5 adrenal glands were enlarged, and their histology suggested *in situ*

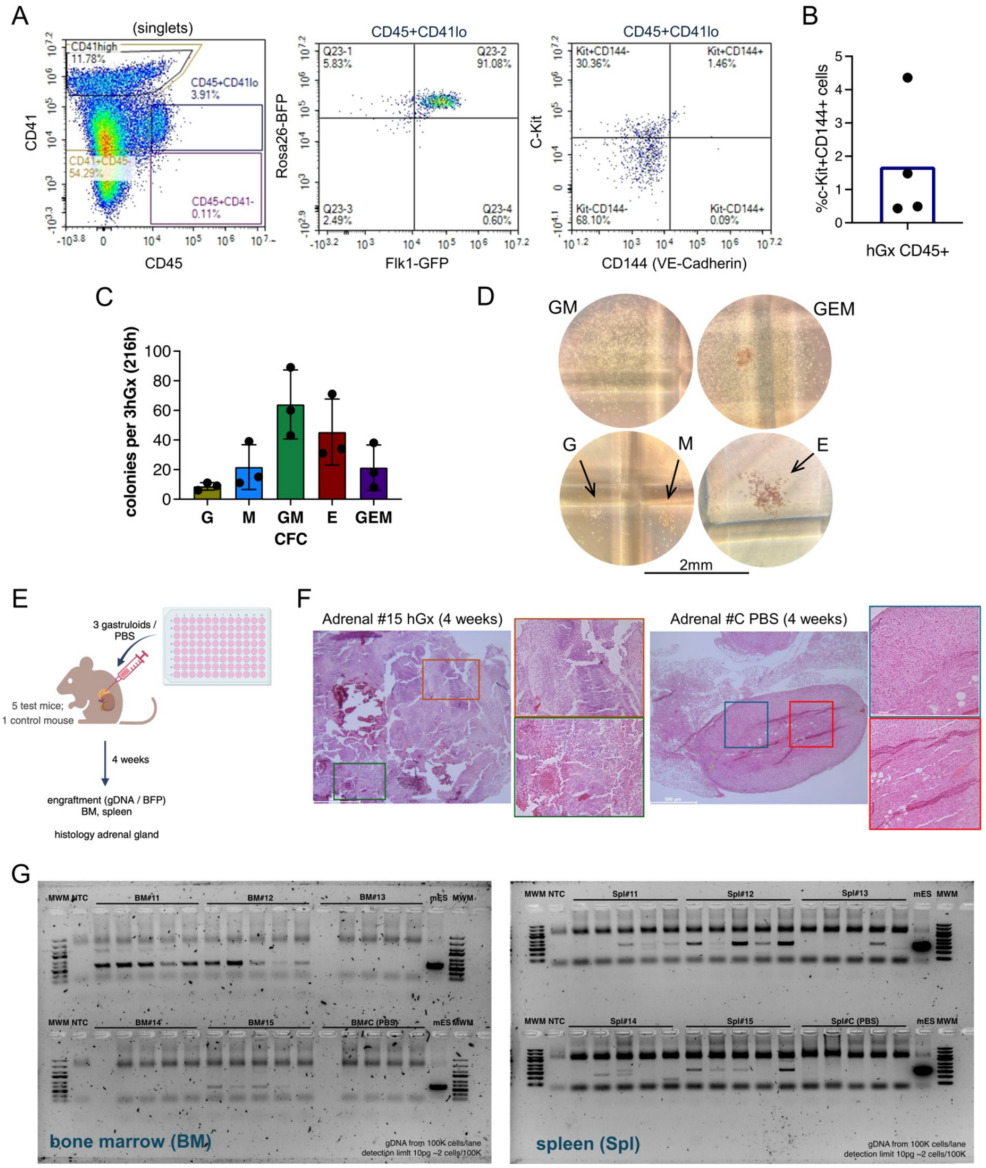


Figure 3

HGxs specify cells with pre-HSC characteristics.

(A) Flow cytometry analysis of 216h *Rosa26-BFP::Flk1-GFP* hGx cells for CD41, CD45, c-Kit and CD144 (VE-Cadherin). **(B)** Quantification of pre-HSC-like c-Kit^{+}CD144^{+} cells within the CD45^{+}CD41^{lo} fraction of 216h-hGx; mean of 4 independent replicate experiments. **(C)** Colony-forming cell (CFC) assay of 216h-hGx in multipotential methylcellulose-based medium. GEM: granulocyte-erythroid-monocyte; GM: granulocyte-monocyte; M: monocyte; E: erythroid. Please note that the medium does not include TPO and does not assess the presence of megakaryocytic progenitors. CFC frequency in 3hGx, n=3, mean±SD. **(D)** Representative photographs of colonies quantified in C; scale bar=2mm. **(E)** Schematic representation of the experimental workflow for execution and analysis of hGx implantation in the adrenal gland of immunodeficient mice. **(E)** Histology of implanted and control adrenal glands; paraffin sections stained with hematoxylin & eosin (H&E). Inserts in implanted animal highlight neural (orange) and hematopoietic and renal tubular epithelium (green); inserts in control animal highlight normal cortical (blue) and medullar (red) adrenal architecture, absent in implanted adrenal glands. **(F)** PCR detection of hGx genomic (g)DNA in the bone marrow (BM) and spleen (Spl) of immunodeficient mice 1 month 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 hGxs. NTC: no template control. See Fig. S4F-H for additional details.

development of the hGx (**Fig. 3F**), particularly mesodermal tissues – somite derivatives (**Fig. 3F** - left, red box), and lateral plate and intermediate mesoderm derivatives, respectively blood cells and renal tubules (**Fig. 3F** - left, green box). The control, PBS-injected glands showed normal tissue architecture (**Fig. 3F** - right). It should be noted that none of the specimens contained glomeruli in addition to the renal tubules, pointing to an origin from the hGx, and not from the adjacent kidney tissue. BFP engraftment of the Spl and BM by flow cytometry was very low level albeit consistently above control (**Fig. S4E**), and we opted for analysing engraftment by PCR analysis of gDNA using primers directed against the BFP insert (**Fig. S4A**). We measured the sensitivity of the amplification through serial dilution of *Flk1-GFP Rosa26-BFP* mES cell gDNA into that of human K562 cells, and determined detection of 10pg/100ng gDNA, equivalent to 2-3 cells (**Fig. S4F**). PCR analysis of spleen and BM gDNA detected *BFP* amplicon in both tissues in 3 of the 5 samples (#11, 12 and 15) (**Fig. 3G**). We sampled the gDNA of each recipient 4-5 times and detected the presence of the amplicon in only a fraction of the replicate samples of the positive animals, reflecting the low-level detection by flow cytometry. Replicate sampling of the control was consistently negative (**Fig. 3G**). Sorting of erythroid Ter119^{+} and lympho-myeloid CD45^{+} cells (**Fig. S4G**) showed PCR engraftment in both lineage fractions (**Fig. S4H**).

Altogether, the data suggest that 216h hGx generate AGM-like pre-HSC capable of at least short-term multilineage engraftment upon maturation. The low-level production of engrafting cells recapitulates their rarity *in vivo*, in agreement with the embryo-like qualities of the gastruloid system, which accommodates modular qualitative bias, but not quantitative changes in tissue specification.

The ability of hGx to reconstitute YS and AGM-like HE and hemopoietic progenitors with temporal coherence lends itself as an attractive model to understand developmental leukemia. In particular, forms of AML unique to the first 2 years of life – infant (inf)AML – have been shown to transform fetal liver (FL) but not adult haematopoietic cells (Waraky *et al*., 2024; Mercher *et al*., 2009; Chen, W. *et al*., 2011; Ragusa *et al*., 2023), compatible with targeting of a transient embryonic cell. In the case of t(7;12) AML (herein MNX1-*r*) (Ragusa *et al*., 2023), which results in a transformation-driving ectopic expression of *MNX1* (Waraky *et al*., 2024), analysis of patient transcriptional signatures (Ragusa *et al*., 2022) for over-representation of cell atlas-defined affiliations, indicates enrichment in cardiomyogenic, endothelial, HSC/progenitor and mast cells (**Fig. 4A**). These cell enrichments are compatible with an early hemogenic cell derivative, capturable in the hGx model. Other cell type-enrichments reflect MNX1 functions in pancreatic and neuronal development (Ragusa *et al*., 2022) (Fig. S5A).

We overexpressed *MNX1* (MNX1-OE) in *Flk1-GFP* mouse ES cells by lentiviral transduction (**Fig. 4B**) and analyzed progression of hGx generation (**Fig. 4C**). We used human *MNX1* cDNA to distinguish from the endogenous gene, but the degree of homology is nevertheless high (84%), supporting functional equivalence. We confirmed that *MNX1* overexpression in hGx is maintained at endpoint (**Fig S5B**). MNX1-OE hGx activated polarized *Flk1-GFP* expression and elongated with similar kinetics to empty-vector (EV) control (**Fig. 4C**), but consistently produced larger hGx (**Fig. 4D**) denoting increased cellularity (**Fig. S5C**). From 192h onwards, MNX1-OE hGx had a higher frequency of spontaneously contractile structures (**Fig. S5D**), compatible with the cardiogenic cell association of patients' transcriptomes (**Fig. 4A**). We interrogated the time-dependent hemogenic cell composition of hGx by flow cytometry, quantifying C-Kit, CD41 and CD45 (**Fig. 4E-G** and **S5E-F**). We observed a significant expansion of the C-Kit^{+} compartment specifically at 144h (**Fig. 4E** and **S5E**). All cell markers were relatively unchanged at later timepoints, although absolute cell numbers were higher in MNX1-OE hGx (**Fig. S5C**).

To better understand the consequences of MNX1-OE on hemogenic development, we performed RNA-sequencing (RNA-seq) of hGx at 144h and 216h, matching YS and AGM-like hemopoietic waves, respectively (**Fig. 4H**). We confirmed expression of the human *MNX1* transgene in the sequenced reads (**Fig. S6A**); in contrast, we could not detect endogenous mouse *Mnx1* in any of

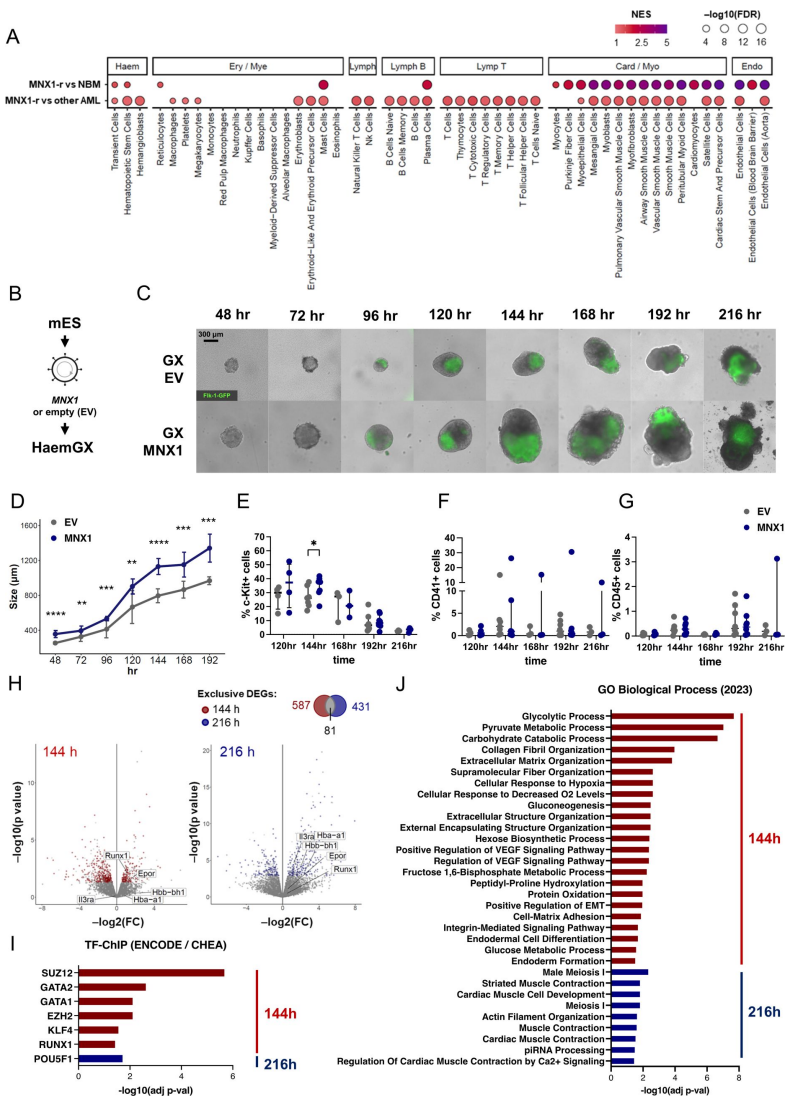


Figure 4

HGxs with MNX1 overexpression have increased cellularity and enhanced hemogenic endothelial potential.

(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)** Schematic representation of generation of *MNX1*-overexpressing and empty vector (EV) mES cells by lentiviral transduction, with subsequent assembly into hGxs (haemGX). **(C)** Imaging of hGxs with *MNX1* overexpression and EV controls at 10x magnification, showing appropriate assembly and polarization of the Flk1-GFP marker. scale bar: 300mm. **(D)** Size of hGx at each timepoint, determined by the distance between the furthest extreme points in μm. Mean ± SD of 3 replicate experiments; 2-tailed t-test, p< 0.05 (*), 0.001 (**), 0.0001 (***), and 0.00001 (****). **(E-G)** Flow cytometry timecourse analysis of **(E)** C-Kit$^+$, **(F)** CD41$^+$, and **(G)** CD45$^+$ cell abundance in MNX1 and EV hGxs (120-216h). Mean ± standard deviation of 3-7 independent experiments; 2-way ANOVA and Sidak's multiple comparison test significant at p<0.05 for C-Kit$^+$ cells only (construct contribution to variance p=0.0191; 144h comparison *p=0.0190). **(H)** Volcano plots of differentially expressed genes (DEGs) by bulk RNA-seq between MNX1 and EV hGxs at 144h and 216h. 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 hGx 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 hGx at 144h (red) and 216 hr (blue) using the ENCODE and ChEA Consensus TFs from ChIP database on EnrichR.

the samples (**Fig. S6A**), confirming that it does not normally play a role in hemogenic development. DEGs between MNX1-OE and EV (**Supplemental File S3**) configured distinct MNX1-driven programs at 144h and 216h, with >80% genes timepoint-specific (**Fig. 4H**). Both programs include up-regulation of hematopoietic-associated genes, capturing EHT and progenitor regulators (e.g. *Runx1*, and *Epor* and *Zfpm1*, respectively) at the earlier timepoint, and more differentiated effectors, e.g. *Hba-a1*, *Hbb-bh1* and *Il3ra* at 216h, compatible with selective targeting and/or expansion of newly specified progenitors at the YS-like stage (**Fig. 4H**). An overview of transcriptional regulatory programs at both timepoints denotes a clearer hemogenic enrichment at 144h, with over-representation of GATA2, GATA1 and RUNX1 binding targets (**Fig. 4I**), putatively capturing an expansion of the HE-to-EMP transition, also represented in enriched GO categories related to VEGF signalling and epithelial-to-mesenchymal transition (EMT) (**Fig. 4J**). Compatible with the increase in contractile foci in MNX1-OE late hGx, there was an enrichment in cardiogenic gene ontologies (GO) at the 216h timepoint (**Fig. 4J**). No hemogenic transcription factors (TF)-ChIP or GO categories were enriched in MNX1-OE hGx at 216h (**Fig. 4J-I**).

We attempted to deconvolute the cellular heterogeneity underlying bulk RNA-seq DEGs by interrogating the cluster signatures obtained from the single-cell analysis of hGx development (**Fig. S6B-C**; refer to **Fig. 2A**). In support of enhanced YS-stage hemato-endothelial specification, MNX1-OE transcriptomes showed enrichment of hGx HE cluster 5 and 0 signatures, and of the EMP-like cluster 4 (**Fig. S6B**). MNX1-OE DEGs at both timepoints also enriched for the relatively less differentiated cluster 10, which predominantly captures unfractionated hGx cells at 120h (**Fig. S6B-C**; refer to **Fig. 2A**). Significantly, clusters 5 and 0 HE signatures were enriched in *MNX1*-r infAML patient signatures (Ragusa et al., 2022) (Fig. S5B-C), positioning MNX1-OE at early stages of hemogenic specification. MNX1-OE DEGs at 216h show enrichment in late-stage clusters 7 and 8 (Fig. S6B-C), denoting enhancement of hGx development, including MLP/MPP in cluster 8, but diverging from *MNX1*-r infAML signatures. Projection of infAML subtype-specific gene signatures onto hGx development separates the lineage affiliations of *MNX1*-r from, for example, the most frequent infAML driver – *KMT2A (MLL)* rearrangements (Fig. S6B-C), which carries myelo-monocytic and/or mixed-lineage affiliation, highlighting the utility of the hGx system for interrogating infAML biology.

We further interrogated the *MNX1*-r-like leukemia-initiating potential of MNX1-OE hGx through serial replating of CFC assays, a classical *in vitro* assay of leukemia transformation. We compared 144h and 216h hGx to match functional and transcriptome data, and plated unfractionated hGx, given the transient nature of C-Kit$^+$ enrichment. Both 144h (**Fig. 5A-B**) and 216h (**Fig. 5C-D**) MNX1-OE hGx replated for at least 5 plates, with significant differences to control (EV) (**Fig. 5B,D**). EV-hGx replating could not be sustained beyond plate 3 in most experiments started from 216h (**Fig. 5C**); 144h EV-hGx exhibited some low-level replating until plate 5 (**Fig. 5A**), but the colonies were formed by few scattered cells clearly distinct from the well-demarcated colonies obtained from MNX1-OE hGx. We analyzed early-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. 5E**), matching the cell phenotype transiently enriched at 144h. Early replating of MNX1-OE hGx also enriched for erythroid-affiliated Ter119$^+$ cells particularly from the 144h time-point (**Fig. 5F**), reflecting the differential erythroid potential of EMP and MLP/MPP at 144h and 216h, respectively. This enrichment is not sustained upon replating, compatible with the undifferentiated nature of *MNX1*-r infAML, which is also reflected by the low levels of CD11b, a marker affiliated with the myelo-monocytic lineages (**Fig. 5E**). Notably, the cells obtained are not CD45$^+$ (**Fig. 5F**), suggesting perpetuation of an early hemogenic state.

We performed RNA-seq analysis of MNX1-OE hGx CFC cells obtained through re-plating and integrated the data with the 144h and 216h hGx time-points to identify genes enriched through the process of transformation. We observed relative enrichment in Hs*MNX1* reads in MNX1-OE CFC samples, supporting selective expansion of transduced cells (**Fig. 6A**). Hierarchical clustering of all DEGs in pairwise comparisons (**Fig. 6B**), identified a subgroup of genes up-regulated upon

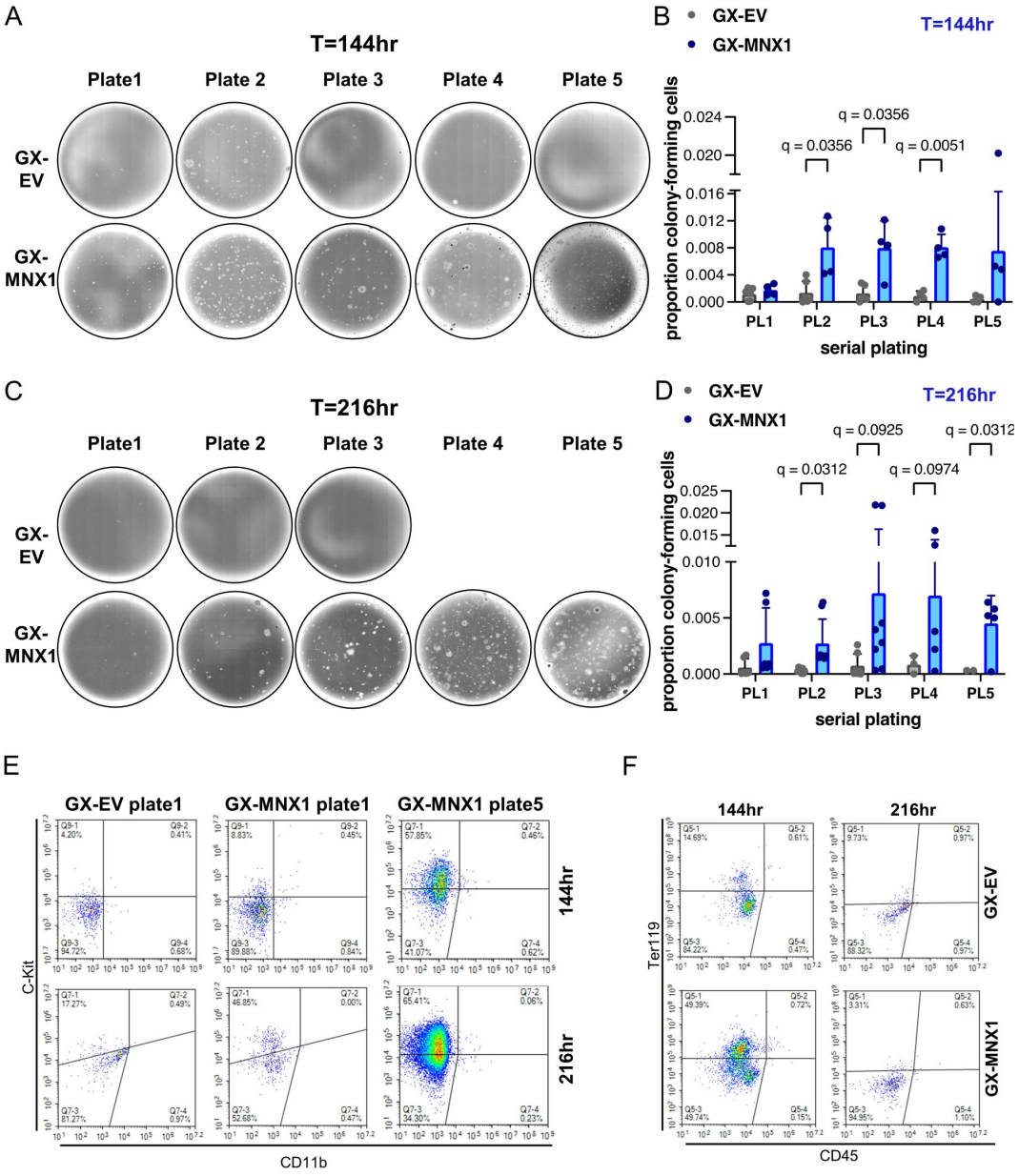


Figure 5

MNX1 selects C-Kit$^+$ clonogenic cells and selectively transforms end-stage hGxs.

(A) Representative photographs of serial replating of colony-forming cell assays initiated from EV or MNX1 cells obtained at 144h of the hGx protocol. **(B)** Quantification of colony-replating efficiency of EV and MNX1 144h-hGx cells. Mean+SD of n>3 replicates; Kruskal-Wallis with Dunn's multiple comparison testing at significant q<0.05. **(C)** Representative photographs of serial replating of colony-forming cell assays initiated from EV or MNX1 cells obtained at 216h of the hGx protocol. **(D)** Quantification of colony-replating efficiency of EV and MNX1 216h-hGx cells. Mean+SD of n=5-8 replicates; Kruskal-Wallis with Dunn's multiple comparison testing at significant q<0.05. **(E)** Flow cytometry analysis of hemogenic progenitor C-Kit$^+$ cells and myeloid-affiliated CD11b$^+$ cells at first round of plating of 144h and 216h-hGxs initiated from EV and MNX1 cells; representative plots. **(F)** Representative flow cytometry plots of serially re-plating MNX1 cells from 144h and 216h-gastruloids (plate 5 cells).

CFC replating which was also up-regulated at MNX1-OE hGx at 144h (K14), thus matching the putative time-dependent enrichment of the MNX1-OE propagating population. To understand the relevance of these genes in *MNX1*-r leukemia, we performed Gene Set Enrichment Analysis (GSEA) of *MNX1*-r patients in comparison to other AML in the same infant (0-2) age group. K14-genes were enriched in MNX1-*r* all comparisons (**Fig. 6C**), with considerable overlap between leading-edge (i.e. significantly-enriched) genes (**Fig. S6D**). Interrogation of cell-type-specific signatures within the K14 MNX1-*r* leading edge genes (**Fig. 6D**) against the single-cell PanglaoDB atlas captured similarities to hemogenic and non-hemogenic cell types, including interneuron, and other neural, signatures, which reflect MNX1 biological functions (Harrison *et al*., 1999; Thaler *et al*., 1999; Ragusa *et al*., 2022). The most highly significant hemogenic signatures capture early developmental components, such as 'endothelial cells', which were specifically up-regulated within MNX1-OE hGx (Fig. S6B). 'Erythroid cell precursors' also reflect MNX1-OE hGx effects (**Fig. 4E-J**). We were intrigued by the 'mast cells' affiliation of *MNX1*-r AML signatures within the hGx K14 cluster, which was more specifically aligned with MNX1-OE hGx isolated at 144h (**Fig. S6E**). Given the characteristic morphology of these cells, we analysed Giemsa-Wright-stained cytospins of CFC replating assays (**Fig. 6E** and **S6F**). Consecutive plates of 144h MNX1-OE-initiated *in vitro* transformation showed a shift from blast-like at plate 3 to precursor cell morphology with some differentiated mast cells at plate 5 (**Fig. S6F**). However, the extent of differentiation was greater, and apparent earlier, in 216h-MNX1-OE initiated colonies (**Fig. 6E**). This suggests that the MNX1-OE-responsive hemogenic C-Kit$^+$ progenitor may experience maturation between 144h and 216h, and is more likely targeted for transformation at the YS EMP stage. The data support the utility of the hGx in dissecting ontogeny of embryonic origin in hematological cancers, allowing us to specifically narrow down the initiating cellular targets of MNX1-*r* AML.

Discussion

In this study, we developed and characterized a hGx model from mouse ES cells, which successfully recapitulates YS-like and AGM-like waves of definitive blood progenitor specification, and can engraft adult hematopoietic tissues upon *in vivo* maturation. We took advantage of the hGx's nearly full recapitulation of embryonic hemogenic specification, including critical early stages of HE production and EHT to EMP and MLP/MPPs, to interrogate the biology of *MNX1*-r AML, a poorly characterized form of leukemia exclusive to the infant age group (Ragusa *et al*., 2023). We showed that MNX1-OE transiently expanded a C-Kit$^+$ cell with hemato-endothelial characteristics at the YS-like stage, and could selectively sustain the phenotype through serial replating of CFC assays. Replating CFCs could be initiated beyond the YS-like stage, albeit at the relative expense of an early differentiation state and of the similarity to patient transcriptome, putatively defining a window of susceptibility to MNX1-driven transformation. The hGx MNX1-OE propagating cell bears transcriptional resemblance to patient *MNX1*-r AML, overall positioning the hGx as an attractive tractable model in which to study developmental leukemias.

MNX1-r has been difficult to characterize due to a combination of rarity, diagnostic difficulty (the underlying t(7;12) can be missed if not specifically looked for) (Tosi *et al*., 2015), and the lack of a fusion protein that can be ectopically expressed and tracked through development, as is the case for other translocations unique to infAML (Alexander and Mullighan, 2021). However, a recent paper using MNX1-OE has shown that MNX1 can initiate a transplantable leukemia in immunodeficient mice by targeting FL, but not BM, cells (Waraky *et al*., 2024). This suggests that ectopic expression of *MNX1* transforms an embryonic cell, rather than depends on an embryonic niche. It also suggests that MNX1-OE can cause leukemia as a single genetic hit, although it cannot be excluded that other mutations are required for, and may indeed have contributed to, full transformation *in vivo*, as observed in a recent report confirming the *in utero* origin of the leukemia (Bousquets-Muñoz *et al*., 2024). However, the mouse model does not address the nature of the embryonic cell targeted by MNX1-OE. Like the MNX1-OE hGx cells selected through

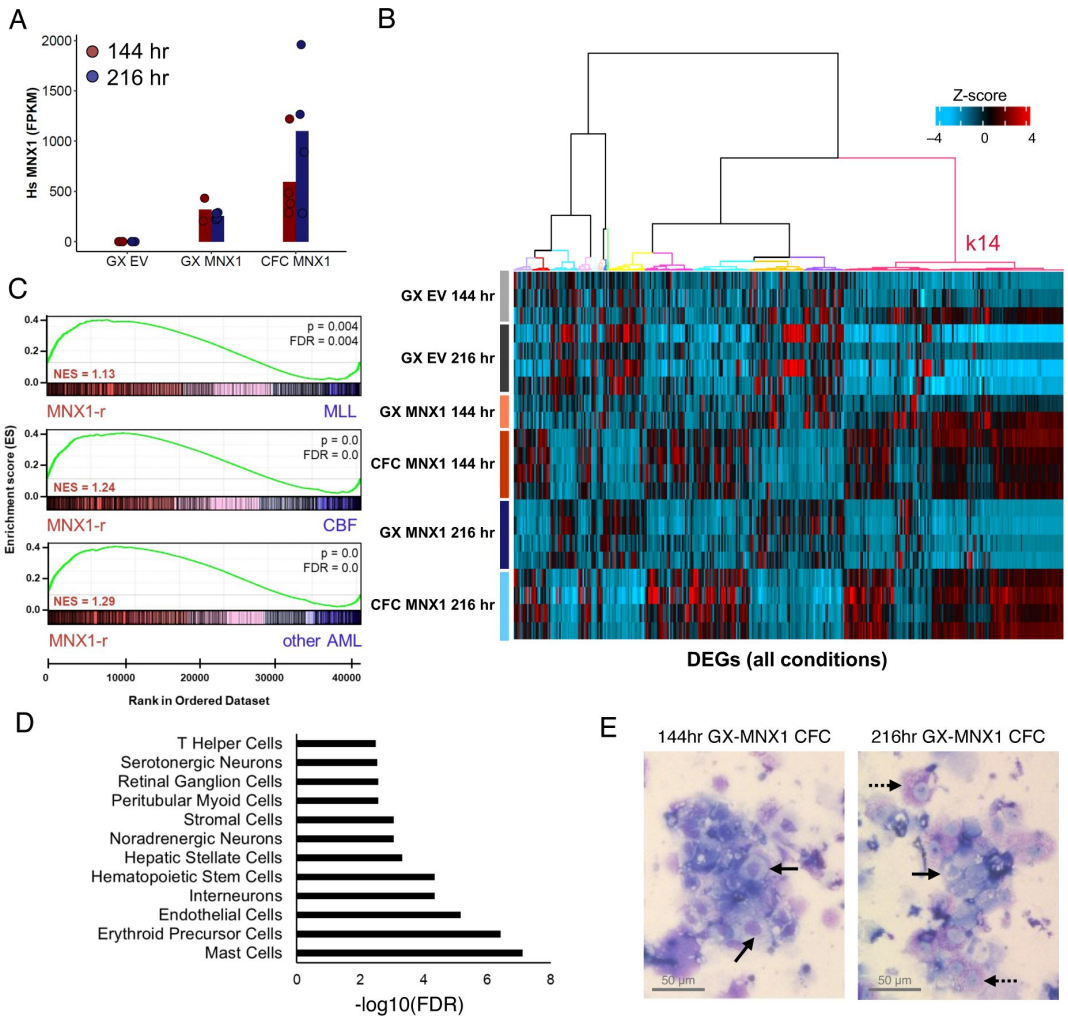


Figure 6

HGxs with *MNX1* overexpression recapitulate MNX1-Acute Myeloid Leukemia patient signatures.

(A) Quantification of human *MNX1* transcripts in FPKM units from RNA-seq of GX-EV, GX-MNX1, and CFC at 144 hr and 216 hr. Bars represent mean of replicates. **(B)** Heatmap comparing the expression of all differentially expressed genes (DEGs) between all conditions (GX-MNX1 144 hr vs GX-EV 144 hr; GX-MNX1 216 hr vs GX-EV 216 hr; CFC MNX1 144 hr vs GX MNX1 144 hr; CFC MNX1 216 hr vs GX MNX1 216 hr) as Z score. Hierarchical clustering by Ward D method on Euclidean distances identifies 14 clusters (k). **(C)** GSEA plots for gene set extracted from k14 from **(B)** against MNX1-r patients RNA-seq counts vs MLL, core-binding factors (CBF), or other pediatric AML. **(D)** Cell type analysis of the intersect of leading edge genes LEGs (n=476) from Fig. S5D in MNX1-r patients (k14 LEGs) **(B)** using the Panglao DB 2021 database. **(E)** Representative Giemsa-Wright stained dissociated CFC replating MNX1 hGx cells from 144h and 216h. Solid black arrows, mast cell precursors; dashed black arrows, mast cells.

serial replating, MNX1-OE mouse leukemia are also C-Kit$^+$ (Waraky *et al*., 2024). Interestingly, clusters of genes up-regulated in the mouse MNX1-OE leukemia (Fig. S7A – K1, 2, 5, 6, and 9) do not robustly separate *MNX1*-r from other infAML by GSEA analysis (Fig. S7B), and have diverged more from the patient data in their cell-type signature analysis than the original MNX1-OE cells in the FL (Fig. S7C), putatively selecting a more generic program of leukemia transformation. We propose that the divergence from *MNX1*-r infAML signatures increases from MNX1-OE hGx 144h (YS-like), to 216h (AGM-like) 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 hGx culture system to the 216h AGM-like time-point, and migrate to the FL in the embryo, eventually extinguishing through differentiation, as suggested by the MNX1-OE hGx CFC assays. Persistence of the original MNX1-OE target cell will determine the sensitivity to MNX1-OE, and can be addressed in the hGx 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 hGx system, in combination with transplantation experiments. Finally, although our study has positioned the MNX1-OE target cell within the YS-EMP stage, the exact nature of the cell requires better definition. The mast cell affiliation, transcriptional and by CFC morphology, is more likely to reflect the period of susceptibility to MNX1 and the multilineage potential of the target cell, than the lineage commitment of the latter. Indeed, blast cells in the leukemia mouse model are morphologically undifferentiated (Waraky *et al*., 2024) and a mast cell morphology has not been described in patients (Ragusa *et al*., 2023; Espersen *et al*., 2018). Although adult HSC have mast cell potential *in vitro*, this is not employed *in vivo* under physiological conditions (Chia *et al*., 2023). Tissue resident mast cells are formed in the embryo at 2 periods: the YS EMP stage, which we propose to have targeted with MNX1-OE in the hGx model; and a transient AGM FL-HSC population which does not persist in post-natal life (Chia *et al*., 2023; Yoshimoto *et al*., 2022), and that may be incipiently specified at the end of the current hGx protocol. Prospective isolation of these cell types from embryos and over-expression of *MNX1* followed by transformation *in vitro* and *in vivo*, will precisely define the embryonic time-window of susceptibility to *MNX1*-r and the nature the leukemia-initiating cell. Harnessing the high-throughput nature of the hGx culture system will help identifying therapeutic vulnerabilities that can improve the prognosis of this often-fatal disease.

The hGx model furthermore constitutes an advanced platform to probe physiological development of the hematopoietic system. Currently, *in vitro* platforms do not recapitulate *in vivo* blood formation sufficiently faithfully (Dijkhuis *et al*., 2023; Fidanza and Forrester, 2021). 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 hGx 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. Although the system is shy of achieving HSC maturation *in situ*, it can mature in a supportive microenvironment, in this case provided by the adrenal gland (Schyrr *et al*., 2023), to produce engrafting multilineage progenitors, and potentially HSCs. This suggests that access to the relevant

hematopoietic molecular programs is preserved. Incipient specification of non-hematopoietic tissues shown to support HSC emergence, including PDGFRA$^+$ mesenchymal (Chandrakanthan *et al*., 2022) and neuro-mesenchymal (Miladinovic *et al*., 2024) cells, and neural-crest derivatives (Kapeni *et al*., 2022; Fitch *et al*., 2012), occurs in sync with development of AGM-like hematopoiesis, to indicate the presence of cues that with further maturation may facilitate HSPC specification at the appropriate developmental time.

We have observed that the nature of the symmetry-breaking pulse affects the specification of hemato-endothelial programs, with a more robust *Flk1*/*Kdr* activation in the presence of Activin A, potentially balancing mesodermal derivatives and the anterior-posterior axis. The hematopoietic output of cardiogenic gastruloids (Rossi *et al*., 2022) does not seem to require Activin A, but their blood formation is of lower efficiency, at least at the comparable timepoints of 144 and 168h, and apparently biased towards erythroid formation, reflecting its YS, and potentially primitive, nature. In the absence of additional cytokines after the CHI99021 pulse, standard gastruloid protocols (Beccari *et al*., 2018) capture a similar transcriptional erythrogenic wave. Although Rossi *et al*. (2022) reported the formation of spleen colonies, these are difficult to ascertain from the data published, and the engraftment capacity of cardiogenic gastruloid-derived blood remains uncertain. Our hGx system has more robust CD41$^+$ cell production at 144-168h, which may reflect distinct tissue induction by the CHI99021+Activin A pulse. Interestingly, through extending the hGx 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. An additional challenge is tissue organisation: non-hematopoietic gastruloid models benefitted from mechanical, or chemo-mechanical, support of artificial ECM embedding (Veenvliet *et al*., 2020; van den Brink, Susanne C *et al*., 2020), 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 by circulatory shear stress (North *et al*., 2009; Diaz *et al*., 2015; Lundin *et al*., 2020), suggesting that introduction of flow may improve hemato-endothelial organization and boost blood production.

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 hGx culture. In conclusion, the hGx represents a faithful *in vitro* reference to replace and complement embryo studies, holding promise to understand the biology of healthy and malignant developmental haematopoiesis and advance blood production for clinical use.

Experimental procedures

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
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

Methods

Cell culture

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

Hemogenic gastruloids assembly, culture, and dissociation

mES were maintained in ES+LIF medium and transferred to 2i+LIF (containing Chiron and MEK inhibitor PD03) for 24 hours prior to the assembly into gastruloids. 250 untransduced mES cells (or 400 for GX-MNX1 and GX-EV) were seeded in each well of a U-bottom, cell-repellent 96-well plate (Greiner Bio-One, Stonehouse, UK) in 40 µl of N2B27 medium [Takara Bio or home-made as Cold Spring Harbour Protocols 'N2B27 Medium' (2017)]. The plate was centrifuged at 750 rpm for 2 minutes to promote deposition and aggregation of the cells and was then incubated at 37°C, 5% CO2 for 48 hours. After 48 hours, 150 µl of N2B27 medium supplemented with 100 ng/ml Activin A (QKine, Cambridge, UK) and 3 µM chiron (Peprotech) was added to each well. At 72 hours, 150 µl of medium were removed, without disrupting the gastruloids in the wells. 100 µl of N2B27 with 5 ng/ml of Vegf and Fgf2 each (Peprotech) were added to each well. From 72 h to 144 h, each day 100 µl of medium were removed and replaced with N2B27 + Vegf + Fgf2. At 144 h, the medium was 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 hGx (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 hGx, 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 hGx 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 described above. Dissociated gastruloid cells from CFC were collected as cell suspensions in PBS and centrifuged onto slides (Shandon Cytospin 2 Cytocentrifuge) at 350rpm for 5-7min, with a PBS-only pre-spin at 1000rpm, 2min. Slides were stained in Giemsa-Wright stain for 30 seconds followed by addition of phosphate buffer pH 6.5 for 5 min.

Immunofluorescence staining

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

Histology and H&E staining

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

Imaging

Images of cultured gastruloids and CFC plates were captured using the Cytation 5 Cell Imaging Multi-Mode Reader (Biotek) plate reader using bright field and FITC channels. ImageJ (Schneider et al., 2012) was used for gastruloid size quantification. Confocal microscopy was performed on LSM700 on a Zeiss Axiovert 200 M with Zeiss EC Plan-Neofluar 10x/0.30 M27 and Zeiss LD Plan-Neofluar 20x/0.4 M27 objective lens. Tissue sections were imaged with a THUNDER Imager 3D microscope system (Leica).

Lentiviral vector packaging and transduction

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

Flow cytometry

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

Single cell RNA sequencing of time-resolved hemogenic gastruloids

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

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

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

Annotation and projection to additional datasets

Raw counts matrices from Thambyrajah *et al*. (2024) were downloaded and processed following the same pipeline as described above for scRNA-seq of gastruloids to generate UMAP and clustering. We implemented the scmap algorithm to compare our scRNA sequencing of gastruloids with the available datasets. We reduced the dimensionality of the space by selecting highly varying genes from the annotated dataset. Then, we constructed a KNN classifier with correlation metric and computed the nearest neighbors of the target data. If neighbors with correlation metrics below 0.7 default standards, the projected cells were not projected onto any cell from the annotated dataset. To visualize the cells over the UMAP plots of the other datasets, we constructed a KNN regressor with three neighbors and a correlation metric. Confusion matrices were used to visualize the overlap between gastruloid clusters and the embryo dataset assigned clusters using scmap.

Bulk RNA sequencing

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

Gene list enrichment analyses

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

Gene Set Enrichment Analysis (GSEA)

Custom gene signatures were used as gene sets for GSEA analysis (Subramanian *et al*., 2005) on the GSEA software v4.2.3 on RNA sequencing expression values in counts units. GSEA was ran in 10000 permutations on gene set using the weighted Signal2Noise metric. Enrichment metrics are shown as normalized enrichment score (NES) and filtered by FDR ≤ 0.05. Leading edge genes (LEGs) are genes with a “Yes” values for core enrichment. For AML patient analysis, clinical

phenotype and expression data (in counts units) were extracted from the GDC TARGET-AML cohorts in the Therapeutically Applicable Research to Generate Effective Treatments project (TARGET, *https://ocg.cancer.gov/programs/target*), downloaded from the University of California Santa Cruz (UCSC) Xena public repository (last accessed 31st August 2022). Patient samples were selected according to the reported karyotype to include t(7;12), inv(16), MLL, normal karyotype, and t(8;21). GSEA was performed comparing RNA sequencing counts of t(7;12) samples against pooled AML subtypes (inv(16), MLL, normal karyotype and t(8;21)) as "other AML".

Real-time polymerase chain reaction (qPCR)

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

Polymerase chain reaction (PCR)

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

Statistical analysis

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

Data availability

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

Acknowledgements

This project was funded by a start-up grant and a BRIEF award from Brunel University London to CP, by a Little Princess Trust (LPT) Project Grant (CCLGA 2023 22 Pina) to CP, 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.

Additional information

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.

Additional files

Supplemental Figures

Supplemental File S1

Supplemental File S2

Supplemental File S3

References

Alexander T.B., Mullighan C.G. 2021) **Molecular biology of childhood leukemia** *Annual Review of Cancer Biology* **5**:95–117

Ballabio E., Cantarella C.D., Federico C., Di Mare P., Hall G., Harbott J., Hughes J., Saccone S., Tosi S. 2009) **Ectopic expression of the HLXB9 gene is associated with an altered nuclear position in t (7; 12) leukaemias** *Leukemia* **23**:1179–1183

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

Boisset J., van Cappellen W., Andrieu-Soler C., Galjart N., Dzierzak E., Robin C. 2010) **In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium** *Nature* **464**: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., Marra M.A 2018) **The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions** *Nature medicine* **24**:103

Bousquets-Muñoz P., Molina O., Varela I., Álvarez-Eguiluz Á, Fernández-Mateos J., Gómez A., Sánchez E.G., Balbín M., Ruano D., Ramírez-Orellana M 2024) **Backtracking NOM1:: ETV6 fusion to neonatal pathogenesis of t (7; 12)(q36; p13) infant AML** *Leukemia* :1–5

Bowie M.B., McKnight K.D., Kent D.G., McCaffrey L., Hoodless P.A., 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**:2808–2816

Brignole C., Calarco E., Bensa V., Giusto E., Perri P., Ciampi E., Corrias M.V., Astigiano S., Cilli M., 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., Crooks G.M 2019) **MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment** *Nature* **576**:281–286

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

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

Chang Y., Hummel S.N., Jung J., Jin G., Deng Q., 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., Gibson B 2018) **Age-specific biological and molecular profiling distinguishes paediatric from adult acute myeloid leukaemias** *Nature communications* **9**:5280

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V., 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., Kersey J. 2011) **'Modeling human infant MLL leukemia in mice: leukemia from fetal liver differs from that originating in postnatal marrow'** *Blood* **117**:3474

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

Ciau-Uitz A., Pinheiro P., Gupta R., Enver T., Patient R 2010) **Tel1/ETV6 specifies blood stem cells through the agency of VEGF signaling** *Developmental cell* **18**:569–578

Ciriza J., Thompson H., Petrosian R., Manilay J.O., 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**:411–423

Costa G., Kouskoff V., Lacaud G 2012) **Origin of blood cells and HSC production in the embryo** *Trends in immunology* **33**:215–223

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

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

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

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

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

Dzierzak E., Bigas A 2018) **Blood development: hematopoietic stem cell dependence and independence** *Cell stem cell* **22**:639–651

Espersen A.D.L., Abrahamsson J., Ha S., Pronk C.J., Jahnukainen K., Jónsson ÓG., Lausen B., Palle J., 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**:359–365

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

Fidanza A., 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**:100050

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

Fowler J.L., Zheng S.L., Nguyen A., Chen A., Xiong X., Chai T., Chen J.Y., Karigane D., Banuelos A.M., 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**:1110–1131

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

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

Fujimoto T., Ogawa M., Minegishi N., Yoshida H., Yokomizo T., Yamamoto M., Nishikawa S 2001) **Step-wise divergence of primitive and definitive haematopoietic and endothelial cell lineages during embryonic stem cell differentiation** *Genes to Cells* **6**:1113–1127

Ganuza M., Hall T., Myers J., Nevitt C., Sánchez-Lanzas R., Chabot A., Ding J., Kooienga E., Caprio C., Finkelstein D 2022) **Murine foetal liver supports limited detectable expansion of life-long haematopoietic progenitors** *Nature cell biology* **24**:1475–1486

Ghiaur G., Ferkowicz M.J., Milsom M.D., Bailey J., Witte D., Cancelas J.A., Yoder M.C., 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**:3313–3321

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

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

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

Jakobsson L., Franco C.A., Bentley K., Collins R.T., Ponsioen B., Aspalter I.M., Rosewell I., Busse M., Thurston G., Medvinsky A 2010) **Endothelial cells dynamically compete for the tip cell position during angiogenic sprouting** *Nature cell biology* **12**:943–953

Kapeni C., Nitsche L., Kilpatrick A.M., Wilson N.K., Xia K., Mirshekar-Syahkal B., Chandrakanthan V., Malouf C., Pimanda J.E., Göttgens B 2022) **'p57Kip2 regulates embryonic blood stem cells**

by controlling sympathoadrenal progenitor expansion *Blood, The Journal of the American Society of Hematology* **140**:464–477

Kikuchi K., 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**:17852–17857

Kiselev V.Y., Yiu A., Hemberg M 2018) **'scmap: projection of single-cell RNA-seq data across data sets'** *Nature methods* **15**: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., Lachmann A 2016) **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update** *Nucleic acids research* **44**:W90–W97

Lacaud G., Kouskoff V 2017) **Hemangioblast, hemogenic endothelium, and primitive versus definitive hematopoiesis** *Experimental hematology* **49**:19–24

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

Li W., Ferkowicz M.J., Johnson S.A., Shelley W.C., Yoder M.C 2005) **Endothelial cells in the early murine yolk sac give rise to CD41-expressing hematopoietic cells** *Stem cells and development* **14**: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., Xiang J 2017) **Conversion of adult endothelium to immunocompetent haematopoietic stem cells** *Nature* **545**:439–445

Love M., Huber W., Anders S 2014) **No title', Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2** *Genome Biology* **15**: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., Goessling W 2020) **YAP regulates hematopoietic stem cell formation in response to the biomechanical forces of blood flow** *Developmental cell* **52**:446–460

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

McGrath K.E., Frame J.M., Fegan K.H., Bowen J.R., Conway S.J., Catherman S.C., Kingsley P.D., Koniski A.D., Palis J 2015) **Distinct sources of hematopoietic progenitors emerge before HSCs and provide functional blood cells in the mammalian embryo** *Cell reports* **11**:1892–1904

McGrath K.E., Palis J 2005) **Hematopoiesis in the yolk sac: more than meets the eye** *Experimental hematology* **33**:1021–1028

Medvinsky A., Dzierzak E 1996) **Definitive hematopoiesis is autonomously initiated by the AGM region** *Cell* **86**:897–906

Medvinsky A., Rybtsov S., Taoudi S 2011) **Embryonic origin of the adult hematopoietic system: advances and questions** *Development* **138**: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., 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**:852–864

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

Moore M.A., 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**:279–296

Moris N., Edri S., Seyres D., Kulkarni R., Domingues A.F., Balayo T., Frontini M., Pina C 2018) **Histone acetyltransferase KAT2A stabilizes pluripotency with control of transcriptional heterogeneity** *Stem cells* **36**:1828–1838

Murry C.E., Keller G 2008) **Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development** *Cell* **132**:661–680

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

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

Niakan K.K., Ji H., Maehr R., Vokes S.A., Rodolfa K.T., Sherwood R.I., Yamaki M., Dimos J.T., Chen A.E., Melton 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**:312–326

North T.E., Goessling W., Peeters M., Li P., Ceol C., Lord A.M., Weber G.J., Harris J., Cutting C.C., Huang P 2009) **Hematopoietic stem cell development is dependent on blood flow** *Cell* **137**:736–748

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

Pastorino F., Brignole C., Marimpietri D., Cilli M., Gambini C., Ribatti D., Longhi R., Allen T.M., Corti A., Ponzoni M 2003) **Vascular damage and anti-angiogenic effects of tumor vessel-targeted liposomal chemotherapy** *Cancer research* **63**:7400–7409

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

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

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

Pina C., May G., Soneji S., Hong D., Enver T 2008) **MLLT3 regulates early human erythroid and megakaryocytic cell fate** *Cell stem cell* **2**:264–273

Quintard C., Tubbs E., Jonsson G., Jiao J., Wang J., Werschler N., Laporte C., Pitaval A., Bah T., Pomeranz G 2024) **A microfluidic platform integrating functional vascularized organoids-on-chip** *Nature Communications* **15**:1452

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

Ragusa D., Dijkhuis L., Pina C., Tosi S 2023) **Mechanisms associated with t (7; 12) acute myeloid leukaemia: from genetics to potential treatment targets** *Bioscience reports* **43**:BSR20220489

Rossi G., Broguiere N., Miyamoto M., Boni A., Guiet R., Girgin M., Kelly R.G., Kwon C., Lutolf M.P 2021) **Capturing cardiogenesis in gastruloids** *Cell stem cell* **28**:230–240

Rossi G., Giger S., Hübscher T., Lutolf M.P 2022) **Gastruloids as in vitro models of embryonic blood development with spatial and temporal resolution** *Scientific Reports* **12**:13380

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

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

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

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

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

Steliarova-Foucher E., Colombet M., Ries L.A., Moreno F., Dolya A., Bray F., Hesseling P., Shin H.Y., Stiller C.A 2017) **International incidence of childhood cancer, 2001–10: a population-based registry study** *The Lancet Oncology* **18**: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., 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**: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., Malleshaiah M. 2017) **Haematopoietic stem and progenitor cells from human pluripotent stem cells** *Nature* **545**:432–438

Taketani T., Taki T., Sako M., Ishii T., Yamaguchi S., 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**:115–119

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

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

Tosi S., Kamel Y.M., Owoka T., Federico C., Truong T.H., 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

Turner D.A., Girgin M., Alonso-Crisostomo L., Trivedi V., Baillie-Johnson P., Glodowski C.R., Hayward P.C., Collignon J., Gustavsen C., 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**:3894–3906

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

Brink Van den, Susanne C Baillie-Johnson, Balayo P., Hadjantonakis T., Nowotschin A., Turner S., Martinez Arias A. 2014) **Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells** *Development* **141**:4231–4242

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

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

Vo L.T., 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**:2641–2648

Waraky A., Östlund A., Nilsson T., Weichenhan D., Lutsik P., Bähr M., Hey J., Tunali G., Adamsson J., 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**:725

Weichenhan D., Riedel A., Sollier E., Toprak U.H., Hey J., Breuer K., Wierzbinska J.A., Touzart A., Lutsik P., Bähr M. 2023) **Altered enhancer-promoter interaction leads to MNX1 expression in pediatric acute myeloid leukemia with t (7; 12)(q36; p13)** *bioRxiv* :2023.09. 13.557546

Wildenhain S., Ingenhag D., Ruckert C., Degistirici O., Dugas M., Meisel R., Hauer J., 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**:40703–40712 https://doi.org/10.1074/jbc.M111.308965

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

Yokomizo T., Ideue T., Morino-Koga S., Tham C.Y., Sato T., Takeda N., Kubota Y., Kurokawa M., Komatsu N., Ogawa M 2022) **Independent origins of fetal liver haematopoietic stem and progenitor cells** *Nature* **609**:779–784

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

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

Author information

Denise Ragusa*
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom, Centre for Genome Engineering and Maintenance, Brunel University London, London, United Kingdom
ORCID iD: 0000-0002-0303-8683

*These authors contributed equally to this work

Chun-Wai Suen*
Department of Genetics, University of Cambridge, Cambridge, United Kingdom

*These authors contributed equally to this work

Gabriel Torregrosa-Cortés*
Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

*These authors contributed equally to this work

Fabio Pastorino*
Laboratory of Experimental Therapies in Oncology, IRCCS Istituto Giannina Gaslini, Genoa, Italy

*These authors contributed equally to this work

Ayona Johns
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom, Centre for Genome Engineering and Maintenance, Brunel University London, London, United Kingdom

Ylenia Cicirò
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom, Centre for Genome Engineering and Maintenance, Brunel University London, London, United Kingdom

Liza Dijkhuis[†]
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom

[†]Present address: Sanquin Research Laboratory, Department of Hematopoiesis, Amsterdam, The Netherlands

Susanne van den Brink
Program in Cancer Research, Hospital de Mar Research Institute, Barcelona, Spain, Josep Carreras Leukemia Research Institute, Barcelona, Spain

Michele Cilli
Animal Facility, IRCCS Policlinico San Martino, Genova, Italy

Connor Byrne[‡]
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom

[‡]Present address: Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

Giulia-Andreea Ionescu[§]
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom

[§]Present address: Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK

Joana Cerveira
Department of Pathology, University of Cambridge, Cambridge, United Kingdom

Kamil R Kranc
Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom, Centre for In Vivo Modelling, Institute of Cancer Research, London, United Kingdom

Victor Hernandez-Hernandez
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom, Centre for Genome Engineering and Maintenance, Brunel University London,

London, United Kingdom

Mirco Ponzoni
Laboratory of Experimental Therapies in Oncology, IRCCS Istituto Giannina Gaslini, Genoa, Italy

Anna Bigas
Program in Cancer Research, Hospital de Mar Research Institute, Barcelona, Spain, Josep Carreras Leukemia Research Institute, Barcelona, Spain
ORCID iD: 0000-0003-4801-6899

Jordi Garcia-Ojalvo
Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

Alfonso Martinez Arias
Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain

Cristina Pina
College of Health, Medicine and Life Sciences, Brunel University London, London, United Kingdom, Centre for Genome Engineering and Maintenance, Brunel University London, London, United Kingdom
ORCID iD: 0000-0002-2575-6301

For correspondence: cristina.pina@brunel.ac.uk

Editors

Reviewing Editor
Owen Tamplin
University of Wisconsin-Madison, Madison, United States of America

Senior Editor
Richard White
University of Oxford, Oxford, United Kingdom

Reviewer #1 (Public review):

Summary

The authors describe a method for gastruloid formation using mouse embryonic stem cells (mESCs) to study YS and AGM-like hematopoietic differentiation. They characterise the gastruloids during nine days of differentiation using a number of techniques including flow cytometry and single-cell RNA sequencing. They compare their findings to a published data set derived from E10-11.5 mouse AGM. At d9, gastruloids were transplanted under the adrenal gland capsule of immunocompromised mice to look for the development of cells capable of engrafting the mouse bone marrow. The authors then applied the gastruloid protocol to study overexpression of Mnx1 which causes infant AML in humans.

In the introduction, the authors define their interpretation of the different waves of hematopoiesis that occur during development. '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.' Herein they designate the yolk sac-derived wave of EMP

hematopoiesis as definitive, according to convention, although paradoxically it does not develop from intra-embryonic mesoderm or give rise to HSCs.

General comments

The authors make the following claims in the paper:

(1) The development of a protocol for hemogenic gastruloids (hGx) that recapitulates YS and AGM-like waves of blood from HE.

(2) The protocol recapitulates both YS and EMP-MPP embryonic blood development 'with spatial and temporal accuracy'.

(3) The protocol generates HSC precursors capable of short-term engraftment in an adrenal niche.

(4) Overexpression of MNX1 in hGx transforms YS EMP to 'recapitulate patient transcriptional signatures'.

(5) hGx is a model to study normal and leukaemic embryonic hematopoiesis.

There are major concerns with the manuscript. The statements and claims made by the authors are not supported by the data presented, data is overinterpreted, and the conclusions cannot be justified. Furthermore, the data is presented in a way that makes it difficult for the reader to follow the narrative, causing confusion. The authors have not discussed how their hGx compares to the previously published mouse embryoid body protocols used to model early development and hematopoiesis.

Specific points

(1) It is claimed that HGxs capture cellularity and topography of developmental blood formation. The hGx protocol described in the manuscript is a modification of a previously published gastruloid protocol (Rossi et al 2022). The rationale for the protocol modifications is not fully explained or justified. There is a lack of novelty in the presented protocol as the only modifications appear to be the inclusion of Activin A and an extension of the differentiation period from 7 to 9 days of culture. No direct comparison has been made between the two versions of gastruloid differentiation to justify the changes.

The inclusion of Activin A at high concentration at the beginning of differentiation would be expected to pattern endoderm rather than mesoderm. BMP signaling is required to induce Flk1+ mesoderm, even in the presence of Wnt. FACS analysis of the hGx during differentiation is needed to demonstrate the co-expression of Flk1-GFP and lineage markers such as CD34 to indicate patterning of endothelium from Flk1+ mesoderm. The FACS plots in Figure 1 show c-Kit expression but very little VE-cadherin which suggests that CD34 is not induced. Early endoderm expresses c-Kit, CXCR4, and Epcam but not CD34 which could account for the lack of vascular structures within the hGx as shown in Figure 1E.

(2) The protocol has been incompletely characterised, and the authors have not shown how they can distinguish between either wave of Yolk Sac (YS) hematopoiesis (primitive erythroid/macrophage and erythro-myeloid EMP) or between YS and intraembryonic Aorta-Gonad-Mesonephros (AGM) hematopoiesis. No evidence of germ layer specification has been presented to confirm gastruloid formation, organisation, and functional ability to mimic early development. Furthermore, differentiation of YS primitive and YS EMP stages of development in vitro should result in the efficient generation of CD34+ endothelial and hematopoietic cells. There is no flow cytometry analysis showing the kinetics of CD34 cell generation during differentiation. Benchmarking the hGx against developing mouse YS and embryo data sets would be an important verification.

Single-cell RNA sequencing was used to compare hGx with mouse AGM. The authors incorrectly conclude that ' ..specification of endothelial and HE cells in hGx follows with time-dependent developmental progression into putative AGM-like HE..' And, '...HE-projected hGx cells.....expressed Gata2 but not Runx1, Myb, or Gfi1b..' Hemogenic endothelium is defined by the expression of Runx1 and Gfli1b is downstream of Runx1.

(3) The hGx protocol 'generates hematopoietic SC precursors capable of short-term engraftment' is not supported by the data presented. Short-term engraftment would be confirmed by flow cytometric detection of hematopoietic cells within the recipient bone marrow, spleen, thymus, and peripheral blood that expressed the BFP transgene. This analysis was not provided. PCR detection of transcripts, following an unspecified number of amplification cycles, as shown in Figure 3G (incorrectly referred to as Figure 3F in the legend) is not acceptable evidence for engraftment. Transplanted hGx formed teratoma-like structures, with hematopoietic cells present at the site of transplant only analysed histologically. Indeed, the quality of the images provided does not provide convincing validation that donor-derived hematopoietic cells were present in the grafts.

There is no justification for the authors' conclusion that '... the data suggest that 216h hGx generate AGM-like pre-HSC capable of at least short-term multilineage engraftment upon maturation...'. Indeed, this statement is in conflict with previous studies demonstrating that pre-HSCs in the dorsal aorta of the mouse embryo are immature and actually incapable of engraftment.

The statement '...low-level production of engrafting cells recapitulates their rarity in vivo, in agreement with the embryo-like qualities of the gastruloid system....' is incorrect. Firstly, no evidence has been provided to show the hGx has formed a dorsal aorta facsimile capable of generating cells with engrafting capacity. Secondly, although engrafting cells are rare in the AGM, approximately one per embryo, they are capable of robust and extensive engraftment upon transplantation.

(4) Expression MNX1 transcript and protein in hematopoietic cells in MNX1 rearranged acute myeloid leukaemia (AML) is one cause of AML in infants. In the hGX model of this disease, Mnx1 is overexpressed in the mESCs that are used to form gastruloids. Mnx1 overexpression seems to confer an overall growth advantage on the hGx and increase the serial replating capacity of the small number of hematopoietic cells that are generated. The inefficiency with which the hGx model generates hematopoietic cells makes it difficult to model this disease. The poor quality of the cytospin images prevents accurate identification of cells. The statement that the kit-expressing cells represent leukemic blast cells is not sufficiently validated to support this conclusion. What other stem cell genes are expressed? Surface kit expression also marks mast cells, frequently seen in clonogenic assays of blood cells. Flow cytometric and gene expression analyses using known markers would be required.

(5) In human infant MNX1 AML, the mutation is thought to arise at the fetal liver stage of development. There is no evidence that this developmental stage is mimicked in the hGx model.

https://doi.org/10.7554/eLife.102324.1.sa3

Reviewer #2 (Public review):

Summary:

In this manuscript, the authors develop an exciting new hemogenic gastruloid (hGX) system, which they claim reproduces the sequential generation of various blood cell types. The key advantage of this cellular system would be its potential to more accurately recapitulate the

spatiotemporal emergence of hematopoietic progenitors within their physiological niche compared to other available in vitro systems. The authors present a large set of data and also validate their new system in the context of investigating infant leukemia.

Strengths:

The development of this new in vitro system for generating hematopoietic cells is innovative and addresses a significant drawback of current in vitro models. The authors present a substantial dataset to characterize this system, and they also validate its application in the context of investigating infant leukemia.

Weaknesses:

The thorough characterization and full demonstration that the cells produced truly represent distinct waves of hematopoietic progenitors are incomplete. The data presented to support the generation of late yolk sac (YS) progenitors, such as lymphoid cells, and aortic-gonad-mesonephros (AGM)-like progenitors, including pre-hematopoietic stem cells (pre-HSCs), by this system are not entirely convincing. Given that this is likely the manuscript's most crucial claim, it warrants further scrutiny and direct experimental validation. Ideally, the identity of these progenitors should be further demonstrated by directly assessing their ability to differentiate into lymphoid cells or fully functional HSCs. Instead, the authors primarily rely on scRNA-seq data and a very limited set of markers (e.g., Ikzf1 and Mllt3) to infer the identity and functionality of these cells. Many of these markers are shared among various types of blood progenitors, and only a well-defined combination of markers could offer some assurance of the lymphoid and pre-HSC nature of these cells, although this would still be limited in the absence of functional assays.

The identification of a pre-HSC-like $CD45^{+}CD41^{-}$/lo c-Kit^{+}VE-$Cadherin^{+}$ cell population is presented as evidence supporting the generation of pre-HSCs by this system, but this claim is questionable. This FACS profile may also be present in progenitors generated in the yolk sac such as early erythro-myeloid progenitors (EMPs). It is only within the AGM context, and in conjunction with further functional assays demonstrating the ability of these cells to differentiate into HSCs and contribute to long-term repopulation, that this profile could be strongly associated with pre-HSCs. In the absence of such data, the cells exhibiting this profile in the current system cannot be conclusively identified as true pre-HSCs.

The engraftment data presented are also not fully convincing, as the observed repopulation is very limited and evaluated only at 4 weeks post-transplantation. The cells detected after 4 weeks could represent the progeny of EMPs that have been shown to provide transient repopulation rather than true HSCs.

https://doi.org/10.7554/eLife.102324.1.sa2

Reviewer #3 (Public review):

In this study, the authors employ a mouse ES-derived "hemogenic gastruloid" model which they generated and which they claim to be able to deconvolute YS and AGM stages of blood production in vitro. This work could represent a valuable resource for the field. However, in general, I find the conclusions in this manuscript poorly supported by the data presented. Importantly, it isn't clear what exactly are the "YS" and the "AGM"-like stages identified in the culture and where is the data that backs up this claim. In my opinion, the data in this manuscript lack convincing evidence that can enable us to identify what kind of hematopoietic progenitor cells are generated in this system. Therefore, the statement that "our study has positioned the MNX1-OE target cell within the YS-EMP stage (line 540)" is not

supported by the evidence presented in this study. Overall, the system seems to be very preliminary and requires further optimization before those claims can be made.

Specific comments below:

(1) The flow cytometric analysis of gastruloids presented in Figure 1 C-D is puzzling. There is a large % of c-Kit+ cells generated, but few VE-Cad+ Kit+ double positive cells. Similarly, there are many CD41+ cells, but very few CD45+ cells, which one would expect to appear toward the end of the differentiation process if blood cells are actually generated. It would be useful to present this analysis as consecutive gating (i.e. evaluating CD41 and CD45 within VE-Cad+ Kit+ cells, especially if the authors think that the presence of VE-Cad+ Kit+ cells is suggestive of EHT). The quantification presented in D is misleading as the scale of each graph is different.

(2) The imaging presented in Figure 1E is very unconvincing. C-Kit and CD45 signals appear as speckles and not as membrane/cell surfaces as they should. This experiment should be repeated and nuclear stain (i.e. DAPI) should be included.

(3) Overall, I am not convinced that hematopoietic cells are consistently generated in these organoids. The authors should sort hematopoietic cells and perform May-Grunwald Giemsa stainings as they did in Figure 6 to confirm the nature of the blood cells generated.

(4) The scRNAseq in Figure 2 is very difficult to interpret. Specific points related to this:
- Cluster annotation in Figure 2a is missing and should be included.
- Why do the heatmaps show the expression of genes within sorted cells? Couldn't the authors show expression within clusters of hematopoietic cells as identified transcriptionally (which ones are they? See previous point)? Gene names are illegible.
- I see no expression of Hlf or Myb in CD45+ cells (Figure 2G). Hlf is not expressed by any of the populations examined (panels E, F, G). This suggests no MPP or pre-HSC are generated in the culture, contrary to what is stated in lines 242-245. (PMID 31076455 and 34589491). Later on, it is again stated that "hGx cells... lacked detection of HSC genes like Hlf, Gfi1, or Hoxa9" (lines 281-283). To me, this is proof of the absence of AGM-like hematopoiesis generated in those gastruloids.

(5) Mapping of scRNA-Seq data onto the dataset by Thambyrajah et al. is not proof of the generation of AGM HE. The dataset they are mapping to only contains AGM cells, therefore cells do not have the option to map onto something that is not AGM. The authors should try mapping to other publicly available datasets also including YS cells.

(6) Conclusions in Figure 3, named "hGx specify cells with preHSC characteristics" are not supported by the data presented here. Again, I am not convinced that hematopoietic cells can be efficiently generated in this system, and certainly not HSCs or pre-HSCs.
- FACS analysis in 3A is again very unconvincing. I do not think the population identified as c-Kit+ CD144+ is real. Also, why not try gating the other way around, as commonly done (e.g. VE-Cad+ Kit+ and then CD41/CD45)?
- The authors must have tried really hard, but the lack of short- or long-engraftment in a number of immunodeficient mouse models (lines 305-313) really suggests that no blood progenitors are generated in their system. I am not familiar with the adrenal gland transplant system, but it seems like a very non-physiological system for trying to assess the maturation of putative pre-HSCs. The data supporting the engraftment of these mice, essentially seen only by PCR and in some cases with a very low threshold for detection, are very weak, and again unconvincing. It is stated that "BFP engraftment of the Spl and BM by flow cytometry was very low level albeit consistently above control (Fig. S4E)" (lines 337-338). I do not think that two dots in a dot plot can be presented as evidence of engraftment.

(7) Given the above, I find that the foundations needed for extracting meaningful data from the system when perturbed are very shaky at best. Nevertheless, the authors proceed to

overexpress MNX1 by LV transduction, a system previously shown to transform fetal liver cells, mimicking the effect of the t(7;12) AML-associated translocation. Comments on this section:
- The increase in the size of the organoid when MNX1 is expressed is a very unspecific finding and not necessarily an indication of any hematopoietic effect of MNX1 OE.
- The mild increase of cKit+ cells (Figure 4E) at the 144hr timepoint and the lack of any changes in CD41+ or CD45+ cells suggests that the increase in Kit+ cells % is not due to any hematopoietic effect of MNX1 OE. No hematopoietic GO categories are seen in RNA seq analysis, which supports this interpretation. Could it be that just endothelial cells are being generated?

(8) There seems to be a relatively convincing increase in replating potential upon MNX1-OE, but this experiment has been poorly characterized. What type of colonies are generated? What exactly is the "proportion of colony forming cells" in Figures 5B-D? The colony increase is accompanied by an increase in Kit+ cells; however, the flow cytometry analysis has not been quantified.

(9) Do hGx cells engraft upon MNX1-OE? This experiment, which appears not to have been performed, is essential to conclude that leukemic transformation has occurred.

https://doi.org/10.7554/eLife.102324.1.sa1

Author response:

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Summary*
>
> *The authors describe a method for gastruloid formation using mouse embryonic stem cells (mESCs) to study YS and AGM-like hematopoietic differentiation. They characterise the gastruloids during nine days of differentiation using a number of techniques including flow cytometry and single-cell RNA sequencing. They compare their findings to a published data set derived from E10-11.5 mouse AGM. At d9, gastruloids were transplanted under the adrenal gland capsule of immunocompromised mice to look for the development of cells capable of engrafting the mouse bone marrow. The authors then applied the gastruloid protocol to study overexpression of Mnx1 which causes infant AML in humans.*
>
> *In the introduction, the authors define their interpretation of the different waves of hematopoiesis that occur during development. '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-gonadmesonephros (AGM) region of the embryo proper.' Herein they designate the yolk sac-derived wave of EMP hematopoiesis as definitive, according to convention, although paradoxically it does not develop from intraembryonic mesoderm or give rise to HSCs.*

The apparent perplexity of the Reviewer with our definition of primitive and definitive waves is somewhat surprising, as it is widely used in the field (e.g. PMID: 18204427; PMID: 28299650; PMID: 33681211). Definitive haematopoiesis, encompassing EMP, MLP, MPP and HSC, highlights their origin from haemogenic hendothelium, generation of mature cells with adult characteristics from progenitors with multilineage potential and direct and indirect developmental contributions to the intra-embryonic and time-restricted generation of HSCs.

> *General comments*
>
> *The authors make the following claims in the paper:*
>
> *(1) The development of a protocol for hemogenic gastruloids (hGx) that recapitulates YS and AGM-like waves of blood from HE.*
>
> *(2) The protocol recapitulates both YS and EMP-MPP embryonic blood development 'with spatial and temporal accuracy'.*
>
> *(3) The protocol generates HSC precursors capable of short-term engraftment in an adrenal niche.*
>
> *(4) Overexpression of MNX1 in hGx transforms YS EMP to 'recapitulate patient transcriptional signatures'.*
>
> *(5) hGx is a model to study normal and leukaemic embryonic hematopoiesis.*
>
> *There are major concerns with the manuscript. The statements and claims made by the authors are not supported by the data presented, data is overinterpreted, and the conclusions cannot be justified. Furthermore, the data is presented in a way that makes it difficult for the reader to follow the narrative, causing confusion. The authors have not discussed how their hGx compares to the previously published mouse embryoid body protocols used to model early development and hematopoiesis. the data is presented in a way that makes it difficult for the reader to follow the narrative, causing confusion. The authors have not discussed how their hGx compares to the previously published mouse embryoid body protocols used to model early development and hematopoiesis.*
>
> *Specific points*
>
> *(1) It is claimed that HGxs capture cellularity and topography of developmental blood formation. The hGx protocol described in the manuscript is a modification of a previously published gastruloid protocol (Rossi et al 2022). The rationale for the protocol modifications is not fully explained or justified. There is a lack of novelty in the presented protocol as the only modifications appear to be the inclusion of Activin A and an extension of the differentiation period from 7 to 9 days of culture. No direct comparison has been made between the two versions of gastruloid differentiation to justify the changes.*

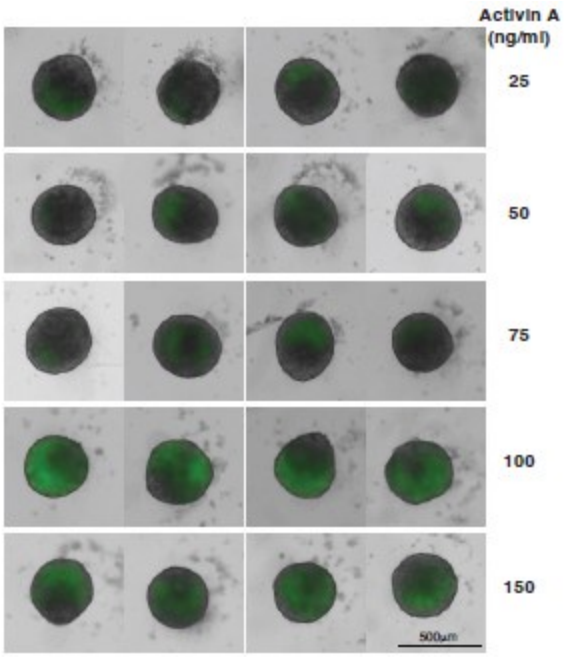
The Reviewer paradoxically claims that the protocol is not novel and that it differs from a previous publication in at least 2 ways – the patterning pulse and the length of the protocol. Of these, the patterning pulse is key. As documented in Fig. S1, we cannot obtain Flk1-GFP expression in the absence of Activin A. Expression of Flk1 is a fundamental step in haemato-endothelial specification and, accordingly, we do not see CD41 or CD45+ cells in the absence of Activin A. Also, in our hands, there is a clear time-dependent progression of marker expression, with sequential acquisition of CD41 and CD45, with the latter not detectable until 192h (Fig. 1C-D), another key difference relative to the Rossi et al (2022) protocol. The 192h-timepoint, we argue in the manuscript, and present further evidence for in this rebuttal, corresponds to the onset of AGM-like haematopoiesis. We have empirically extended the protocol to maximise the CD45+ cell output (Fig. S1B-D).

> *The inclusion of Activin A at high concentration at the beginning of differentiation would be expected to pattern endoderm rather than mesoderm. BMP signaling is required to induce Flk1+ mesoderm, even in the presence of Wnt.*

Again, we call the Reviewer's attention to Fig. S1 which clearly shows that Activin A (with no BMP added) is required for induction of Flk1 expression, in the presence of Wnt. Activin A in combination with Wnt, is used in other protocols of haemato-endothelial differentiation from pluripotent cells, with no BMP added in the same step of patterning and differentiation (PMID: 39227582; PMID: 39223325). In the latter protocol, we also call the Reviewer's attention to the fact that a higher concentration of Activin A precludes the need for BMP4 addition. Finally, one of us has recently reported that Activin A, on its own, will induce FLK1, as well as other anterior mesodermal progenitors (https://www.biorxiv.org/content/10.1101/2025.01.11.632562v1). In addressing the Reviewer's concerns with the dose of Activin A used, we titrated its concentration against activation of Flk1, confirming optimal Flk1-GFP expression at the 100ng/ml dose used in the manuscript.

Author response image 1.

Dose-dependent requirement of Activin A for induction of Flk1 expression in haemogenic gastruloids. Composite GFP and brightfield live imaging of Flk1-GFP haemogenic gastruloids at 96h. Images were acquired using a Cytation5 instrument (Thermo). Images are representative of 12 gastruloids per condition.



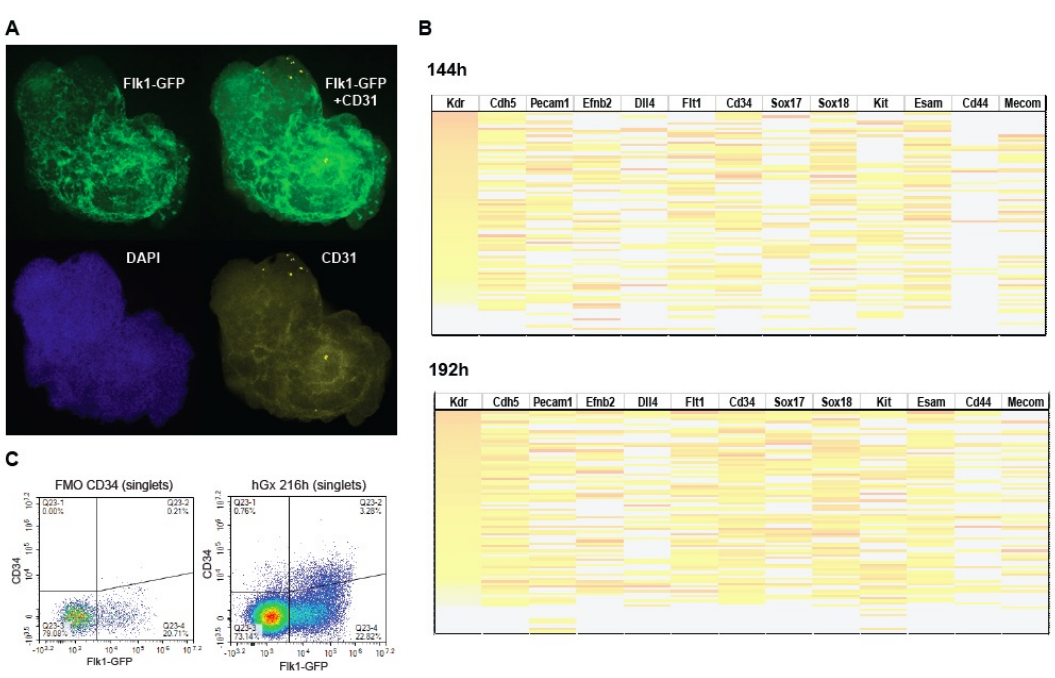
> *FACS analysis of the hGx during differentiation is needed to demonstrate the co-expression of Flk1-GFP and lineage markers such as CD34 to indicate patterning of endothelium from Flk1+ mesoderm. The FACS plots in*
>
> *Fig. 1 show c-Kit expression but very little VE-cadherin which suggests that CD34 is not induced. Early endoderm expresses c-Kit, CXCR4, and Epcam, but not CD34 which could account for the lack of vascular structures within the hGx as shown in Fig. 1E.*

We were surprised by the Reviewer's comment that there are no endothelial structures in our gastruloids. The presence of a Flk1-GFP+ network is visible in the GFP images in Fig.1B, from 144h onwards, also shown in Author response image 2A. In addition, our single-cell RNA-seq data, included in the manuscript, confirms the presence of endothelial cells with a developing endothelial, including arterial, programme. This can be seen in Fig. 2B, F of the manuscript

and is represented in Author response image 2B. In contrast with the Reviewer's claims that no endothelial cells are formed, the data show that Kdr (Flk1)+ cells co-express Cdh5/VE-Cadherin and indeed Cd34, attesting to the presence of an endothelial programme. Arterial markers Efnb2, Flt1, and Dll4 are present. A full-blown programme, which also includes haemogenic markers including Sox17, Esam, Cd44 and Mecom is clear at early (144h) and, particularly at late (192h) timepoints in cells sorted on detection of surface c-Kit (Author response image 2B). Further to the data shown in B, already present in the manuscript, we also document co-expression of Flk1-GFP and CD34 by flow cytometry (Author response image 2C).

Author response image 2.

Haemogenic gastruloids have a branched vascular network. A. Whole-mount confocal imaging of 144h-haemogenic gastruloids. B. Differentiation of an arterial endothelial programme in haemogenic gastruloids; singlecell RNA-seq data of differentiating haemogenic gastruloids, sorted on cell surface expression of c-Kit at 144 and 192h; gene expression colour scale from yellow (low) to orange (high); grey = no detectable expression. C. Flow cytometry plots of 216h-haemogenic gastruloids showing detection of haemato-endothelial marker CD34.



(2) The protocol has been incompletely characterised, and the authors have not shown how they can distinguish between either wave of Yolk Sac (YS) hematopoiesis (primitive erythroid/macrophage and erythro-myeloid EMP) or between YS and intraembryonic Aorta-Gonad-Mesonephros (AGM) hematopoiesis. No evidence of germ layer specification has been presented to confirm gastruloid formation, organisation, and functional ability to mimic early development. Furthermore, differentiation of YS primitive and YS EMP stages of development in vitro should result in the efficient generation of CD34+ endothelial and hematopoietic cells. There is no flow cytometry analysis showing the kinetics of CD34 cell generation during differentiation. Benchmarking the hGx against developing mouse YS and embryo data sets would be an important verification.

The Reviewer is correct that we have not provided detailed characterisation of the different germ layers, as this was not the focus of the study. In that context, we were surprised by the earlier comment assuming co-expression of c-kit, Cxcr4 and Epcam, which we did not show, while overlooking the endothelial programme reiterated above, which we have presented.

Given our focus on haemato-endothelial specification, we have started the single-cell RNA-seq characterisation of the haemogenic gastruloid at 120h and have not looked specifically at earlier timepoints of embryo patterning.

This said, we show the presence of neuroectodermal cells in cluster 9; on the other hand, cluster 7 includes hepatoblast-like cells, denoting endodermal specification. We are happy to include this characterisation, to the extent that it is present, in a revised version of the manuscript. However, in the absence of earlier timepoints and given the bias towards mesodermal specification, we expect that specification of ectodermal and endodermal programmes may be incomplete.

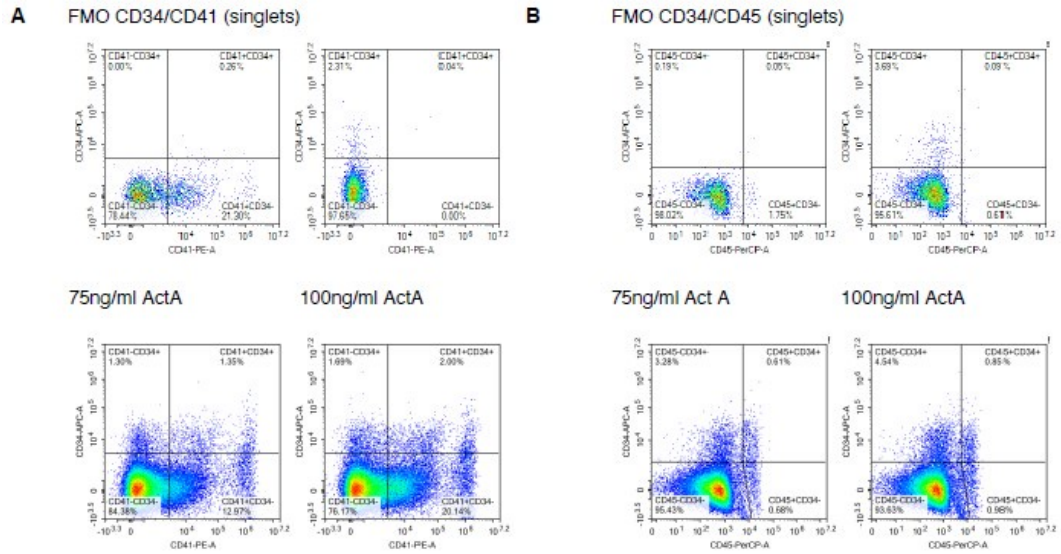
In respect of the contention regarding the capture of YS-like and AGM-like haematopoiesis, we have presented evidence in the manuscript that haemogenic cells generated during gastruloid differentiation, particularly at late 192h and 216h timepoints project onto highly purified c-Kit+ CD31+ Gfi1-expressing cells from mouse AGM (PMID: 38383534), providing support for the recapitulation of the corresponding developmental stage. In distinguishing between YS-like and AGM-like haematopoiesis, we call the Reviewer's attention to the replotting of the single-cell RNA-seq data already in the manuscript, which we provided in response to point 1 (Author response image 2B), which highlights an increase in Sox17, but not Sox18, expression in the 192h haemogenic endothelium, which suggests an association with AGM haematopoiesis (PMID: 20228271). A significant association of Cd44 and Procr expression with the same time-point (Fig. 2F in the manuscript), further supports an AGM-like endothelial-to-haematopoietic transition at the 192h timepoint.

Following on the Reviewer's comments about CD34, we also inspected co-expression of CD34 with CD41 and CD45, the latter co-expression present in, although not necessarily exclusive to, AGM haematopoiesis.

Reassuringly, we observed clear co-expression with both markers (Author response image 3), in addition to a CD41+CD34-population, which likely reflects YS EMP-independent erythropoiesis. Interestingly, marker expression is responsive to the levels of Activin A used in the patterning pulse, with the 100ng/ml Activin A used in our protocol superior to 75ng/ml.

Author response image 3.

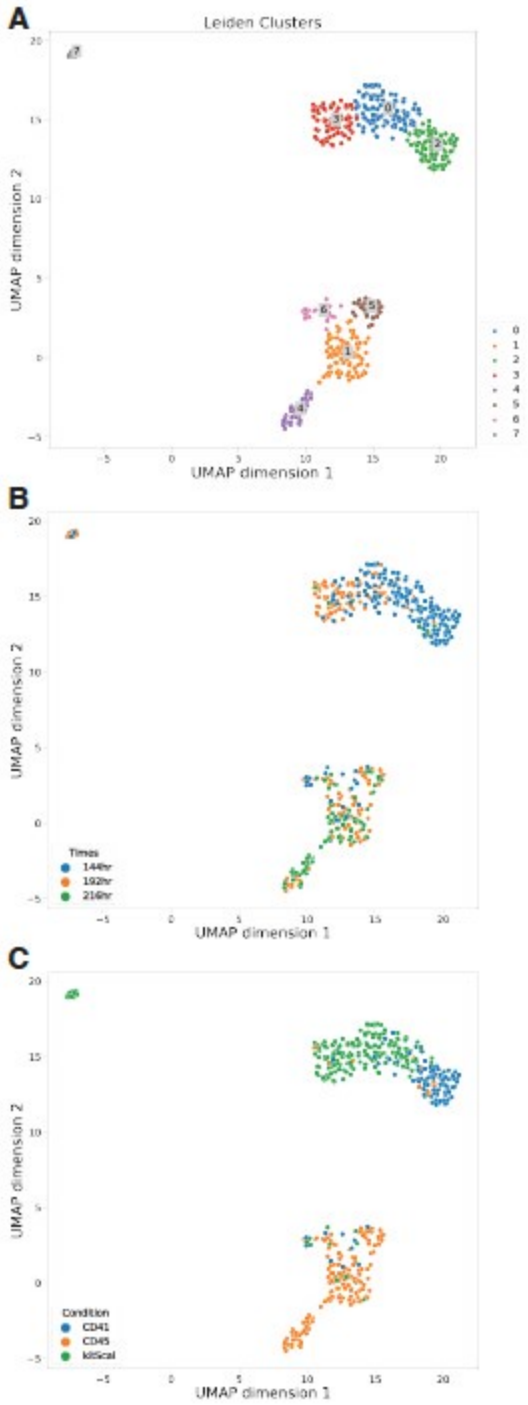
Association of CD34 with CD41 and CD45 expression is Activin A-responsive and supports the presence of definitive haematopoiesis. A. Flow cytometry analysis of CD34 and CD41 expression in 216h-haemogenic gastruloids; two doses of Activin A were used in the patterning pulse with CHI99021 between 48-72h. FMO controls shown. B. Flow cytometry analysis of CD34 and CD45 at 216h in the same experimental conditions.



We agree that it remains challenging to identify markers exclusive to AGM haematopoiesis, which is operationally equated with generation of transplantable haematopoietic stem cells. While HSC generation is a key event characteristic of the AGM, not all AGM haematopoiesis corresponds to HSCs, an important point in evaluating the data presented in the manuscript, and indeed acknowledged by us.

Author response image 4.

Clustering of haemogenic gastruloid cells sorted on the basis of haemato-endothelial surface markers CD41, C-Kit and CD45. A. Leiden clustering to single-cell RNA-seq data. B. Time stamps of sorted haemogenic gastruloid cells in A. C. Surface marker stamps of cells in A.



Given the centrality of this point in comments by all the Reviewers, we have conducted projections of our single-cell RNA-seq data against two studies which (1) capture arterial and haemogenic specification in the para-splanchnopleura (pSP) and AGM region between E8.0 and E11 (Hou et al, PMID: 32203131), and (2) uniquely capture YS, AGM and FL progenitors and the AGM endothelial-to-haematopoietic transition (EHT) in the same scRNA-seq dataset (Zhu et al, PMID: 32392346).

Focusing the analysis on the subsets of haemogenic gastruloid cells sorted as CD41+ (144h) CKit+ (144h and 192h) and CD45+ (192h and 216h) (Author response image 4AC), we show:

(1) That a subset of haemato-endothelial cells from haemogenic gastruloids at 144h to 216h project onto intra-embryonic cells spanning E8.25 to E10 (Author response image 5A-B). This is in agreement with our interpretation that 216h are no later than the MPP/pre-HSC state of embryonic development, requiring further maturation to generate long-term engrafting HSC.

(2) That haemogenic gastruloids contain YS-like (including EMP-like) and AGM-like haematopoietic cells (Author response image 6A-B). Significantly, some of the cells, particularly c-Kit-sorted cells with a candidate endothelial and HE-like signature project onto AGM pre-HE and HE, as well as IAHC, and later, predominantly 216h cells, have characteristics of MPP/LMPP-like cells from the FL.

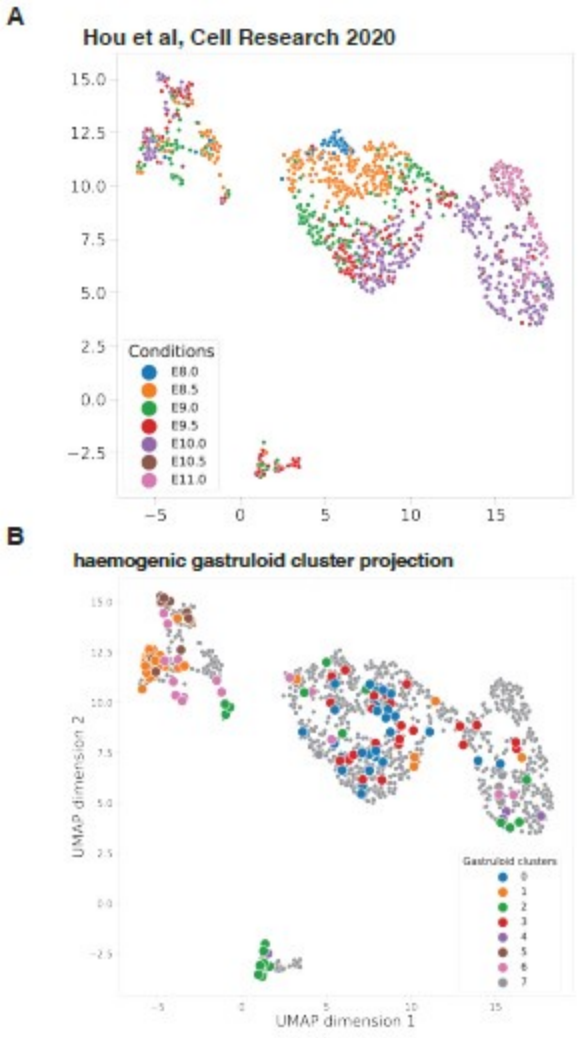
Altogether, the data support the notion that haemogenic gastruloids capture YS and AGM haematopoiesis until E10, as suggested by us in the manuscript. We thought it was important to share this preliminary data with the Editors at an early stage, and we will incorporate a deeper analysis in a revised version of the manuscript.

> *Single-cell RNA sequencing was used to compare hGx with mouse AGM. The authors incorrectly conclude that ' ..specification of endothelial and HE cells in hGx follows with time-dependent developmental progression into putative AGM-like HE..' And, '...HE-projected hGx cells.....expressed Gata2 but not Runx1, Myb, or Gfi1b..' Hemogenic endothelium is defined by the expression of Runx1 and Gfli1b is downstream of Runx1.*

As a hierarchy of regulation, Gata2 precedes and drives Runx1 expression at the specification of HE (PMID: 17823307; PMID: 24297996), while Runx1 drives the EHT, upstream of Gfi1b in haematopoietic clusters (PMID: 34517413).

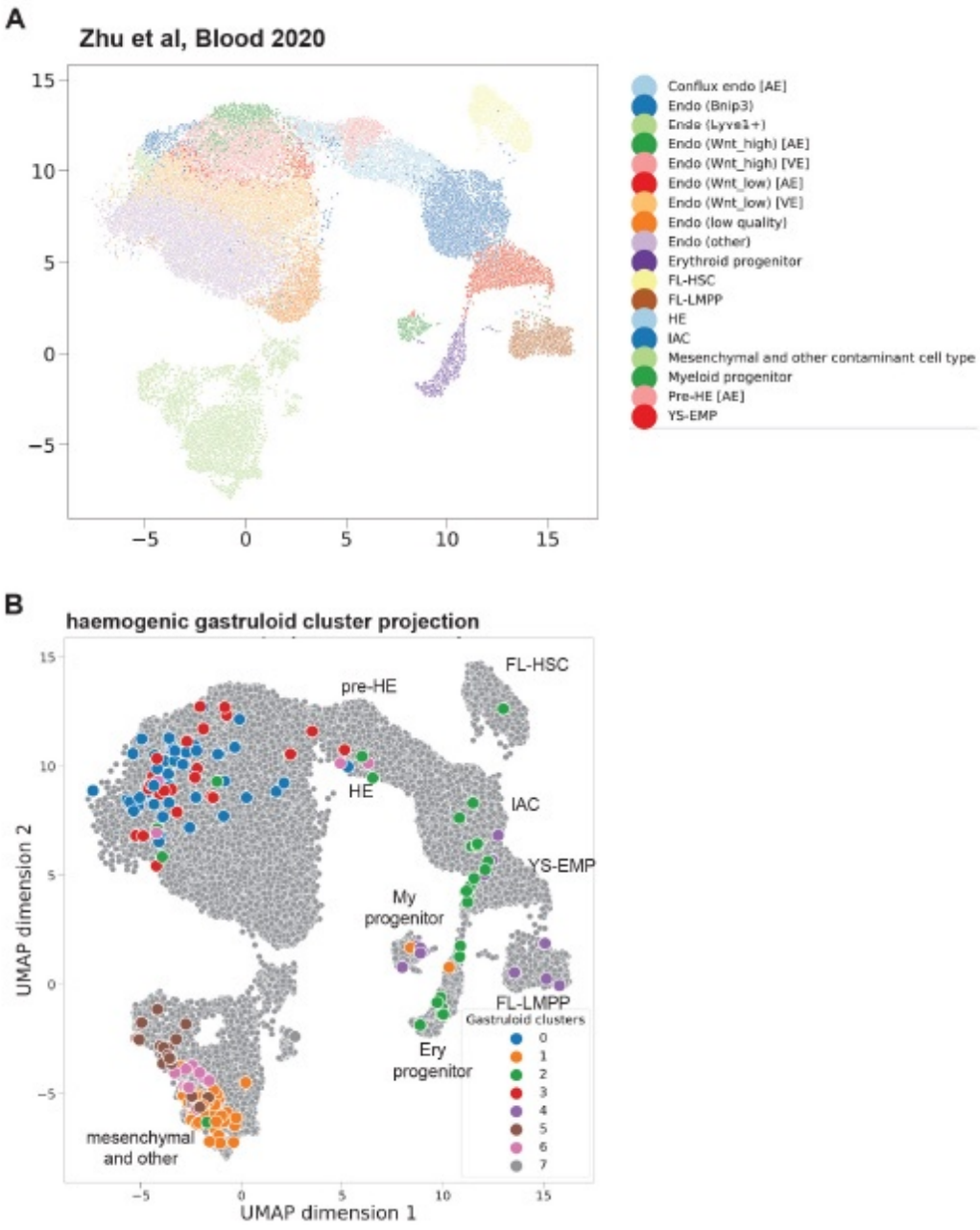
Author response image 5.

Projection of sorted haemogenic gastruloid cells onto Hou et al dataset (PMID: 32203131) analysing development of mouse intra-embryonic haematopoiesis. A. Time signatures of Hou et al data. B. Projection of Leiden clusters in Author response image 4A. Methodology as described in our manuscript; 68% gastruloid cells projected.



Author response image 6.

Projection of sorted haemogenic gastruloid cells onto Zhu et al dataset (PMID: 32392346), capturing arterial endothelial and haemogenic endothelial development, in reference to YS, AGM and FL haematopoietic progenitors. A. Functional cluster classification as per Zhu et al. B. Projection of Leiden clusters in Author response image 4A. Methodology as detailed in our manuscript; 58% gastruloid cells projected. Haematopoietic clusters annotated as in A.



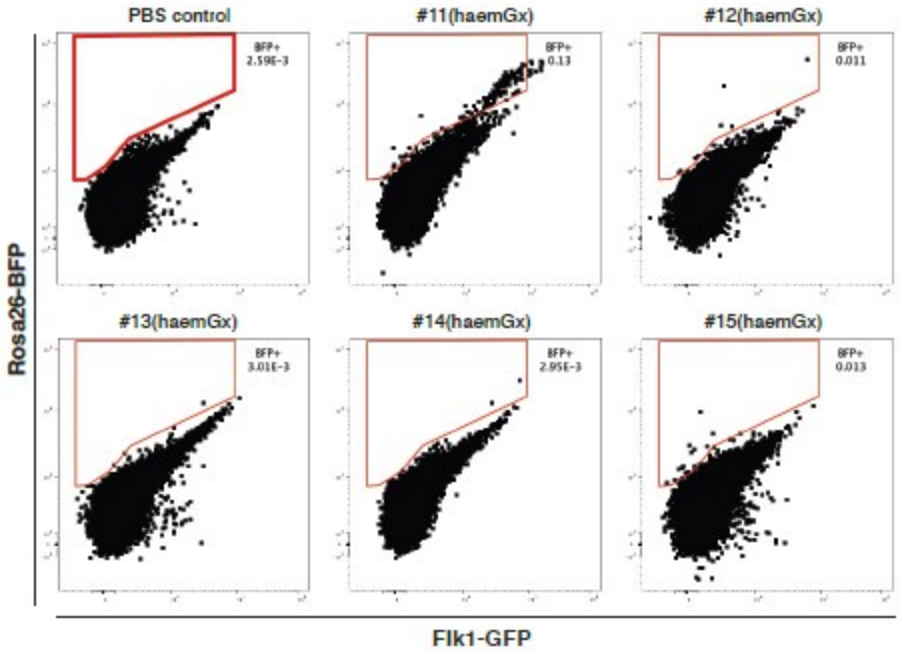
> *(3) The hGx protocol 'generates hematopoietic SC precursors capable of short-term engraftment' is not supported by the data presented. Short-term engraftment would be confirmed by flow cytometric detection of hematopoietic cells within the recipient bone marrow, spleen, thymus, and peripheral blood that expressed the BFP transgene. This analysis was not provided. PCR detection of transcripts, following an unspecified number of amplification cycles, as shown in Figure 3G (incorrectly referred to as Figure 3F in the legend) is not acceptable evidence for engraftment.*

We provide the full flow cytometry analysis of spleen engraftment in the 5 mice which received implantation of 216h-haemogenic gastruloids in the adrenal gland; an additional (control) animal received adrenal injection of PBS (Author response image 7). The animals were analysed at 4 weeks. In this experiment, the bone marrow collection was limiting, and material was prioritised for PCR.

We had previously provided only representative plots of flow cytometry analysis of bone marrow and spleen in Fig. S4E, which we described as low-level engraftment. The analysis was complemented with genomic DNA PCR, where detection was present in only some of the replicates tested per animal. We confirm that PCR analysis used conventional 40 cycles; the sensitivity was shown in Fig. S4F. As shown in Fig. 3 A-C, no more than 7 CD45+CD144+ multipotent cells are present per haemogenic gastruloid, with 3 haemogenic gastruloids implanted in the adrenal gland of each transplanted animal. We argue that the low level of cytometric and molecular engraftment at 4 weeks, from haemogenic gastruloid-derived progenitors that have not progressed beyond a stage equivalent to E10 Author response image 5A-B) and that we have described as requiring additional maturation in vivo, are not surprising.

Author response image 7.

BFP engraftment of Nude recipient mice 4 weeks after unilateral adrenal implantation of 216h-haemogenic gastruloids. Flow cytometry analysis of spleen engraftment. Genomic PCR analysis is shown in Fig. 3G of the manuscript.



> *Transplanted hGx formed teratoma-like structures, with hematopoietic cells present at the site of transplant only analysed histologically. Indeed, the quality of the images provided does not provide convincing validation that donor-derived hematopoietic cells were present in the grafts.*

As stated in the text, the images mean to illustrate that the haemogenic gastruloids developed in situ. The observation of donor-derived blood cells in the implanted haemogenic gastruloids would not correspond to engraftment, as we have amply demonstrated that they have generated blood cells in vitro. There is no evidence that there are remaining pluripotent cells in the haemogenic gastruloid after 9 days of differentiation, and it is therefore not clear that these are teratomas

> *There is no justification for the authors' conclusion that '... the data suggest that 216h hGx generate AGM-like pre-HSC capable of at least short-term multilineage engraftment*

> *upon maturation...'. Indeed, this statement is in conflict with previous studies demonstrating that pre-HSCs in the dorsal aorta of the mouse embryo are immature and actually incapable of engraftment.*

We have clearly stated that we do not see haematopoietic engraftment through transplantation of dissociated haemogenic gastruloids, which reach the E10 state containing pre-HSC (Author response image 5). Instead, we observed rare myelo-erythroid (in the manuscript) and myelo-lymphoid (Author response image 9 below, in response to Reviewer 2) engraftment upon in vivo maturation of haemogenic gastruloids with preserved 3D organisation. These statements are not contradictory.

> *The statement '...low-level production of engrafting cells recapitulates their rarity in vivo, in agreement with the embryo-like qualities of the gastruloid system....' is incorrect. Firstly, no evidence has been provided to show the hGx has formed a dorsal aorta facsimile capable of generating cells with engrafting capacity. Secondly, although engrafting cells are rare in the AGM, approximately one per embryo, they are capable of robust and extensive engraftment upon transplantation.*

We are happy to rephrase the statement to simply say that "...the data suggest that 216h haemogenic gastruloids contain candidate AGM-like progenitors with some short-term engraftment potential but incomplete functional maturation." To be clear, with our existing statement we meant to highlight that the production of definitive AGM-like haematopoietic progenitors (not all of which are engrafting) in haemogenic gastruloids does not correspond to non-physiological single-lineage programming. We did not claim that we achieved production of HSC, which would be long-term engrafting.

> *(4) Expression MNX1 transcript and protein in hematopoietic cells in MNX1 rearranged acute myeloid leukaemia (AML) is one cause of AML in infants. In the hGX model of this disease, Mnx1 is overexpressed in the mESCs that are used to form gastruloids. Mnx1 overexpression seems to confer an overall growth advantage on the hGx and increase the serial replating capacity of the small number of hematopoietic cells that are generated. The inefficiency with which the hGx model generates hematopoietic cells makes it difficult to model this disease. The poor quality of the cytospin images prevents accurate identification of cells. The statement that the kit-expressing cells represent leukemic blast cells is not sufficiently validated to support this conclusion. What other stem cell genes are expressed? Surface kit expression also marks mast cells, frequently seen in clonogenic assays of blood cells. Flow cytometric and gene expression analyses using known markers would be required.*

The haemogenic gastruloid model generates haematopoietic and haemato-endothelial cells. MNX1 expands Kit+ cells at 144h, which we show to have a haemato-endothelial signature (manuscript Fig. 2B, which we replotted in Author response image 2B).

Serial replating of CFC assays is a conventional in vitro assay of leukaemia transformation. Critically, colony replating is not maintained in EV control cells, attesting to the transformation potential of MNX1.

Although we have not fully-traced the cellular hierarchy of MNX1-driven transformation in the haemogenic gastruloid system, the in vitro replating expands a Kit+ cell (Fig. 5E), which reflects the surface phenotype of the leukaemia, also recapitulated in the mouse model initiated by MNX1-overexpressing FL cells. Importantly, it recapitulates the transcriptional profile of MNX1-leukaemia patients (Fig. 6C), which is uniquely expressed by MNX1144h and replated colony cells, but not to MNX1 216h gastruloid cells, arguing against a generic signature of MNX1 overexpression (Fig. 6B). Importantly, the MNX1-transformation of haemogenic gastruloid cells is superior to the FL leukaemia model at capturing the unique

transcriptional features of MNX1-driven leukaemia, distinct from other forms of AML in the same age group (Fig S7). It is possible that this corresponds to a preleukaemia event, and we will explore this in future studies, which are beyond the proof-of-principle nature of this paper.

> *(5) In human infant MNX1 AML, the mutation is thought to arise at the fetal liver stage of development. There is no evidence that this developmental stage is mimicked in the hGx model.*

We never claim that the haemogenic gastruloid model mimics the foetal liver. We propose that susceptibility to MNX1 is at the HE-to-EMP transition. Moreover, and importantly, contrary to the Reviewer's statement, there is no evidence in the literature that the mutation arises in the foetal liver stage, just that the mutation arises before birth (PMID: 38806630), which is different. In a mouse model of MNX1 overexpression, the authors achieve leukaemia engraftment upon MNX1 overexpression in foetal liver, but not in bone marrow cells (PMID: 37317878). This is in agreement with a vulnerability of embryonic / foetal, but not adult cells to the MNX1 expression caused by the translocation. However, haematopoietic cells in the foetal liver originate from YS and AGM precursors, so the origin of the MNX1-susceptible cells can be in those locations, rather than the foetal liver itself.

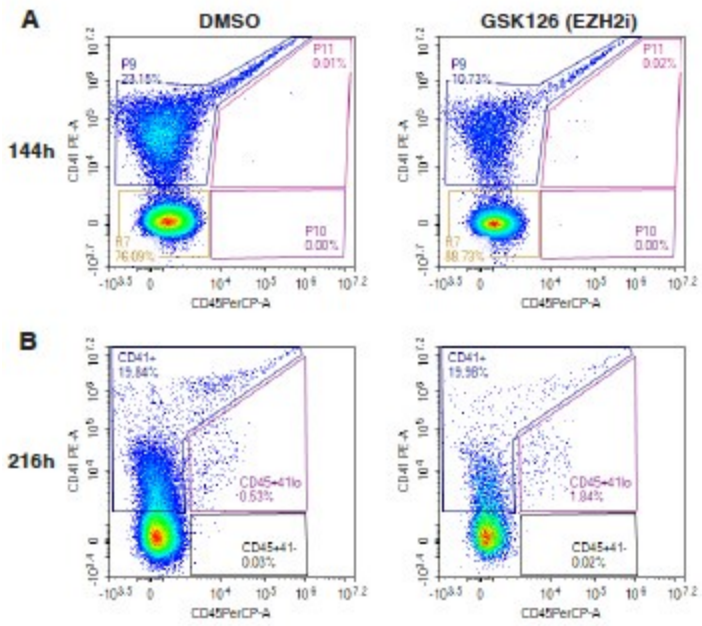
> ***Reviewer #2 (Public review):***
> *Summary:*
> *In this manuscript, the authors develop an exciting new hemogenic gastruloid (hGX) system, which they claim reproduces the sequential generation of various blood cell types. The key advantage of this cellular system would be its potential to more accurately recapitulate the spatiotemporal emergence of hematopoietic progenitors within their physiological niche compared to other available in vitro systems. The authors present a large set of data and also validate their new system in the context of investigating infant leukemia.*
> *Strengths:*
> *The development of this new in vitro system for generating hematopoietic cells is innovative and addresses a significant drawback of current in vitro models. The authors present a substantial dataset to characterize this system, and they also validate its application in the context of investigating infant leukemia.*
> *Weaknesses:*
> *The thorough characterization and full demonstration that the cells produced truly represent distinct waves of hematopoietic progenitors are incomplete. The data presented to support the generation of late yolk sac (YS) progenitors, such as lymphoid cells, and aortic-gonad-mesonephros (AGM)-like progenitors, including pre-hematopoietic stem cells (pre-HSCs), by this system are not entirely convincing. Given that this is likely the manuscript's most crucial claim, it warrants further scrutiny and direct experimental validation. Ideally, the identity of these progenitors should be further demonstrated by directly assessing their ability to differentiate into lymphoid cells or fully functional HSCs. Instead, the authors primarily rely on scRNA-seq data and a very limited set of markers (e.g., Ikzf1 and Mllt3) to infer the identity and functionality of these cells. Many of these markers are shared among various types of blood progenitors, and only a well-defined combination of markers could offer some assurance of the lymphoid and pre-HSC nature of these cells, although this would still be limited in the absence of functional assays.*
> *The identification of a pre-HSC-like CD45^{+}CD41$^{-/lo}$ c-Kit^{+}VE-Cadherin^{+} cell population is presented as evidence supporting the generation of pre-HSCs by this system, but this claim is questionable. This FACS profile may also be present in progenitors generated in the yolk sac such as early erythro-myeloid progenitors (EMPs). It is only within the AGM context, and in conjunction with further functional assays demonstrating the ability of*

> *these cells to differentiate into HSCs and contribute to long-term repopulation, that this profile could be strongly associated with pre-HSCs. In the absence of such data, the cells exhibiting this profile in the current system cannot be conclusively identified as true pre-HSCs.*

At this preliminary response stage, we present 2 additional pieces of evidence to support our claims that we capture YS and AGM stages of haematopoietic development. In future experiments, we can complement these with functional assays, including co-culture with OP9 and OP9-DL stroma.

Author response image 8.

EZH2 inhibition affects CD41+ cellular output in haemogenic gastruloids at 144, but not 216h. A. Flow cytometry analysis of CD41 expression in 144h-haemogenic gastruloid treated with 0.5μM EZH2 inhibitor GSK126 from 120h. DMSO (0.05%), vehicle. 1 of 2 independent experiments (average CD41+: DMSO, 21.20%; GSK126, 12.10%; CD45 not detected). B. Flow cytometry analysis of CD41 and CD45 expression in 216h gastruloids, treated with DMSO or GSK216. (DMSO: average CD41+, 15.28%; average CD45+ 0.46%. GSK126: average CD41+, 23.78%; average CD45+, 2.08%).



In Author response images 5 and 6, we project our single-cell RNA-seq data onto (1) developing intra-embryonic pSP and AGM between E8 and E11 (Author response image 5) and (2) a single-cell RNA-seq study of HE development which combines haemogenic and haematopoietic cells from the YS, the developing HE and IAHC in the AGM, and FL (Author response image 6). Our data maps E8.25-E10 (Author response image 5) and captures YS EMP and erythroid and myeloid progenitors, as well as AGM pre-HE, HE and IAHC, with some cells matching HSPC and LMPP (Author response image 6), as suggested by the projection onto the Thambyrajah et al data set (Fig. S3 in the manuscript).

Given the difficulty in finding markers that specifically associate with AGM haematopoiesis, we inspected the possibility of capturing different regulatory requirements at different stages of gastruloid development mirroring differential effects in the embryo. Polycomb EZH2 is specifically required for EMP differentiation in the YS, but does not affect AGM-derived haematopoiesis; it is also not required for primitive erythroid cells (PMID: 29555646; PMID:

34857757). We treated haemogenic gastruloids from 120h onwards with either DMSO (0.05%) or GSK126 (0.5μM), and inspected the cellularity of gastruloids at 144h, which we equate with YS-EMP, and 216h – putatively AGM haematopoiesis (Author response image 8). We show that EZH2 inhibition / GSK126 treatment specifically reduces %CD41+ cells at 144h (Author response image 8A), but does not reduce %CD41+ or %CD45+ cells at 216h (Author response image 8B).

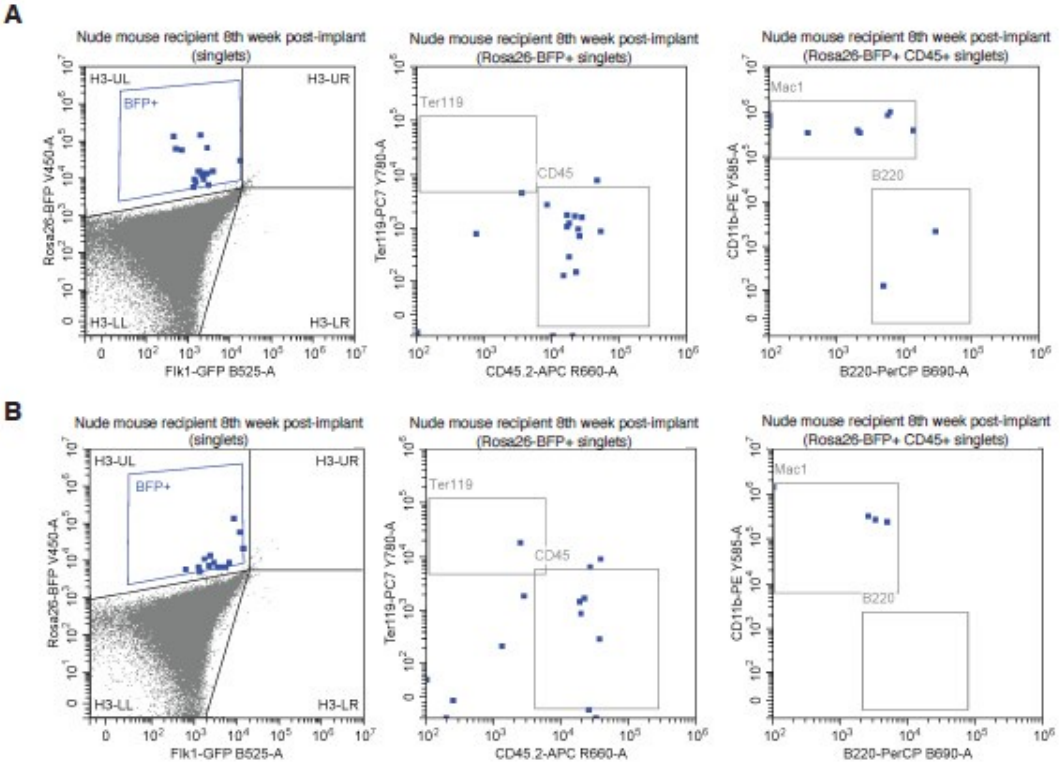
Although preliminary, these data, together with the scRNA-seq projections described, provide evidence to our claim that 144h haemogenic gastruloids capture YS EMPs, while CD41+ and CD45+ cells isolated at 216h reflect AGM progenitors. We cannot conclude as to the functional nature of the AGM cells from this experiment.

> *The engraftment data presented are also not fully convincing, as the observed repopulation is very limited and evaluated only at 4 weeks post-transplantation. The cells detected after 4 weeks could represent the progeny of EMPs that have been shown to provide transient repopulation rather than true HSCs.*

We clearly state that there is low level engraftment and do not claim to have generated HSC. We describe cells with short-term engraftment potential. Although the cells we show in the manuscript at 4 weeks could be EMPs (Author response image 7 and Fig. 3 and S3), we now have 8-week analysis of implant recipients, in which we observed, again low-level, engraftment of the recipient bone marrow in 1:3 animals (Author response image 9). This engraftment is myeloid-lymphoid and therefore likely to have originated in a later progenitor. To be clear, we do not claim that this corresponds to the presence of HSC. It nevertheless supports the maturation of progenitors with engraftment potential.

Author response image 9.

Flow cytometry BFP engraftment of recipient bone marrow 8-weeks post implantation of 216hhaemogenic gastruloids in the adrenal gland of Nude mice. 1:3 animals show BFP CD45+ engraftment in the myeloid (Mac1+) and B-lymphoid (B220+) lineages. 3 haemogenic gastruloids were implanted unilaterally in the adrenal gland of each animal. A. Engrafted animal, showing CD45+ BFP cells of myeloid (CD11b) and B-lymphoid affiliation (B220). B. Non-engrafted mouse recipient of haemogenic gastruloid implants.



> ***Reviewer #3 (Public review):***
> *In this study, the authors employ a mouse ES-derived "hemogenic gastruloid" model which they generated and which they claim to be able to deconvolute YS and AGM stages of blood production in vitro. This work could represent a valuable resource for the field. However, in general, I find the conclusions in this manuscript poorly supported by the data presented. Importantly, it isn't clear what exactly are the "YS" and the "AGM"-like stages identified in the culture and where is the data that backs up this claim. In my opinion, the data in this manuscript lack convincing evidence that can enable us to identify what kind of hematopoietic progenitor cells are generated in this system. Therefore, the statement that "our study has positioned the MNX1-OE target cell within the YS-EMP stage (line 540)" is not supported by the evidence presented in this study. Overall, the system seems to be very preliminary and requires further optimization before those claims can be made.*
> *Specific comments below:*
> *(1) The flow cytometric analysis of gastruloids presented in Figure 1 C-D is puzzling. There is a large % of c-Kit+ cells generated, but few VE-Cad+ Kit+ double positive cells. Similarly, there are many CD41+ cells, but very few CD45+ cells, which one would expect to appear toward the end of the differentiation process if blood cells are actually generated. It would be useful to present this analysis as consecutive gating (i.e. evaluating CD41 and CD45 within VE-Cad+ Kit+ cells, especially if the authors think that the presence of VE-Cad+ Kit+ cells is suggestive of EHT). The quantification presented in D is misleading as the scale of each graph is different.*

Fig. 1C-D provide an overview of haemogenic markers during the timecourse of haemogenic gastruloid differentiation, and does indeed show a late up-regulation of CD45, as the Reviewer points out would be expected. The %CD45+ cells is indeed low. However, we should point out that the haemogenic gastruloid protocol, although biased towards mesodermal outputs, does not aim to achieve pure haematopoietic specification, but rather place it in its embryo-like context. Consecutive gating at the 216h-timepoint is shown and quantified in Fig.

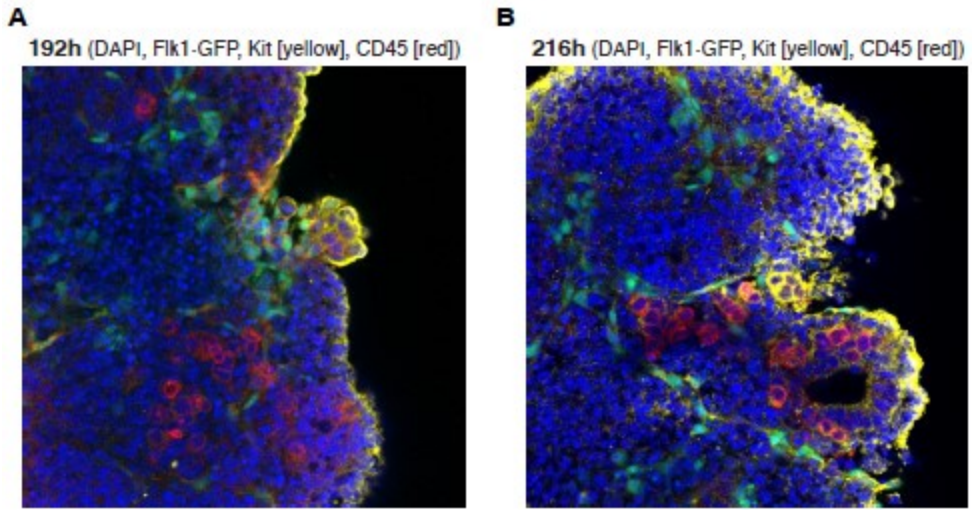
3A-B. We refute that the scale is misleading. It is a necessity to represent the data in a way that is interpretable by the reader: the gates (in C) are truly representative and annotated, as are the plot axes (in D).

> *(2) The imaging presented in Figure 1E is very unconvincing. C-Kit and CD45 signals appear as speckles and not as membrane/cell surfaces as they should. This experiment should be repeated and nuclear stain (i.e. DAPI) should be included.*

We include the requested images below (Author response image 10).

Author response image 10.

Confocal images of haematopoietic production in haemogenic gastruloids. Wholemount, cleared haemogenic gastruloids were stained for CD45 (pseudo-coloured red) and c-Kit antigens (pseudo-coloured yellow) with indirect staining, as described in the manuscript. Flk1-GFP signal is shown in green. Nuclei are contrasted with DAPI. (A) 192h. (B) 216h.

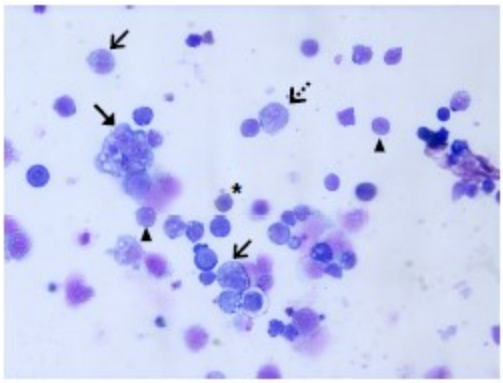


> *(3) Overall, I am not convinced that hematopoietic cells are consistently generated in these organoids. The authors should sort hematopoietic cells and perform May-Grunwald Giemsa stainings as they did in Figure 6 to confirm the nature of the blood cells generated.*

It is factual that the data are reproducible and complemented by functional assays shown in Fig. 3, which clearly demonstrate haematopoietic output. The single-cell RNA-seq data also show expression of a haematopoietic programme. Nevertheless, we include Giemsa-Wright's stained cytospins obtained at 216h to illustrate haematopoietic output (Reviewer Fig. 11). Inevitably, the cytospins will be inconclusive as to the presence of endothelial-to-haematopoietic transition or the generation of haematopoietic stem/progenitor cells, as these cells do not have a distinctive morphology.

Author response image 11.

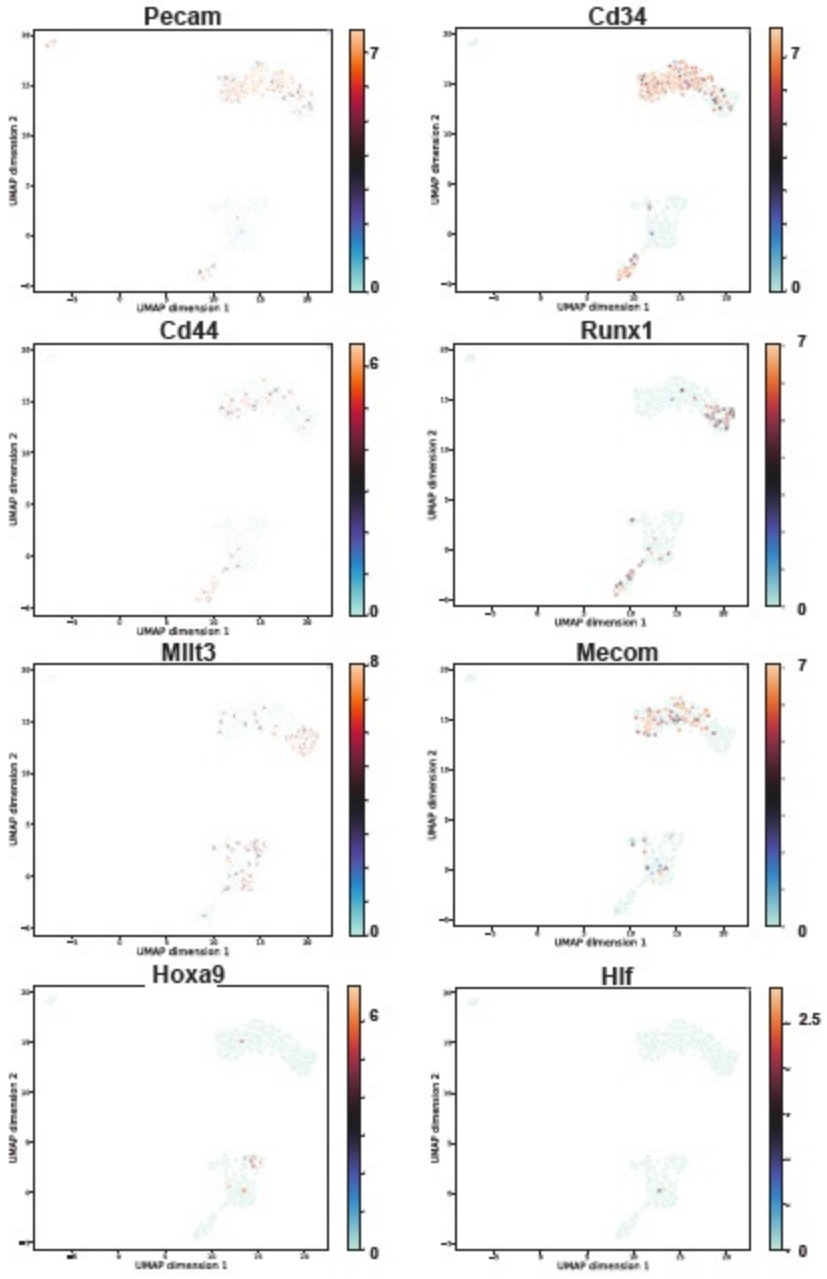
Cytospin of dissociated haemogenic gastruloids at 216h. Cytospins were stained with Giemsa-Wright's stain and are visualised with a 40x objective. Annotated are cells in the monocytic (dashed open arrow), granulocytic (solid open arrow), megakaryocytic (solid arrow) and erythroid (asterisk) lineages; arrowheads indicate cells with a non-specific blast-like morphology. Representative image.



> *(4) The scRNAseq in Figure 2 is very difficult to interpret. Specific points related to this:*
> *- Cluster annotation in Figure 2a is missing and should be included.*
> *- Why do the heatmaps show the expression of genes within sorted cells? Couldn't the authors show expression within clusters of hematopoietic cells as identified transcriptionally (which ones are they? See previous point)? Gene names are illegible.*
> *- I see no expression of Hlf or Myb in CD45+ cells (Figure 2G). Hlf is not expressed by any of the populations examined (panels E, F, G). This suggests no MPP or pre-HSC are generated in the culture, contrary to what is stated in lines 242-245. (PMID 31076455 and 34589491).*
> *Later on, it is again stated that "hGx cells... lacked detection of HSC genes like Hlf, Gfi1, or Hoxa9" (lines 281-283). To me, this is proof of the absence of AGM-like hematopoiesis generated in those gastruloids.*

Author response image 12.

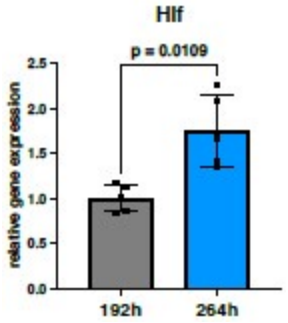
Expression of endothelial, haemogenic and haematopoietic genes in haemogenic gastruloid cells sorted at 144h, 192h and 216h. UMAP as in Author response image 4. Pecam (CD31) and CD34 represent endothelial genes also detected in haemogenic endothelium. CD44 is specifically enriched at the endothelial-to-haemogenic transition. Mecom is detected in haemogenic endothelium and haematopoietic progenitors. Mllt3 and Runx1 are haematopoietic markers. Hoxa9 and Hlf are associated with haematopoietic stem and progenitor cells and their detection is rare in haemogenic gastruloids at 216h.



For a combination of logistic and technical reasons, we performed single-cell RNA-seq using the Smart-Seq2 platform, which is inherently low throughput. We overcame the issue of cell coverage by complementing whole-gastruloid transcriptional profiling at successive time-points with sorting of subpopulations of cells based on individual markers documented in Fig. 1. We clearly stated which platform was used as well as the number and type of cells profiled (Fig. S2A and lines 172-179 of the manuscript), and our approach is standard. We will review our representation of the data in a revised manuscript. Nevertheless, at this stage, we provide plots of the expression of key haematopoietic markers over UMAPs of haemogenic gastruloid timecourse (Author response image 12). We also show preliminary qRT-PCR data with increased Hlf expression upon extension of the protocol to 264h (Author response image 13), further confirming haematopoietic specification, including of candidate definitive progenitor cells, in the haemogenic gastruloid model.

Author response image 13.

Hlf expression is up-regulated in late stage haemogenic gastruloids. Quantitative RT-PCR analysis of Hlf expression in unfractionated haemogenic gastruloids cultured for 264h. From 168h onwards, haemogenic gastruloids were cultured in N2B27 in the presence of VEGF, SCF, FLT3L and TPO, all recombinant mouse cytokines, as described in the manuscript. Shown are mean±standard deviation of n=5 replicates from 2 mouse ES cell lines, respectively Flk1-GFP and Rosa26-BFP::Flk1-GFP, reported in the manuscript; 2-tailed unpaired t-test with Welch correction.



> *(5) Mapping of scRNA-Seq data onto the dataset by Thambyrajah et al. is not proof of the generation of AGM HE. The dataset they are mapping to only contains AGM cells, therefore cells do not have the option to map onto something that is not AGM. The authors should try mapping to other publicly available datasets also including YS cells.*

We have done this and the data are presented in Author response image 5 and 6. As detailed in response to Reviewer 1, we have conducted projections of our single-cell RNA-seq data against two studies which (1) capture arterial and haemogenic specification in the para-splanchnopleura (pSP) and AGM region between E8.0 and E11 (Hou et al, PMID: 32203131) (Author response image 5), and (2) uniquely capture YS, AGM and FL progenitors and the AGM endothelial-to-haematopoietic transition (EHT) in the same scRNA-seq dataset (Zhu et al, PMID: 32392346) (Author response image 6). Specifically in answering the Reviewers' point, we show that different subsets of haemogenic gastruloid cells sorted on haemogenic surface markers c-Kit, CD41 and CD45 cluster onto pre-HE and HE, intra-aortic clusters and FL progenitor compartments, and to YS EMP and erythroid and myeloid progenitors. This lends support to our claim that the haemogenic gastruloid system specifies both YS-like and AGM-like cells.

> *(6) Conclusions in Figure 3, named "hGx specify cells with preHSC characteristics" are not supported by the data presented here. Again, I am not convinced that hematopoietic cells can be efficiently generated in this system, and certainly not HSCs or pre-HSCs.*

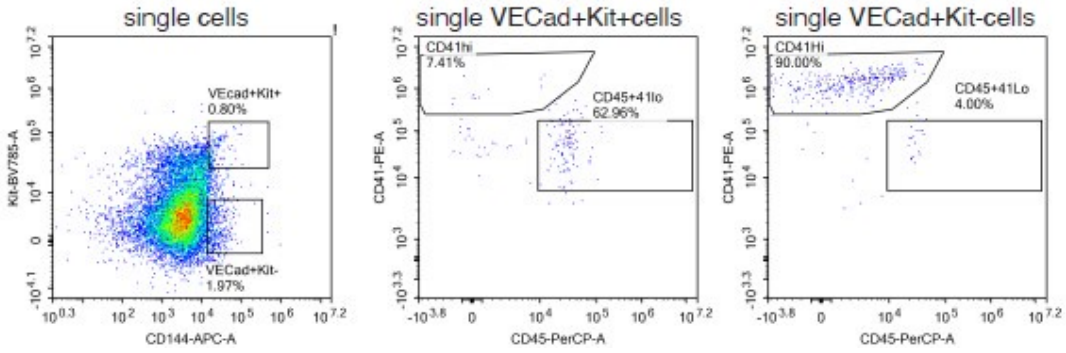
We have provided evidence, both in the manuscript and in this response to Reviewers, that there is haematopoietic specification, including of progenitor cells, in the haemogenic gastruloid system (Fig. 3 and Author response image 7,9). We have added data in this response that supports the specification of YS-like and AGM-like cells (Author response image 5-6, 8). Importantly, we have never claimed that haemogenic gastruloids generate HSC. We accept the Reviewer's comment that we have not provided sufficient evidence for the specification of pre-HSC-like cells. We will re-phrase Fig. 3 conclusion as "Haemogenic gastruloids specify cells with characteristics of definitive haematopoietic progenitors".

> *- FACS analysis in 3A is again very unconvincing. I do not think the population identified as c-Kit+ CD144+ is real. Also, why not try gating the other way around, as commonly done (e.g. VE-Cad+ Kit+ and then CD41/CD45)?*

There is nothing unconventional about our gating strategy, which was done from a more populated gate onto the less abundant one to ensure that the results are numerically more robust. In the case of haemogenic gastruloids, unlike the AGM preparations the Reviewer may be referring to, CD41 and CD45+ cells are more abundant as there is no circulation of more differentiated haematopoietic cells away from the endothelial structures. This said, we did perform the gating as suggested (Author response image 14), indeed confirming that most VE-cad+ Kit+ cells are CD45+. Interestingly VE-cad+Kit- are predominantly CD41+, reinforcing the true haematopoietic nature of these cells.

Author response image 14.

Flow cytometry analysis of VE-cadherin+ cells in haemogenic gastruloids at 216h of the differentiation protocol, probing co-expression of CD45, CD41 and c-Kit.



> *- The authors must have tried really hard, but the lack of short- or long-engraftment in a number of immunodeficient mouse models (lines 305-313) really suggests that no blood progenitors are generated in their system. I am not familiar with the adrenal gland transplant system, but it seems like a very non-physiological system for trying to assess the maturation of putative pre-HSCs. The data supporting the engraftment of these mice, essentially seen only by PCR and in some cases with a very low threshold for detection, are very weak, and again unconvincing. It is stated that "BFP engraftment of the Spl and BM by flow cytometry was very low level albeit consistently above control (Fig. S4E)" (lines 337-338). I do not think that two dots in a dot plot can be presented as evidence of engraftment.*

We have presented the data with full disclosure and do not deny that the engraftment achieved is low-level and short-term, indicating incomplete maturation of definitive haematopoietic progenitors in the current haemogenic gastruloid system. However, we call the Reviewer's attention to the fact that detection of BFP+ cells by PCR and flow cytometry in the recipient animals at 4 weeks is consistent between the 2 methods (Author response image 7).

Furthermore, we have now also been able to detect low-level myelo-lymphoid engraftment in the bone marrow 8 weeks after adrenal implantation, again suggesting the presence of a small number of definitive haematopoietic progenitors that potentially mature from the 3 haemogenic gastruloids implanted (Author response image 9).

> *(7) Given the above, I find that the foundations needed for extracting meaningful data from the system when perturbed are very shaky at best. Nevertheless, the authors proceed to overexpress MNX1 by LV transduction, a system previously shown to transform fetal liver cells, mimicking the effect of the t(7;12) AML-associated translocation. Comments on this section:*
> *- The increase in the size of the organoid when MNX1 is expressed is a very unspecific finding and not necessarily an indication of any hematopoietic effect of MNX1 OE.*

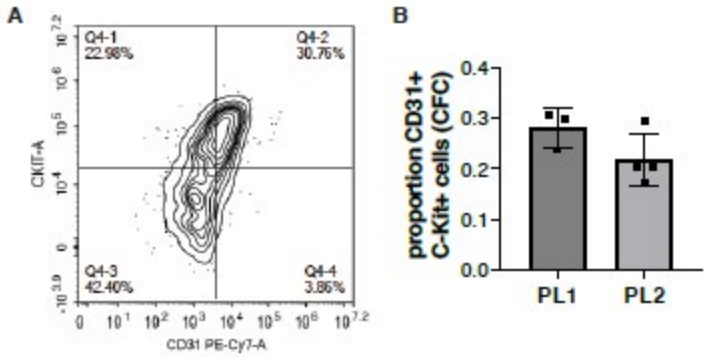
We agree with the Reviewer on this point; it is nevertheless a reproducible observation which we thought relevant to describe for completeness and data reproducibility.

> *- The mild increase of cKit+ cells (Figure 4E) at the 144hr timepoint and the lack of any changes in CD41+ or CD45+ cells suggests that the increase in Kit+ cells % is not due to any hematopoietic effect of MNX1 OE. No hematopoietic GO categories are seen in RNA seq analysis, which supports this interpretation. Could it be that just endothelial cells are being generated?*

The Reviewer is correct that the MNX1-overexpressing cells have a strong endothelial signature, which is present in the patients (Fig. 4A). We investigated a potential link with c-Kit by staining cells from the replating colonies during the process of in vitro transformation with CD31. We observed that 40-50% of c-Kit+ cells (20-30% total colony cells) co-expressed CD31(Author response image 15), at least at early plating. These cells co-exist with haematopoietic cells, namely Ter119+ cells, as expected from the YS-like erythroid and EMP-like affiliation of haematopoietic output from 144h-haemogenic gastruloids (Fig. 5F).

Author response image 15.

Endothelial affiliation of MNX1-oe replating cells from haemogenic gastruloid. A. Representative flow cytometry plot of plate 1 CFC from MNX1-overexpressing haemogenic gastruloids at 144h. B. Quantification of the proportion of CD31+c-Kit+ cells in plates 1 and 2 of MNX1-oe-driven in vitro transformation.



> *(8) There seems to be a relatively convincing increase in replating potential upon MNX1-OE, but this experiment has been poorly characterized. What type of colonies are generated? What exactly is the "proportion of colony forming cells" in Figures 5B-D? The colony increase is accompanied by an increase in Kit+ cells; however, the flow cytometry analysis has not been quantified.*

Given the inability to replate control EV cells, there is not a population to compare with in terms of quantification. The level of c-Kit+ represented in Fig. 5E is achieved at plate 2 or 3

(depending on the experiment), both of which are significantly enriched for colony-forming cells relative to control (Fig. 5B, D).

> *(9) Do hGx cells engraft upon MNX1-OE? This experiment, which appears not to have been performed, is essential to conclude that leukemic transformation has occurred.*

For the purpose of this study, we are satisfied with confirmation of in vitro transformation potential of MNX1 haemogenic gastruloids, which can be used for screening purposes. Although interesting, in vivo leukaemia engraftment from haemogenic gastruloids is beyond the scope of this study.

https://doi.org/10.7554/eLife.102324.1.sa0