SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

University of Milano-Bicocca

Department of Medicine and Surgery

PhD program in Translational and Molecular Medicine

XXXVIII Cycle

Dissecting pre-natal hematopoiesis in health and disease: from the emergence of embryonic stem and progenitor cells to the origins of pediatric leukemia

Doctoral candidate: Giulia Quattrini

Registration number: 896542

Tutor: Prof. Emanuele Azzoni

Supervisor: Cristiana Barone, PhD

Coordinator: Prof. Francesco Mantegazza

ACADEMIC YEAR 2024/2025

Table of contents

Abstract

Embryonic hematopoiesis is a complex process consisting in the emergence of three consecutive and partially overlapping waves of hematopoietic cells, which culminates with the generation of hematopoietic stem cells (HSCs). However, several aspects of this process are still incompletely understood.

In this context, this work initially focused on investigating normal embryonic hematopoiesis. By combining lineage-tracing strategies, functional assays, and single-cell transcriptomics in mice, we characterized the heterogeneity of the hemogenic endothelium (HE), a specialized endothelial population that gives rise to all embryonic hematopoietic stem and progenitor cells (HSPCs).

Through genetic fate mapping strategies that allow labeling and tracking of distinct subsets of fetal HSPCs, we characterized the contribution of different subsets of fetal HSPCs to embryonic hematopoiesis. Notably, we were able to capture a wave of fetal-restricted hematopoietic progenitors emerging from a subset of HE localized in the vitelline and umbilical arteries between embryonic day (E)8.5 and E9.5, which represents a predominant contributor to fetal lympho-myelopoiesis.

This knowledge served us as a platform to generate a novel $Kras^{G12D}$-driven mouse model of juvenile myelomonocytic leukemia (JMML) with prenatal onset. JMML is a RAS-driven pediatric myeloproliferative disease that often originates before birth. With current models, investigating the *in utero* origins of the disease is particularly challenging. Thus, we exploited our newly generated model to tackle this understudied aspect of JMML. Our model resembles the main features of the human disease. Our results demonstrated that JMML can originate not only from fetal HSCs but also from HSC-independent progenitors, which show differential susceptibility to $Kras^{G12D}$-induced leukemic transformation. We were able to identify a *in utero* pre-leukemic stage of the disease and conclusively demonstrate that mutation acquisition in embryonic HSPCs can result in JMML disease development after birth. Using a different model, we also showed that HSC-independent progenitors can undergo leukemic transformation upon acquisition of the Mll-Af9 translocation.

In summary, by characterizing the identity and the contribution of fetal HSPCs to fetal hematopoiesis, and studying their susceptibility to leukemic transformation, this work highlights how gaining deeper insights in basic developmental research is essential for translational advance in pediatric leukemia.

Introduction

Embryonic hematopoiesis

Embryonic hematopoiesis is the complex and dynamic process that ultimately gives rise to all blood and immune cell lineages in the embryo. During life, hematopoiesis occurs in three major phases, embryonic, fetal, and adult, each characterized by distinct sites of origin. During embryogenesis, hematopoietic progenitors arise in the yolk sac (YS) and in the aorta–gonad–mesonephros (AGM) region. All these progenitors transiently migrate to the fetal liver (FL), where they proliferate, and at late stage of gestation, they migrate to the bone marrow (BM), where they will maintain adult hematopoiesis (**Figure 1**).

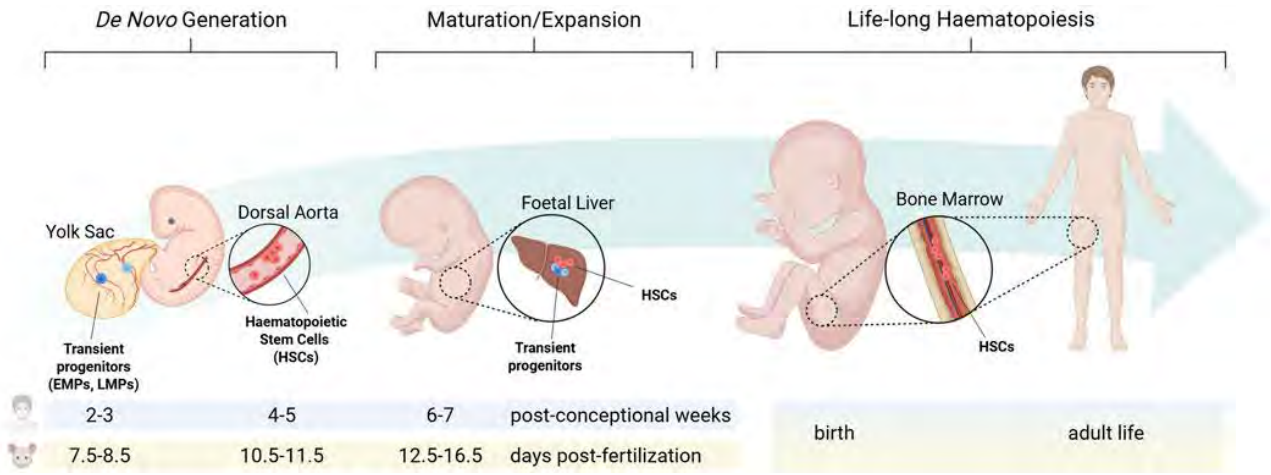


***Figure 1**: Hematopoietic system development from embryonic development to adult (Sigimura et al., 2025)*

Although less well studied compared to adult hematopoiesis, recent technological advances, including single-cell RNA sequencing, multi-omics approaches, high-resolution imaging and lineage tracing strategies have deepened the understanding and the detailed knowledge of this process, allowing the identification of different intermediate cell populations and clarified the temporal and spatial overlap between hematopoietic waves. A comprehensive understanding of embryonic hematopoiesis is essential for translational medicine because of multiple reasons. Understanding how blood progenitors develop in the embryo can inform the techniques aimed at the generation of these cells from pluripotent stem cells, for therapeutic purposes. Moreover, alterations in this process can lead to disease in fetal, pediatric and adult life.

In mammalians, embryonic hematopoiesis is based on the emergence of three consecutive and partially overlapping waves of progenitors (**Figure 2**), characterized by increasing lineage potential[1].

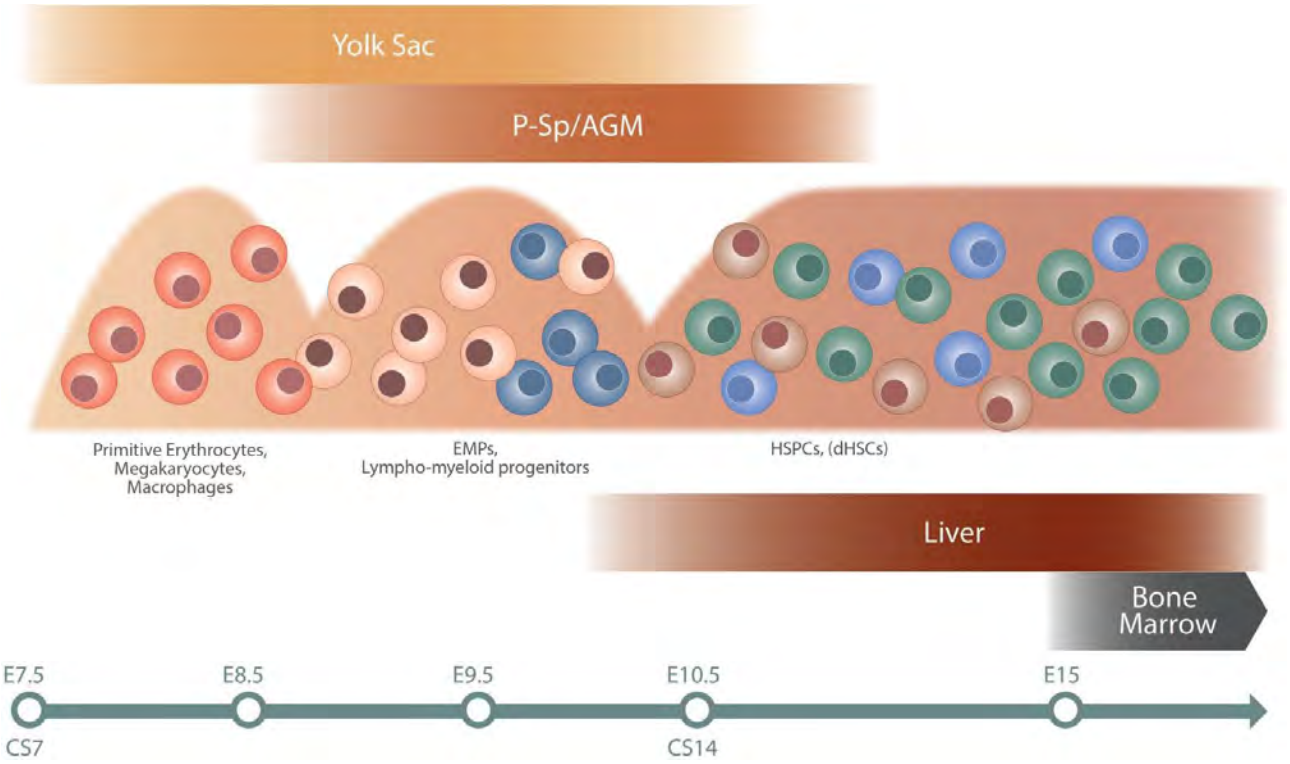


Figure 2: Schematic overview of the three embryonic hematopoietic waves (Karlsson et al., 2021)

The first hematopoietic wave, or primitive wave, arises in the extra-embryonic YS at approximately embryonic day (E) 7.0 in the mouse, equivalent to 2–3 weeks post-conception in humans. Here, extra-embryonic mesoderm differentiates into haemato-vascular progenitors which organize in blood islands that generate nucleated erythrocytes, megakaryocytes, and macrophages. Primitive erythrocytes (EryP) are large, nucleated cells expressing embryonic globin, and originate from EryP-CFC progenitors, arising from E7.25 and reaching their peak at E8.25[2]. They mature in circulation between E12.5 and E16.5, undergoing enucleation and hemoglobinization, and they are necessary for embryo oxygen diffusion. They then disappear at perinatal stage[2,3]. Despite the debate about their origin, recent studies demonstrate a possible derivation from hemogenic endothelium[4]. Concomitantly, primitive megakaryocytes (E7) and primitive macrophages (E7.25) also originate from the YS[3]. Primitive megakaryocytes, around E9, colonize all tissues, particularly the brain, giving rise to the microglia, the only long-lived population of extra-embryonic origin[5].

The second hematopoietic wave, or pro-definitive wave, initially originates from the yolk sac, around E8.25, corresponding to 3-4 weeks post conception in humans. During this phase, a highly specialized

endothelium type is formed, called hemogenic endothelium (HE), from which clusters of multipotent hematopoietic progenitors emerge, by undergoing endothelial to hematopoietic transition (EHT)[6]. They are mainly composed of two different populations of progenitors, erythro-myeloid progenitors (EMPs) and lympho-myeloid progenitors (LMPs)[3]. In mouse, EMPs emerge around E8.25 and LMPs at E9.5. EMPs give rise to tissue-resident macrophages that persist throughout adult life and make significant contributions to fetal erythropoiesis[7,8] and innate lymphoid cells. However, the extent of their postnatal persistence is still debated. LMPs transiently contribute to lympho-myelopoiesis during fetal development, giving rise to early B and T cells before the emergence of hematopoietic stem cells (HSCs)[1]. As the circulatory system is established (at E8.25 in the mouse), both EMPs and LMPs migrate through the bloodstream to the FL, where they sustain hematopoiesis until birth. So, FL represents the primary hematopoietic organ from mid-gestation.

Recently we demonstrated the emergence of a wave of fetal-restricted hematopoietic stem and progenitor cells (HSPCs), emerging between E8.5 and E9.5 from HE, in particular form the hematopoietic clusters of the vitelline and umbilical arteries[9]. They are functionally distinct from definitive MPPs and HSCs and lack long-term engraftment potential, but they represent the main contributor to fetal lympho-myelopoiesis (See Barone et al., *Biorxiv*, 2024, under revision; included in this thesis)[9].

The third hematopoietic wave, or definitive wave, originates intra embryonically, specifically in the AGM from E10.5 in the mouse, corresponding to 4-5 weeks in humans. In this phase HE undergoes EHT to generate HSCs[10,11]. HSCs are capable of self-renewal and multilineage differentiation throughout life. Their emergence depends on a complex AGM niche, integrating biomechanical forces such as blood flow-induced shear stress and signaling pathways including Notch, Wnt, BMP, and inflammatory cytokines[12]. These cells progress through hierarchies of intermediate states of pre-HSCs and pro-HSCs before acquiring full self-renewal and multilineage potential[13,14]. From E12, HSCs migrate and colonize the fetal liver, where they mature and expand, and finally migrate to the bone marrow (BM) at the end of gestation, which becomes the definitive site of hematopoiesis throughout postnatal life. The transition from fetal to adult HSC identity occurs postnatally, around three weeks after birth, marked by type I interferon gene activation and gradual acquisition of adult-like transcriptional and functional signatures[15].

The mechanism regulating the HSPCs fate choice upon EHT are still incompletely understood. Recent studies demonstrate that HE can give rise to different subsets of specialized progenitors, each associated with a distinct molecular and transcriptional pattern. In particular, each population expresses unique chromatin modifiers and spliceosome components, correlating with distinct isoform

expression patterns, including multiple stemness associated factors, which may represent one of the mechanisms regulating HSPCs fate choice (Neo, Fadlullah et al., *Nature Cardiovascular Research*, in press; included in this thesis).

In summary, embryonic hematopoiesis is a crucial step of development, and it represents a complex and highly regulated process. In the last decades, the knowledge about the underlying mechanism has significantly expanded, and ongoing studies continue to refine this knowledge. The understanding of embryonic hematopoiesis is crucial for gaining deeper insight in the developmental origins of several types of childhood leukemias and bone marrow failure, and for driving HSCs regenerative strategies for clinical use, also involving transplantation and gene therapy.

Pediatric hematopoietic malignancies

Hematologic malignancies represent the most common cancers in childhood, accounting for approximately one third of all pediatric malignancies[16]. They are characterized by highly heterogenous biological and clinical manifestations. Compared to adult hematological disorders, their pediatric counterparts were historically less explored. An increasing number of studies have recently focused on their pathogenesis and they have ultimately led to an improvement of their clinical outcomes. However, the relatively high incidence rates among pediatric malignancies underline the need to gain deeper insight into their molecular landscape, in order to develop specific therapies.

The cellular origins of pediatric leukemias are generally still incompletely characterized, but, given the complexity of embryonic hematopoiesis, they are likely distinct from the adult disease. Accumulating evidence suggests that many cases of pediatric hematopoietic malignancies have a prenatal origin, with chromosomal translocations or mutations occurring *in utero*. In fact, genetic lesions have been identified in neonatal blood spots, before clinical manifestation of the disease[17]. As embryonic hematopoiesis is sustained by different populations of HSC-dependent and -independent progenitors, it is possible to hypothesize that, differently from adult leukemias, the cellular origin of pediatric leukemia could be related also to HSC-independent progenitors. Moreover, during fetal development, hematopoietic stem cells display high proliferative potential and a degree of epigenetic plasticity, therefore they can be particularly susceptible to leukemic transformation.

Pediatric hematologic malignancies represent a heterogeneous group of diseases, which include acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), as well as myelodysplastic

syndromes (MDS) and myeloproliferative neoplasms (MPN). ALL is the most common leukemia in children, and it derives from the malignant transformation of B- and T- lymphoid progenitor cells. Clinically, it typically presents with bone marrow failure and cytopenia, while genetically is highly heterogeneous[18]. AML results from the clonal proliferation of myeloid precursors with impaired differentiation, leading to accumulation of immature myeloid blasts[19]. Notably, rearrangements that involve the histone lysine methyltransferase 2A, *MLL*, are found in most of the pediatric acute leukemias, with evidence of its prenatal origin[20,21]. Chronic myeloid leukemia (CML), rare in children, is characterized by the presence of the BCR::ABL1 fusion gene and usually follows a more indolent chronic course[22]. Pediatric MDS is a rare group of disorders, characterized by ineffective hematopoiesis, morphologic dysplasia, and risk for progression to acute myeloid leukemia[23]. MPN are characterized by excessive proliferation of one or more mature myeloid cell lineages. In addition, some pediatric malignancies exhibit intermediate features, representing overlapping disease between MDS and MPN[24]. Among these, the most common MDS/MPN overlapping disease in children is juvenile myelomonocytic leukemia (JMML).

Juvenile myelomonocytic leukemia (JMML)

Juvenile myelomonocytic leukemia (JMML) is a myelodysplastic/myeloproliferative neoplasm (MDS/MPN) overlap syndrome, affecting children with a median age at diagnosis of 2 years[25]. JMML account for 1% of all pediatric leukemias, with an incidence rate of 1.2 cases per million/year[26]. JMML is a rare and heterogeneous disease, characterized by the hyperactivation of the RAS/MAPK signaling pathway[27,28]. Notably, nearly half of the cases of JMML have been reported to originate *in utero*[29], highlighting the complexity of the disease and the challenges in investigating its early developmental and pathogenic mechanisms.

The typical clinical features observed in patients are related with the uncontrolled proliferation of malignant myeloid cells, and include leukocytosis, monocytosis, thrombocytopenia, splenomegaly, anemia and pulmonary infiltrations by myeloid cells[25,30,31]. Formal diagnostic criteria include prominent monocytosis in peripheral blood ($\geq 1 \times 10^9$/L) low frequency of blasts in peripheral blood and bone marrow (<20%), splenomegaly and absence of Philadelphia chromosome (BCR/ABL1 rearrangement)[24]. Additional criteria, considered common hallmarks of JMML, are the presence of elevated fetal hemoglobin normalized by age and Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF) hypersensitivity in vitro[24].

The prognosis is generally poor and currently the therapeutic options are scarce, with allogenic stem cell transplantation (HSCT) being the only effective curative option[31,32]. However, after HSCT relapse occurs frequently, with rates reaching around 50% of cases[31]. Azacitidine, an hypomethylating agent, demonstrated positive results in stabilizing the disease and serving as a bridge to HSCT, and it is now being tested for stand-alone treatment[33]. Promising pre-clinical studies evaluating Chimeric Antigen Receptor-T (CAR-T) cells therapy potential are ongoing[34].

The vast majority of JMML cases (around 95%) is caused by a mutation in a gene of the RAS pathway, including PTPN11, KRAS, NRAS, CBL and NF1[27]. These mutations are usually mutually exclusive.

Under normal conditions, RAS pathway regulates several cellular processes, such as cell proliferation, survival, differentiation and cell migration, through the activation of a phosphorylation signaling cascade (**Figure 3**).

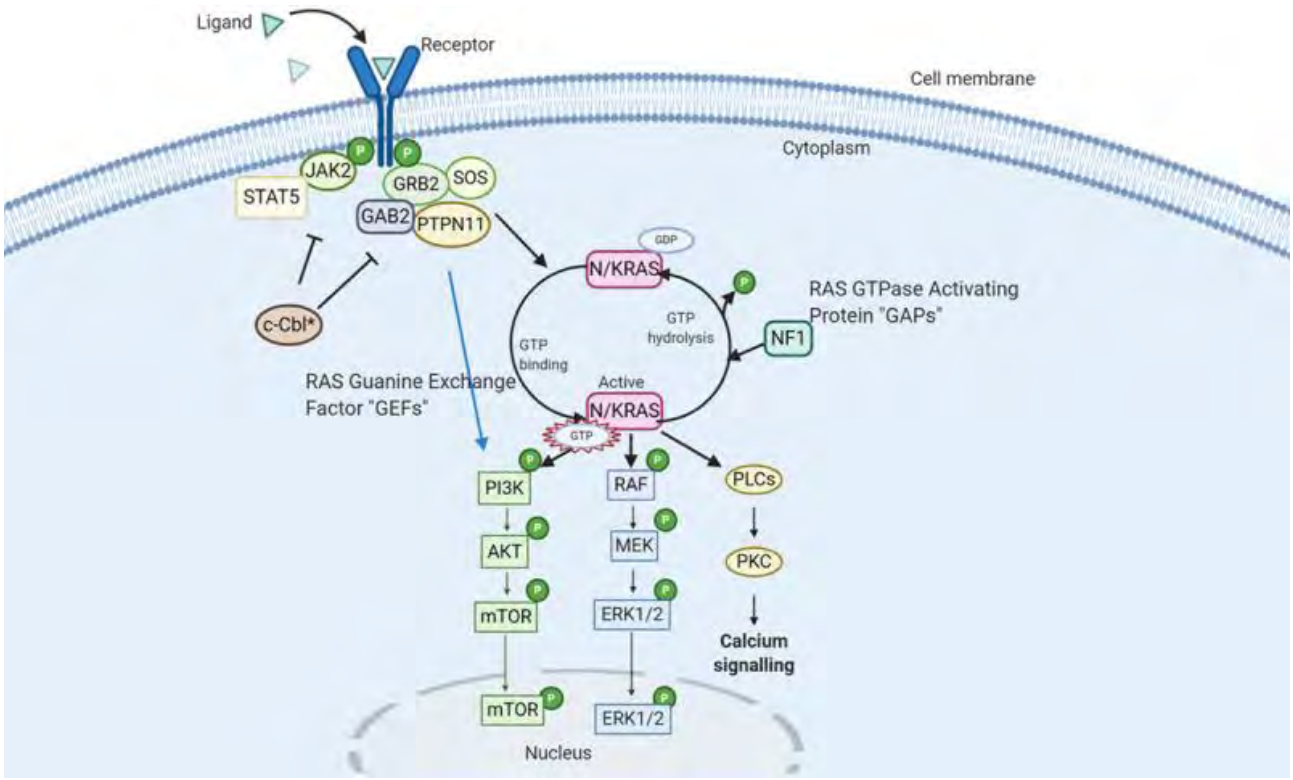


Figure 3: RAS signaling pathway (Gupta et al., 2021)

NRAS and KRAS are small GTPase proteins that serve as molecular switches downstream of receptor and non-receptor tyrosine kinases (RTKs and TKs). RAS activity is regulated by a dual system, in which phosphorylation and dephosphorylation are controlled by the opposite actions of guanine nucleotide-exchange factors (GEFs) and GTPase-activating proteins (GAPs). Activation of RTKs or TKs induces the recruitment of adaptor proteins, including GRB2, GAB2, and SHP2 (encoded by the

PTPN11 gene), as well as GEFs, which facilitate the exchange of GDP for GTP on RAS, converting it into its active form. Once activated, RAS triggers a signaling cascade that sequentially engages RAF and phosphorylates MEK and ERK, ultimately conveying signals to the nucleus. The inactivation of the RAS pathway is mediated by GAPs, including NF1, which promote the hydrolysis of RAS-GTP back to the inactive RAS-GDP form. Furthermore, the ubiquitin ligase CBL can serve as an additional negative regulator by targeting activated RTKs for proteasomal degradation, thereby dampening RAS signaling.

Given its central role in cellular signaling, mutations affecting the RAS pathway can severely affect normal cellular functions and are frequently implicated in malignant transformation.

PTPN11 somatic mutations are the most common among JMML patients, and are found in up to 30% of patients[35]. PTPN11-mutated JMML is typically associated with an aggressive course and poor prognosis. The PTPN11 gene encodes for the SHP2 protein, a critical positive regulator of the pathway. JMML-associated PTPN11 variants (e.g. E76K) are gain of function mutations conferring a stronger effect on phosphatase activity and therefore a constitutive activation of the downstream target of the RAS pathway[36]. Interestingly, PTPN11 germline variants can also be associated with Noonan Syndrome (NS), a rare developmental disorder characterized dysmorphism, a growth delay, and cardiovascular defects. Children with NS often develop a myeloproliferative disorder clinically identical to JMML, but it is usually milder and spontaneously regressing[37].

NRAS and KRAS somatic mutations are each detected in about 15% of JMML cases[35]. JMML-associated mutations in KRAS and NRAS impair GAP-mediated inactivation, stabilize the active GTP-bound state and also the nucleotide exchange rate[38]. Germline mutations in these genes have been reported in different RASopathies, including NS. However, similar to PTPN11, they are usually distinct from those reported in JMML, and only a minority of KRAS- or NRAS-driven RASophaties have been linked to JMML.

CBL mutations account for 10% of patients and are usually missense mutations, leading to a loss of function[35]. This causes a dysregulation of the conversion of RAS-GDP to RAS-GDP and activation of the RAS/RAF/MEK/ERK pathway. Germline mutations in CBL are associated with CBL syndrome, a disease characterized by neurologic, vasculitis, Noonan syndrome-like features and an increased risk of JMML. CBL-driven JMML usually have an indolent course, with most of them resolving spontaneously[39].

NF1 mutations are detected in about 10% of patients[35], and most of them are frameshift mutations resulting in a loss of function mutation. This causes a reduced dephosphorylation of RAS-GTP

activated proteins, resulting in a constant activation of the RAS signaling pathway[40]. Germline mutations are associated with neurofibromatosis type 1, a congenital disorder characterized by the presence of café-au-lait macules, skinfold freckling, development of tumors of the nervous system, and overlapping features with other RASopathies. These patients have a high predisposition to develop JMML[25].

These mutations are typically mutually exclusive, demonstrating that a single one of these mutations can be sufficient to drive JMML development. However, in a minority of cases, secondary mutations can be detected. Around 10% of JMML cases show co existing mutations in RAS pathway, with PTPN11 and NF1 being the most common[35]. Moreover, in 30% of cases secondary mutations outside the RAS pathway have been identified. Among these, mutation in SETBP1 and JAK3 are the most common and usually correlate with a rapidly progressing course and an overall worse prognosis[41–43].

Additionally, epigenetic alterations in JMML have been identified and they have been shown to correlate with disease course and prognosis. Moreover, accumulating evidence show that different methylation profiles strictly correlate with survival rate, suggesting their potential as diagnostic and prognostic biomarkers for clinical decision-making[35,44,45].

Despite advances in understanding genetic and epigenetic drivers of JMML, many aspects of its biology remain poorly understood. Interestingly, recent studies highlighted the presence of marked inflammatory gene expression signature in patients-derived HSPCs, suggesting a possible role in disease initiation and progression[46]. Preclinical studies also demonstrated that inflammation can sustain disease development but also impair normal hematopoiesis, further supporting the central role of inflammation in JMML[47–49], which may represent a promising therapeutic target.

A comprehensive understanding of developmental hematopoiesis is essential to fully comprehend the biology of pediatric leukemias and to uncover its developmental origins. In this context, this work was initially focused on dissecting embryonic hematopoiesis, with the aim of gaining a deeper insight into the cellular hierarchies and the complex molecular programs that regulate the emergence of different subsets of HSPCs. By integrating lineage tracing strategies, phenotypic and functional assays and single-cell transcriptomics in mice, our studies have provided novel insights into the heterogeneity of HE and the complex regulation of HSC and progenitor fate decision. In one of our studies, we investigated the contribution of different subpopulations of HE to fetal and postnatal hematopoiesis. Importantly, we discovered that fetal-restricted hematopoietic stem/progenitor cells, which we identified as the main contributor to fetal lympho-myelopoiesis, originate from a distinct subset of HE in the vitelline and umbilical arteries. This knowledge and the lineage tracing models generated in this study served us as the basis for investigating the prenatal origins of JMML. We

generated a novel $Kras^{G12D}$-driven transgenic murine model of JMML with prenatal origin, which recapitulated the main features of the human disease . We showed that the cellular origin of JMML is not restricted to the HSCs compartment, and that distinct subpopulations of embryonic HSPCs exhibit differential susceptibility to $Kras^{G12D}$. Moreover, through a combination of functional and single-cell transcriptomic analyses, we identified a *in utero* pre-leukemic stage of the disease. Notably, in our JMML model we identified the presence of enhanced inflammation, which may represent a potential therapeutic vulnerability.

Overall, our work highlights the critical role of basic developmental research for driving translational advances in pediatric leukemia and potentially identifying new therapeutic approaches.

References

1. Barone, C., Orsenigo, R., Meneveri, R., Brunelli, S. & Azzoni, E. One Size Does Not Fit All: Heterogeneity in Developmental Hematopoiesis. *Cells* **11**, 1061 (2022).

2. Palis, J., Malik, J., McGrath, K. E. & Kingsley, P. D. Primitive erythropoiesis in the mammalian embryo. *Int J Dev Biol* **54**, 1011–1018 (2010).

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

4. Garcia-Alegria, E. *et al.* Early Human Hemogenic Endothelium Generates Primitive and Definitive Hematopoiesis In Vitro. *Stem Cell Reports* **11**, 1061–1074 (2018).

5. Ovchinnikov, D. A. Macrophages in the embryo and beyond: much more than just giant phagocytes. *Genesis* **46**, 447–462 (2008).

6. Ottersbach, K. Endothelial-to-haematopoietic transition: an update on the process of making blood. *Biochem Soc Trans* **47**, 591 (2019).

7. Perdiguero, E. G. & Geissmann, F. Development and maintainance of resident macrophages. *Nat Immunol* **17**, 2 (2016).

8. Gomez Perdiguero, E. *et al.* Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547–551 (2015).

9. Barone, C. *et al.* Hemogenic endothelium of the vitelline and umbilical arteries is the major contributor to mouse fetal lympho-myelopoiesis. *bioRxiv* 2024.07.11.603050 (2024) doi:10.1101/2024.07.11.603050.

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

11. De Bruijn, M. F. T. R., Speck, N. A., Peeters, M. C. E. & Dzierzak, E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J* **19**, 2465–2474 (2000).

12. Sánchez-Lanzas, R., Jiménez-Pompa, A. & Ganuza, M. The evolving hematopoietic niche during development. *Front Mol Biosci* **11**, 1488199 (2024).

13. Taoudi, S. *et al.* Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-Cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell* **3**, 99–108 (2008).

14. Rybtsov, S. *et al.* Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43- embryonic precursor. *Stem Cell Reports* **3**, 489–501 (2014).

15. Li, Y. *et al.* Single-Cell Analysis of Neonatal HSC Ontogeny Reveals Gradual and Uncoordinated Transcriptional Reprogramming that Begins before Birth. *Cell Stem Cell* **27**, 732-747.e7 (2020).

16. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA Cancer J Clin* **71**, 7–33 (2021).

17. Greaves, M. In utero origins of childhood leukaemia. *Early Hum Dev* **81**, 123–129 (2005).

18. Onciu, M. Acute Lymphoblastic Leukemia. *Hematol Oncol Clin North Am* **23**, 655–674 (2009).

19. Elgarten, C. W. & Aplenc, R. Pediatric acute myeloid leukemia: updates on biology, risk stratification, and therapy. *Curr Opin Pediatr* **32**, 57–66 (2020).

20. Harrison, C. J. *et al.* Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12. *J Clin Oncol* **28**, 2674–2681 (2010).

21. Gale, K. B. *et al.* Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. *Proc Natl Acad Sci U S A* **94**, 13950–13954 (1997).

22. Gotesman, M., Raheel, S. & Panosyan, E. H. Chronic Myeloid Leukemia in Children and Adolescents. *Adv Pediatr* **70**, 145–155 (2023).

23. Chisholm, K. M. & Bohling, S. D. Childhood Myelodysplastic Syndrome. *Clin Lab Med* **43**, 639–655 (2023).

24. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).

25. Chronic Myelomonocytic Leukemia in Childhood: A Retrospective Analysis of 110 Cases. *Blood* **89**, 3534–3543 (1997).

26. Chan, R. J., Cooper, T., Kratz, C. P., Weiss, B. & Loh, M. L. Juvenile myelomonocytic leukemia: A report from the 2nd International JMML Symposium. *Leuk Res* **33**, 355–362 (2009).

27. Stieglitz, E. *et al.* The genomic landscape of juvenile myelomonocytic leukemia. *Nat Genet* **47**, 1326–1333 (2015).

28. Fiñana, C., Gómez-Molina, N., Alonso-Moreno, S. & Belver, L. Genomic and Epigenomic Landscape of Juvenile Myelomonocytic Leukemia. *Cancers* vol. 14 Preprint at https://doi.org/10.3390/cancers14051335 (2022).

29. Behnert, A. *et al.* Exploring the genetic and epigenetic origins of juvenile myelomonocytic leukemia using newborn screening samples. *Leukemia* **36**, 279 (2021).

30. Niemeyer, C. M. & Flotho, C. *Review Series MYELODYSPLASTIC SYNDROMES Juvenile Myelomonocytic Leukemia: Who's the Driver at the Wheel?* (2019).

31. Locatelli, F. & Niemeyer, C. M. How I treat juvenile myelomonocytic leukemia. *Blood* **125**, 1083–1090 (2015).

32. Wintering, A., Dvorak, C. C., Stieglitz, E. & Loh, M. L. Juvenile myelomonocytic leukemia in the molecular era: a clinician's guide to diagnosis, risk stratification, and treatment. *Blood Adv* **5**, 4783–4793 (2021).

33. Niemeyer, C. M. *et al.* Response to upfront azacitidine in juvenile myelomonocytic leukemia in the AZA-JMML-001 trial. *Blood Adv* **5**, 2901–2908 (2021).

34. Werner, J. *et al.* Cellular immunotherapy targeting CLL-1 for juvenile myelomonocytic leukemia. *Nat Commun* **16**, (2025).

35. Murakami, N. *et al. Integrated Molecular Profiling of Juvenile Myelomonocytic Leukemia*. http://genomon.hgc.jp/exome/ (2018).

36. LaRochelle, J. R. *et al.* Structural and Functional Consequences of Three Cancer-Associated Mutations of the Oncogenic Phosphatase SHP2. *Biochemistry* **55**, 2269–2277 (2016).

37. Strullu, M. *et al.* Juvenile myelomonocytic leukaemia and Noonan syndrome. *J Med Genet* **51**, 689–697 (2014).

38. Hobbs, G. A., Der, C. J. & Rossman, K. L. RAS isoforms and mutations in cancer at a glance. *J Cell Sci* **129**, 1287–1292 (2016).

39. Niemeyer, C. M. *et al.* Germline CBL mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia. *Nat Genet* **42**, 641 (2010).

40. Side, L. E. *et al.* Mutations of the NF1 Gene in Children With Juvenile Myelomonocytic Leukemia Without Clinical Evidence of Neurofibromatosis, Type 1. *Blood* **92**, 267–272 (1998).

41. Sakaguchi, H. *et al.* Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. *Nat Genet* **45**, 937–941 (2013).

42. Wakamatsu, M. *et al.* Detection of subclonal SETBP1 and JAK3 mutations in juvenile myelomonocytic leukemia using droplet digital PCR. *Leukemia* **35**, 259–263 (2021).

43. Bresolin, S. *et al.* Mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia: a report from the Italian AIEOP study group. *Oncotarget* **7**, 28914–28919 (2016).

44. Lipka, D. B. *et al.* RAS-pathway mutation patterns define epigenetic subclasses in juvenile myelomonocytic leukemia. *Nat Commun* **8**, (2017).

45. Stieglitz, E. *et al.* Genome-wide DNA methylation is predictive of outcome in juvenile myelomonocytic leukemia. *Nat Commun* **8**, 1–8 (2017).

46. Solman, M. *et al.* Inflammatory response in hematopoietic stem and progenitor cells triggered by activating SHP2 mutations evokes blood defects. *Elife* **11**, (2022).

47. Solman, M. *et al.* Inflammatory response in hematopoietic stem and progenitor cells triggered by activating SHP2 mutations evokes blood defects. *Elife* **11**, (2022).

48. Yan, Y. *et al.* JMML tumor cells disrupt normal hematopoietic stem cells by imposing inflammatory stress through overproduction of IL-1β. *Blood Adv* **6**, 200–206 (2022).

49. Yu, W. M., Daino, H., Chen, J., Bunting, K. D. & Qu, C. K. Effects of a leukemia-associated gain-of-function mutation of SHP-2 phosphatase on interleukin-3 signaling. *Journal of Biological Chemistry* **281**, 5426–5434 (2006).

nature cardiovascular research

Article

https://doi.org/10.1038/s44161-025-00740-z

Single-cell profiling reveals three endothelial-to-hematopoietic transitions with divergent isoform expression landscapes

Received: 17 September 2024

Accepted: 30 September 2025

Published online: 11 November 2025

Check for updates

Wen Hao Neo[1,14], Muhammad Zaki Hidayatullah Fadlullah[1,2,14], Harshangda Bhatnagar[1,3], Cristiana Barone[4], Giulia Quattrini[4], Filipa Timóteo-Ferreira[4], Joana Carrelha[5,6], Gianluca Sala[4], Robert Sellers[7], John Weightman[8], Wolfgang Breitwieser[8], Natalia Moncaut[9], Roshana Thambyrajah[1,10], Sten Eirik W. Jacobsen[5,11,12], Mudassar Iqbal[3], Syed Murtuza Baker[3], Emanuele Azzoni[4,13], Michael Lie-A-Ling[1] ✉ & Georges Lacaud[1] ✉

Hemogenic endothelium (HE) is recognized as the origin of all definitive blood cells, including hematopoietic stem cells (HSCs); however, the mechanisms governing the hematopoietic progenitor versus HSC fate choice within the HE remain unknown. Here we combine differentiation assays with full-length single-cell transcriptome data for extra-embryonic yolk sac (YS) and intra-embryonic aorta–gonad–mesonephros (AGM) region HE populations. We identified and localized three differentiation trajectories, each containing a distinct HE subset: erythromyeloid progenitor-primed HE in the YS plexus, lymphomyeloid progenitor-primed HE in large YS arteries and hematopoietic stem and progenitor cell-primed HE in the AGM. Chromatin modifiers and spliceosome components were enriched in AGM HE. This correlated with a higher isoform complexity of the AGM HE transcriptome. Distinct AGM HE-specific isoform expression patterns were observed for a broad range of genes, including stemness-associated factors like *Runx1*. Our data form a unique resource for studying cell fate decisions in different HE populations.

A pivotal step during mammalian ontogeny is the establishment of the hematopoietic system, which unfolds in three successive, partially overlapping waves[1,2]. The first two waves takes place in the yolk sac (YS). Wave 1 generates primitive erythrocytes and macrophages (E7.5)[3,4]. Wave 2 sequentially gives rise to erythromyeloid progenitors (EMPs; E8.25)[4,5] and lymphomyeloid progenitors (LMPs; E9.5)[6,7]. The final wave, in the intra-embryonic aorta–gonad–mesonephros (AGM) region, produces hematopoietic stem and progenitor cells (HSPCs; wave-3, E10.5)[8,9]. In recent years, it has become evident that wave 2 cells not only play a role in wave 3 HSC generation but can also persist into adulthood[2,10,11]. The hematopoietic cells in wave 2 and wave 3, also known as the definitive waves, arise from a specific endothelium, called hemogenic endothelium (HE), through a process called endothelial-to-hematopoietic transition (EHT)[12–16] orchestrated by the transcription factors RUNX1 (refs. 13,17–19) and GFI1 (refs. 20,21).

A critical question in hematopoietic development is why the HE in the extra-embryonic space is skewed toward EMP and LMP generation, whereas intra-embryonic HE, primarily localized in the dorsal aorta, can efficiently give rise to HSCs. The spatiotemporal difference in emergence suggests that HE cells from distinct waves are intrinsically

A full list of affiliations appears at the end of the paper. ✉e-mail: Michael.Lie-a-ling@cruk.manchester.ac.uk; Georges.Lacaud@cruk.manchester.ac.uk

different, leading to divergent molecular dependencies. Indeed, NOTCH signaling is essential for HSC development but not for EMP generation[22,23]. Conversely, *Ezh2* is essential for the generation of functional EMPs, whereas it is dispensable for AGM HSC development[24,25].

Single-cell RNA sequencing (scRNA-seq) is ideally suited to identify intrinsic differences between rare cell populations. We previously characterized a granular full-length (Smart-seq2) single-cell transcriptomic profile of the AGM EHT trajectory, defining a HE continuum (HE^{AGM}) encompassing HE cells at various stages of commitment[26]. Here, we present the acquisition and analysis of a complementary full-length transcriptome of extra-embryonic EHT populations. We identified two distinct extra-embryonic HE populations, both residing within the $Runx1^{pos}KIT^{pos}$ endothelial population. The first, HE^{YSP}, is contained within $CD24^{neg}Vwf^{neg}$ $LYVE1^{pos}$ endothelial cells, is dominant before E9.5, has EMP potential, and is localized throughout the YS endothelial plexus. The second, HE^{YSA}, is contained within $CD24^{pos}Vwf^{pos}$ $LYVE^{neg}$ endothelial cells, is dominant after E9.5, has LMP potential, and is exclusively found in large extra-embryonic arteries.

Our data reveal both striking similarities and differences between extra- and intra-embryonic HE populations. While all HE populations share a common signature marked by the expression of the transcription factors *Gfi1* and *Mycn*, there are pronounced differences with regard to the expression of chromatin modifiers and genes involved in RNA processing. This correlates with increased isoform complexity in the HE^{AGM} transcriptome. Distinct HE^{AGM}-specific isoform expression patterns are observed across a broad range of genes, suggestive of a stochastic transcriptional environment guiding the unique HSPC cell fate choices made within the AGM. Notably, multiple stemness-associated factors, such as *Runx1*, display differential isoform expression profiles when compared to the YS HE populations.

The dataset presented here forms a comprehensive full-length scRNA-seq atlas of three distinct definitive hematopoietic EHT trajectories giving rise, respectively to EMPs, LMPs and HSPCs, which can be accessed and queried at https://shiny.cruk.manchester.ac.uk/AGM_YS_dataset_final/.

Results

Extra-embryonic HE potential resides within the KIT^{pos} population

Although HE activity was previously reported to reside within KIT^{neg} cells in the AGM[12,27], it is associated with KIT^{pos} cells during in vitro mouse embryonic stem cell differentiation[16], recapitulating YS hematopoiesis[16]. To determine whether extra-embryonic HE is mainly found within the $KIT^{neg/low}$ or KIT^{pos} endothelium (defined as $CD31^{pos}$ and hematopoietic lineage/LIN^{neg}:$CD41^{neg}CD45^{neg}TER119^{neg}$), we examined the hematopoietic potential of E9.5 and E10.5 YS endothelial cells from $Runx1b^{RFP}/Gfi1^{GFP}$ reporter mice[26,28,29]. *Runx1* and *Gfi1* expressions are robust indicators of HE identity[20,21,26,28]. YS $KIT^{neg/low}$ or KIT^{pos} FACS-sorted single endothelial cells were co-cultured on OP9 cells for 7 days to support EHT and hematopoietic expansion (Fig. 1a and Extended Data Fig. 1a). Hematopoietic potential was only detected in wells seeded with endothelial cells expressing KIT and RUNX1 (Figs. 1b; E9.5 and E10.5). No hematopoietic cells were generated from either the $KIT^{pos}Runx1^{neg}$ or KIT^{neg} endothelial populations. Robust colony formation of the $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}$Runx1b:$RFP^{pos}$ cells was only observed after maturation/EHT on OP9 cells (Fig. 1c and Extended Data Figs. 1a and 2a) indicating that this population contains true HE cells and not already committed hematopoietic cells. Together, these data establish that at E9.5 and E10.5 YS extra-embryonic HE predominantly resides within the $KIT^{pos}Runx1^{pos}CD31^{pos}LIN^{neg}$ population.

scRNA-seq profiling of the extra-embryonic EHT trajectory

To construct a comprehensive full-length Smart-seq2 scRNA-seq dataset capturing the extra-embryonic EHT process, akin to our previous AGM HE study[26], we isolated individual cells of extra-embryonic populations across E9.0, E9.5 and E10.5. These included cells from HE-enriched (FACS-HE) populations from single and double *Runx1* and *Gfi1* reporter mice, non-HE endothelial cells (FACS-ENDO), and committed EMP (FACS_EMP) and LMP (FACS-LMP) hematopoietic progenitors (Fig. 1d and Extended Data Fig. 1a,b). Overall, 960 sequenced cells (100 FACS-ENDO, 660 FACS-HE, 118 EMP and 82 LMP) passed quality control with a median of 6,553 genes detected per individual cell (Extended Data Fig. 2b and Methods).

Unsupervised hierarchical clustering separated the cells into five clusters (K1–K5), with two main dendrogram branches (Fig. 1e). The first branch (K1–K2) exhibits a strong endothelial identity, with K1 containing the majority of FACS-ENDO (Fig. 1e-f). The second branch (K3–K5) has a pronounced hematopoietic identity, with the FACS-EMP and the FACS-LMP cells localizing within K4 and K5, respectively (Fig. 1e,f). K3 also displays a strong hematopoietic profile, including expression of *Ptprc* (*CD45*) and *Myb*, and markedly reduced expression of endothelial genes (*Cdh5*, *Kdr/Flk1*, *Pecam1* and *Procr*) compared to K1–K2 (Fig. 1e,f). These data indicate that HE cells reside in K2, and that K3 consists of committed early hematopoietic progenitors.

Finally, we reclustered the above YS populations with the addition of 115 YS cells, which were sorted using the established AGM HE phenotype: $KIT^{neg}CD41^{neg}CD45^{neg}CDH5^{pos}Gfi1/^{Gfi1bpos}$ (refs. 21,26). More than 95% of these cells clustered together with FACS-ENDO cells, further confirming that within the YS, HE resides within the KIT^{pos} population (Extended Data Fig. 2c–f). Altogether, these analyses suggest that we captured the full extra-embryonic YS EHT process.

Integration of YS and AGM scRNA-seq data reveals three distinct EHT trajectories

To compare extra-embryonic with intra-embryonic EHT, we conducted a joint analysis with our previously published AGM EHT dataset[26] (Extended Data Fig. 3a). We utilized a semi-supervised clustering approach and focused on populations that retain some endothelial characteristics: extra-embryonic clusters K1–K3 and AGM $CDH5^{pos}$ clusters (Fig. 2a,b). The data integration revealed three parallel sets of EHT clusters (Fig. 2b–d) with minimal overlap between extra-embryonic and AGM-derived cells (Fig. 2a,b). We designated the three trajectories as trajectory A and B for the YS-derived cells and trajectory C for the AGM-derived cells. Overall, the integration resulted in 13 clusters (Fig. 2b), which were named based on their known identity/trajectory within the Uniform Manifold Approximation and Projection (UMAP); AGM clusters[26] (c1_arterial endothelium, c2_pre-HE, c3_HE, c4_EHT, c5_intra-aortic-hematopoietic-clusters), YS-A trajectory clusters (a1 and a2), YS-B trajectory clusters (b1, b2 and b3), YS-progenitor clusters (p1 and p2). The only cluster that demonstrated an appreciable overlap between YS and AGM-derived cells was called Mix (Fig. 2b). Cells from YS endothelial (K1) and hematopoietic (K3) clusters, respectively, contributed to b1 and p1/p2 populations. Most cells from the YS HE population (K2) contributed to two distinct pairs of clusters (a1–a2 and b2–b3), situated parallel to AGM c3_HE and c4_EHT. Another scRNA-seq profiling study annotated cells similar to c3_HE as pre-HE, and c4_EHT as HE[26,30]. To reconcile semantic differences in HE definitions across studies, we considered both c3 and c4 as a single HE entity or continuum (HE^{AGM}). Using HE^{AGM} and the coexpression of hematopoietic and endothelial genes (Fig. 2c,d), we inferred that extra-embryonic clusters a1, a2, b2 and b3, likely possess HE properties. Overall, the integration of AGM and extra-embryonic EHT datasets suggests the existence of three distinct EHT trajectories.

Differential spatiotemporal emergence of extra-embryonic EHT trajectories

To unravel the characteristics of the extra-embryonic EHT clusters, we first examined the relative prevalence of each cluster present in the FACS-HE population from E9.0 to E10.5 (Figs. 1d–f and 2e). Cells in clusters b2–b3 were more prevalent at earlier (E9.0) developmental

https://doi.org/10.1038/s44161-025-00740-z

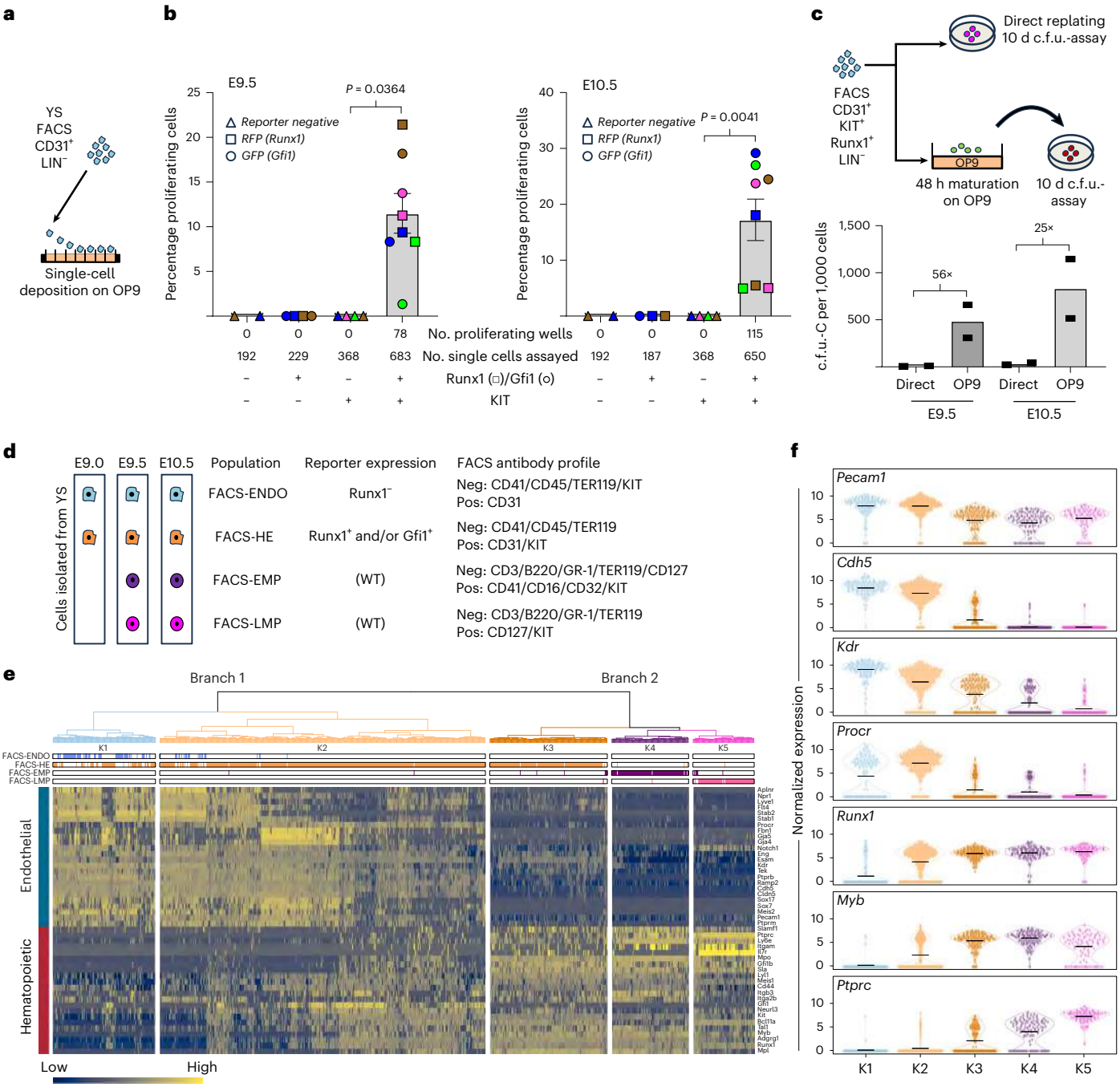


Fig. 1 | Single-cell profiling of the extra-embryonic KIT^{pos} endothelial fraction to characterize extra-embryonic EHT. a–**c**, Extra-embryonic HE potential resides within the KIT^{pos} population. Schematic of single-cell hematopoietic assays on the CD31^{pos}Lineage^{neg} (CD41^{neg}CD45^{neg}TER119^{neg}) extra-embryonic populations (**a**). Sorted KIT^{pos} and KIT^{neg} extra-embryonic single cells were co-cultured on OP9 feeder cells for 7 days. Hematopoietic activity was only observed in KIT^{pos}RUNX1^{pos} and KIT^{pos}RUNX1^{pos}GFI1^{pos} cells (**b**). No hematopoietic activity was observed in either the KIT^{pos}RUNX1^{neg} or the KIT^{neg}RUNX1^{pos} cells. Squares represent *Runx1* positive cells isolated from *Runx1*:RFP reporters, circles represent *Runx1/Gfi1* double positive cells isolated from *Runx1*:RFP/*GFi1*:GFP double reporters, triangle indicate *Runx1 / Gfi1* negative cells. Different biological experiments (for each reporter used) are indicated by color (brown, blue, green, magenta). KIT^{pos} cells were obtained from $n = 4$ biological experiments, KIT^{neg} cells were obtained from $n = 2$ biological experiments. Bars represent the average percentage of proliferating cells ± s.e.m. Statistical test was a two-tailed paired *t*-test. Error bars are not displayed for reporter KIT^{neg} samples and these samples were not tested for statistical significance. Hematopoietic colony-forming unit (c.f.u.) assay on KIT^{pos}Runx1^{pos}CD31^{pos}Lin^{neg} (CD41^{neg}CD45^{neg}TER119^{neg}) extra-embryonic cells (**c**). Cells were either directly replated or co-cultured with OP9 feeder cells for 48 h before replating. Hematopoietic colonies were quantified after 10 days. $n = 2$ biological experiments. Bars represent the average number of c.f.u. per 1000 cells seeded. Numbers above the bars represent the fold increase in hematopoietic output. **d**–**f**, Single-cell profiling of extra-embryonic EHT. Schematic of the cell populations FACS sorted from dissected E9, E9.5 and E10.5 YS and processed for full-length single-cell Smart-seq2 RNA sequencing (**d**). Endo, endothelium. Tree dendrogram generated by hierarchical clustering of the sorted populations in **c** (**e**). Two main branches are identified (K1–K2 and K3–K5). Below the dendrogram, the contribution of the different FACS-sorted populations to each cluster is shown. Bottom: heatmap depicting the expression of endothelial (top) and hematopoietic (bottom) genes across clusters K1–K5. Violin plots depicting the expression of selected endothelial (*Pecam1*, *Cdh5*, *Kdr* and *Procr*) and hematopoietic genes (*Runx1*, *Myb* and *Ptprc*) across clusters K1–K5 (**f**). Black bars represent the mean expression level.

https://doi.org/10.1038/s44161-025-00740-z

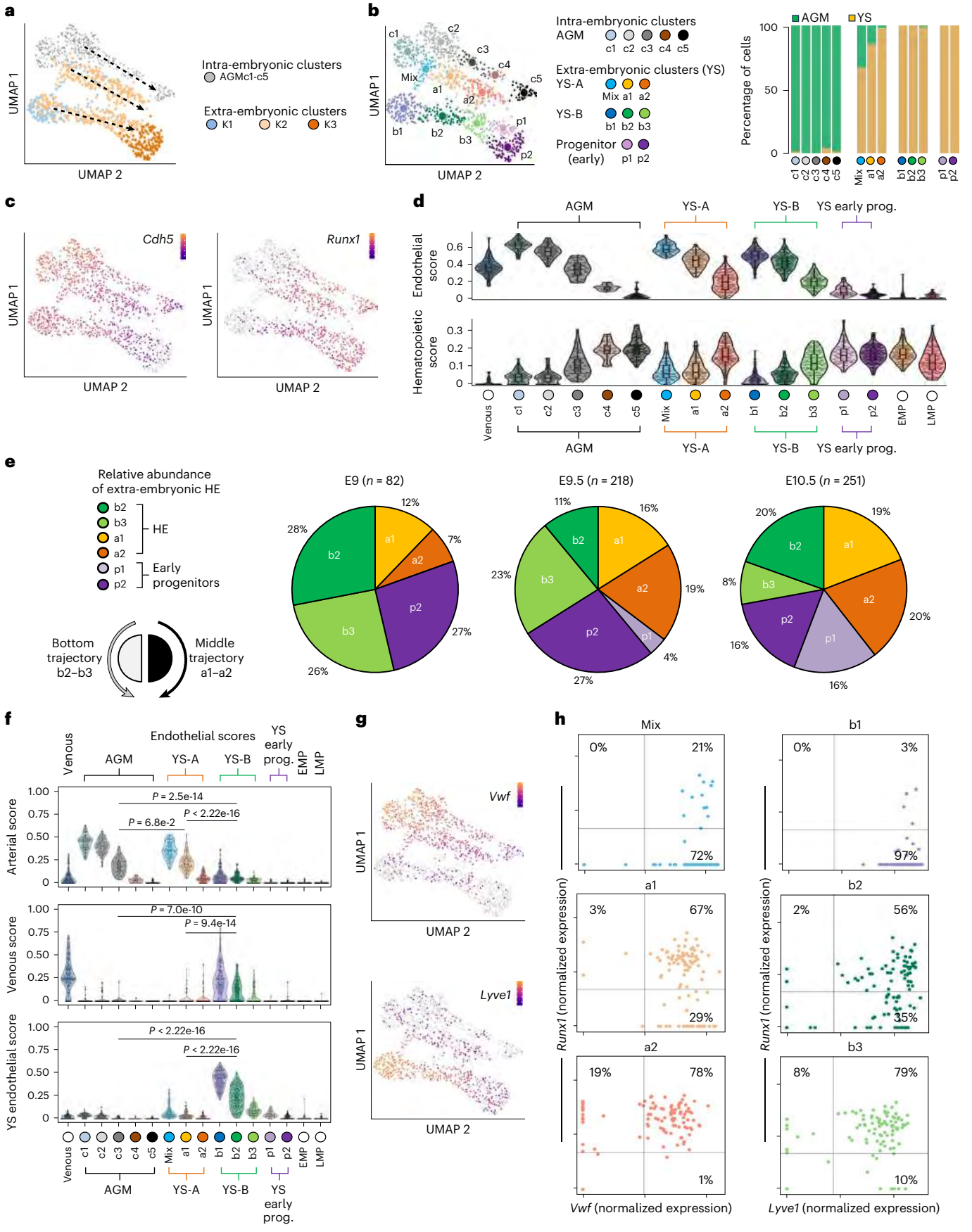


Fig. 2 | Two extra-embryonic EHT trajectories with distinct endothelial signatures. Semi-supervised clustering of intra-embryonic (AGM-derived) and extra-embryonic EHT scRNA-seq datasets. **a**, UMAP of the integrated data overlayed with the K1–K3 extra-embryonic YS clusters defined in Fig. 1d. Arrows indicate the presence of three EHT trajectories (one AGM trajectory and two YS trajectories) **b**, UMAP of the integrated data depicting the 13 clusters spread across intra-embryonic (AGM-derived) and extra-embryonic cells (YS-derived) cells (left). Clusters AGMc_3 and AGMc_4 form the intra-embryonic/AGM HE continuum (HE^{AGM}). There are two putative HE continua in the extra-embryonic space: YSc_a1, YSc_a2 and YSc_b2, YSc_b3. Graph depicting the contribution of intra-embryonic derived cells (AGM) and extra-embryonic derived cells (YS) to each cluster (right). **c**, UMAPs depicting the expression of the endothelial gene *Cdh5* and the hematopoietic gene *Runx1*. **d**, Violin plot depicting endothelial (top) and hematopoietic (bottom) signature scores across all 13 clusters defined in **b**. The signature scores were calculated using the genes depicted in Fig. 1d. Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5 × interquartile range. **e**, Relative abundance of extra-embryonic clusters YSc-b2 YSc-b3, YSc-a1 YSc-a2 (putative HE) and YSc-p1, YSc p2 (early hematopoietic progenitors). Numbers depict the percentage of the total number FACS-HE cells across all analyzed clusters at each embryonic stage. **f**, Violin plots depicting arterial, venous and YS endothelial scores across all clusters defined in **b**. For reference, AGM-derived venous endothelial cells (left column) and extra-embryonic-derived EMP and LMP populations (right columns) are also included. Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5 × interquartile range. Two-sided Wilcoxon rank-sum tests were used (with *P* values adjusted via the Benjamini–Hochberg procedure to control the FDR) to compare the early HE clusters (c3, a1 and b2). **g**, UMAPs depicting the expression of the *Vwf* (marking the AGM and YS-A clusters) and *Lyve1* (marking YS-B clusters). **h**, Correlation between transcript expression of *Runx1* and *Vwf* in clusters mix, a1 and a2 (YS-A trajectory) (top). Correlation between *Runx1* and *Lyve1* transcript expression in clusters b1, b2 and b3 (YS-B trajectory) (bottom).

stages than cells in a1–a2, suggesting that the YS-B EHT trajectory is established before the YS-A trajectory. The appearance of the two extra-embryonic progenitor clusters followed a similar sequential pattern, with p2 preceding p1.

As endothelial gene expression plays a pivotal role in defining HE identity, we evaluated whether the different trajectories could be segregated based on endothelial profiles (arterial, venous and YS) (Fig. 2f and Supplementary Table 1). The AGM EHT trajectory exhibited a robust arterial endothelial identity, similar to extra-embryonic clusters mix, a1 and a2, whereas b1, b2 and b3 displayed venous and especially YS endothelial profiles (Fig. 2f). Mix and b1 likely represent non-HE endothelial cells, as they display the strongest arterial and YS endothelial profiles within their respective trajectories while also lacking *Runx1* expression in most of the cells that make up the cluster (Fig. 2c,f and Extended Data Fig. 3b).

Finally, we screened for specific markers allowing us to determine the spatial localization, within the YS, of cells representing these different EHT trajectories. Differential gene expression analysis identified the endothelial genes *Vwf* (Von Willebrand factor) and *CD24a* (a glycosylphosphatidylinositol (GPI)-anchored cell surface protein) as good markers for the YS-A trajectory. *Lyve1* (lymphatic vessel endothelial hyaluronan receptor 1) was associated with the YS-B trajectory (Fig. 2g,h and Extended Data Fig. 3c). Whole-mount staining of E9.5 and E10.5 YS, obtained from a Vwf^{eGFP} reporter[31] mouse model, for RUNX1, LYVE1 and eGFP revealed their distinct spatial expression patterns within the extra-embryonic vasculature. While high LYVE1 expression was evident throughout the YS plexus and in large veins, eGFP (*Vwf* expression) staining was confined to large arterial vessels (Fig. 3a,b). Additionally, putative HE cells expressing both RUNX1 and *Vwf* were primarily observed in large arteries and infrequently in large veins. In the plexus $RUNX1^{pos}Vwf^{pos}$ cells were absent at E9.5 and infrequent at E10.5. Putative HE cells expressing RUNX1 and LYVE1 were distributed throughout the plexus (Fig. 3c–e). Altogether, these results suggest the presence of two separate extra-embryonic HE populations. The first (b2 and b3) is LYVE1 positive, dominant until E9.5 and can be found throughout the YS endothelial plexus. The second population (a1 and a2) expresses CD24a and *Vwf*, is prevalent after E9.5 and is found in large extra-embryonic arteries.

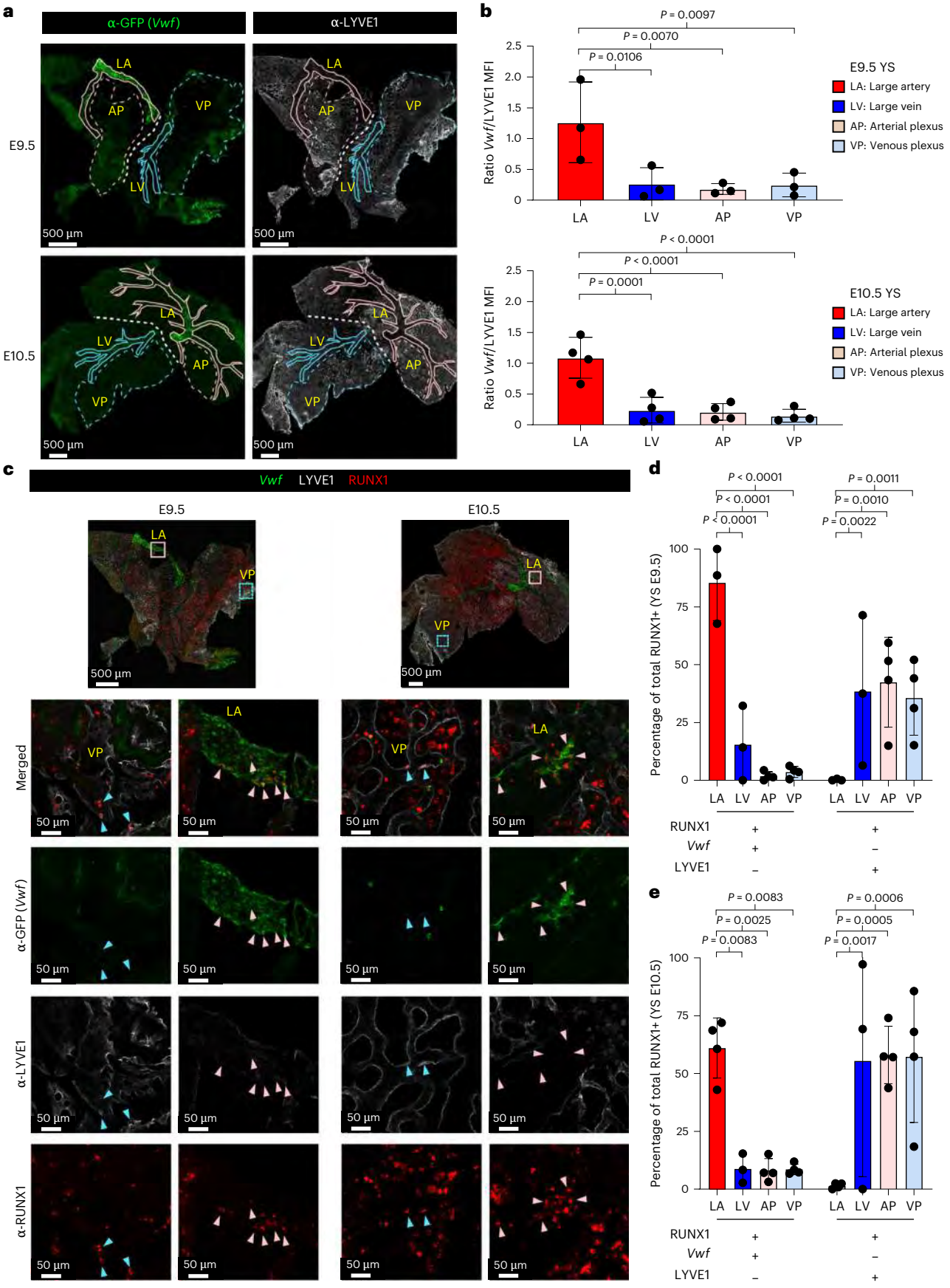
The two extra-embryonic HE populations have distinct hematopoietic potentials

To isolate and functionally characterize the two extra-embryonic HE populations, we screened our data for cell surface markers suitable for FACS enrichment from wild-type (WT) embryos devoid of fluorescent reporters. This highlighted the previously identified *Lyve1* and *CD24a*, as potential markers for respectively the YS-B and YS-A HE trajectory. *Mcam* (melanoma cell adhesion molecule) was expressed at early stages of both trajectories (Fig. 4a,b and Extended Data Fig. 3b,c). Next, we examined by flow cytometry $Runx1^{RFP}$ expression, a strong indicator of HE identity, in extra-embryonic $KIT^{pos}CD31^{pos}LIN^{neg}$ YS endothelial subpopulations defined by a combination of these markers: $LYVE1^{neg}CD24^{pos}MCAM^{pos}$, $LYVE1^{neg}CD24^{pos}MCAM^{neg}$, $LYVE1^{pos}CD24^{neg}MCAM^{pos}$ and $LYVE1^{pos}CD24^{neg}MCAM^{neg}$. Within the $MCAM^{pos}$ cell populations, few cells displayed transcription of the *Runx1* locus (RFP 1.5–23%), suggesting limited HE enrichment. In contrast, both $MCAM^{neg}$ populations were highly enriched for cells with an active *Runx1* locus ($CD24^{pos}LYVE1^{neg}MCAM^{neg}$ 72 – 88% and $CD24^{neg}LYVE1^{pos}MCAM^{neg}$ 62-67%) (Fig. 4c). Subsequent scRNA-seq of cells in these populations confirmed that these $MCAM^{neg}$ populations are enriched for the extra-embryonic HE (Fig. 4d).

To functionally assess the hematopoietic potential of the two YS HE populations, single cells were sorted, co-cultured on OP9, and evaluated for myeloid (GR1, MAC1/CD11b), erythroid (TER119) and lymphoid (CD19) potential by flow cytometry after 14 days of co-culture. Single cells from both YS HE populations displayed hematopoietic activity regardless of the developmental stage (Fig. 4e). Wells seeded with $CD24^{pos}$ cells contained myeloid, lymphoid and mixed lymphoid/myeloid cells, whereas wells seeded with $LYVE1^{pos}$ cells predominantly

Fig. 3 | Spatial separation between transcriptomically different EHT trajectories in the yolk sac. a, Confocal whole-mount immunofluorescence (WM IF) analysis of E9.5 (top) and E10.5 (bottom) Vwf^{eGFP} YS. Images show maximum intensity three-dimensional (3D) projections. Representative areas where fluorescence has been quantified are delimited by lines. Pink solid line, large artery (LA); turquoise solid line, large vein (LV); pink dashed line, arterial plexus (AP); turquoise dashed line, vein plexus (VP). Scale bars, 500 µm. **b**, The ratio of *Vwf*-associated MFI to LYVE1-associated MFI is plotted on the *y* axis, reflecting the relative fluorescence intensities within selected areas in Vwf^{eGFP} YS as displayed in Fig. 3a. $n = 3$ E9.5 and $n = 4$ E10.5 YS were analyzed (6–10 areas per YS). Error bars represent mean ± s.d. Statistical test used was a one-way analysis of variance (ANOVA) (Fisher's least significant difference). **c**, WM IF analysis of E9.5 (left) and E10.5 (right) Vwf^{eGFP} YS. Whole YS images show maximum intensity 3D projections. The boxed area in the merged image is magnified in the lower panel and shows a single 2.5-mm-thick optical slice. Turquoise arrowheads indicate $RUNX1^{pos}Vwf^{neg}LYVE1^{pos}$ putative HE cells; pink arrowheads indicate $RUNX1^{pos}Vwf^{pos}LYVE1^{neg}$ putative HE cells. Scale bars, 500 µm (3D), 50 µm (slice). **d**,**e**, Quantification of the percentage of $RUNX1^{pos}Vwf^{pos}LYVE1^{neg}$ and $RUNX1^{pos}Vwf^{neg}LYVE1^{pos}$ cells on the total of RUNX1 positive cells in LA, LV, AP and VP of E9.5 (**d**) and E10.5 (**e**) Vwf^{eGFP} YS (displayed in Fig. 3c). Each dot represents measurements from an individual YS. E9.5 LA and LV $n = 3$, AP and VP $n = 4$. E10.5 LV $n = 3$, LA, AP and VP $n = 4$ 5–17 areas per YS were analyzed. Error bars represent mean ± s.d. Statistical test used was a two-way ANOVA (Fisher's least significant difference).

https://doi.org/10.1038/s44161-025-00740-z



https://doi.org/10.1038/s44161-025-00740-z

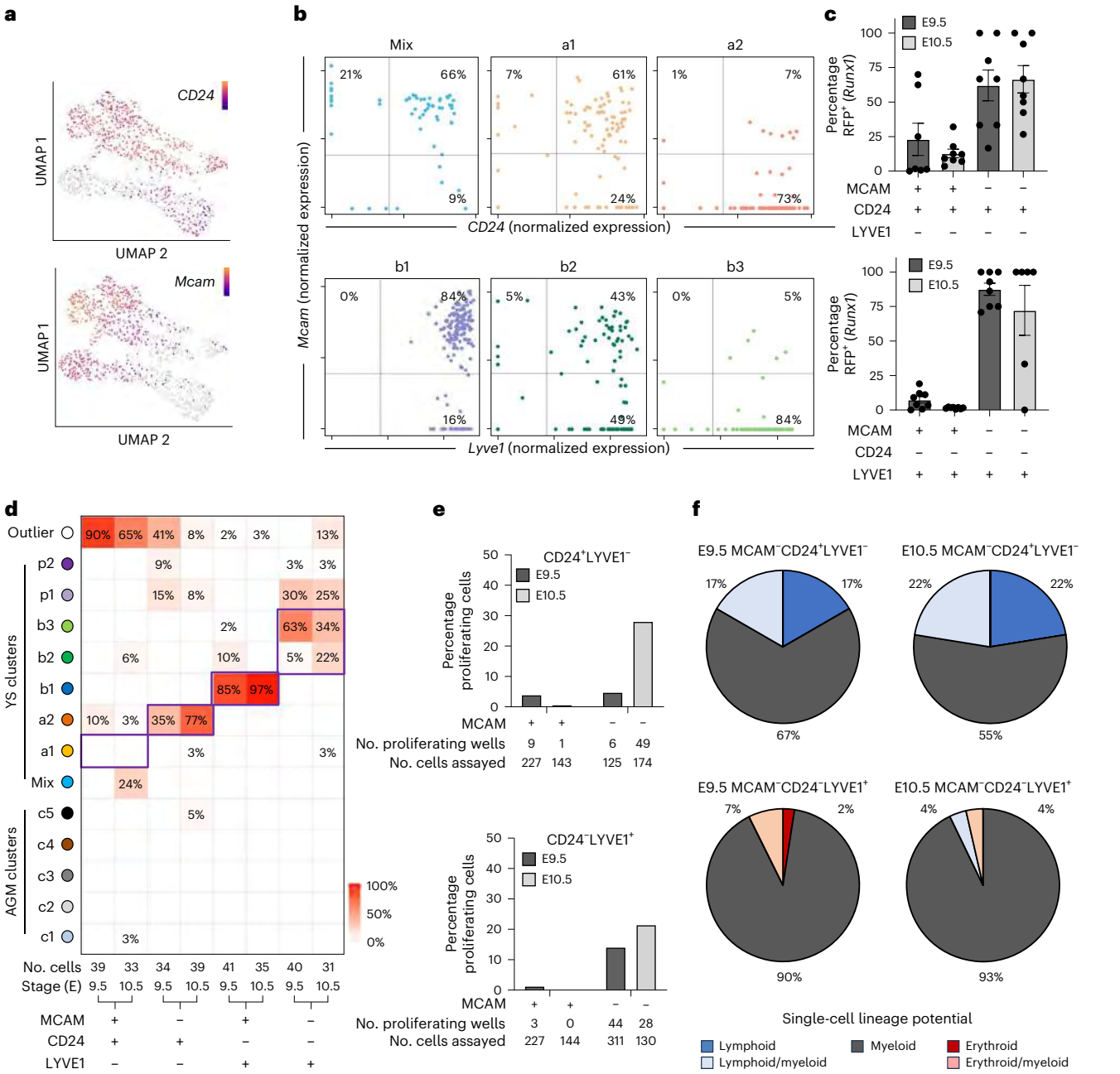


Fig. 4 | CD24^{pos} YS HE has lymphoid-myeloid potential and LYVE1^{pos} YS HE has erythroid-myeloid potential. a, UMAPs depicting the expression of *Cd24a* (marking the AGM and YS-A clusters) and *Mcam* (marking cells toward the endothelial end of all three trajectories). **b**, Correlation between *Cd24a* and *Mcam*, transcript expression in clusters mix, a1 and a2 (YS-A trajectory) (top). Correlation between *Lyve1* and *Mcam*, transcript expression in clusters b1, b2 and b3 (YS-B trajectory) (bottom). **c**, Flow cytometry on extra-embryonic CD45^{neg}CD41^{neg}TER119^{neg} (Lineage negative) CD31^{pos}KIT^{pos} cells from *Runx1*:RFP reporter embryos. MCAM, LYVE1 and CD24 antibodies were used to analyze the proportion of *Runx1* (RFP) expressing cells in different subpopulations. Each dot represents cells from a single YS. E9.5 MCAM^{pos}CD24^{pos}LYVE1^{neg} $n = 7$, E10.5 MCAM^{neg}CD24^{neg}LYVE1^{pos} $n = 6$, all other samples $n = 8$. Bars represent the average ± s.e.m. **d**, Heatmap displaying the distribution (as percentage) of different CD45^{neg}CD41^{neg}TER119^{neg}CD31^{pos}KIT^{pos} FACS-sorted, scRNA-sequenced cell populations across the in silico EHT clusters defined in Fig. 2b. Based on *k*-nearest-neighbor classifier approach. MCAM, LYVE1, CD24 sorting profiles are depicted on the *x* axis. Purple boxes indicate the expected/predicted cluster (*y* axis) for the sorted population (*x* axis) based on the data presented in Fig. 4b. **e**, Single-cell hematopoietic assays of YS-A HE (KIT^{pos}CD31^{pos}LIN^{neg}LYVE1^{neg}CD24^{pos}MCAM^{neg}) and YS-B HE (KIT^{pos}CD31^{pos}LIN^{neg}LYVE1^{pos}CD24^{neg}MCAM^{neg}) cultured on OP9 feeder cells for 14 days. The percentage of wells with proliferating hematopoietic cells is shown. **f**, Lineage distribution of the hematopoietic cells shown in **e** as determined by flow cytometry for myeloid (GR1 and MAC1/CD11b), erythroid (TER119) and lymphoid (CD19) markers. **g**, Violin plots depicting LMP (top) and EMP (bottom) scores across early progenitor clusters p1 and p2. EMP and LM signatures have been previously published and are listed in Supplementary Table 1. Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5 × interquartile range. Two-sided Wilcoxon rank-sum tests were used (with *P* values adjusted via the Benjamini–Hochberg procedure to control the FDR) to compare EMP and LMP as well as clusters p1 and p2. **h**, Violin plots depicting prospective LMP fate (top) and EMP fate (bottom) scores across early progenitor clusters p1 and p2 as well as EMP and LMP populations. EMP fate (8 genes) and LMP fate (14 genes) signatures (Supplementary Table 1) were extracted by intersecting pairwise differential gene expression results (EMP versus LMP and P1 versus P2; Extended Data Fig. 3f). Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5 × interquartile range. Two-sided Wilcoxon rank-sum tests were used (with *P* values adjusted via the Benjamini–Hochberg procedure to control the FDR) to compare EMPs and LMPs as well as clusters p1 and p2.

gave rise to myeloid, erythroid and erythroid/myeloid cells (Fig. 4f). We also utilized our Runx1b^{RFP} reporter model in conjunction with this HE marker panel to enrich for the least progressed MCAM^{pos} endothelial cells within the YS-A (a1, LYVE1^{neg}CD24^{pos}MCAM^{pos}*Runx1*RFP^{pos}) and YS-B (b2, LYVE1^{pos}CD24^{neg}MCAM^{pos}*Runx1*RFP^{pos}) trajectories. The hematopoietic potential of both MCAM^{pos} cell populations was lower than that of the respective MCAM^{neg} populations (Fig. 4e and Extended Data Fig. 3d,e), but the hematopoietic identity of the output was similar (Fig. 4f and Extended Data Fig. 3d,e).

Finally, as the emergence of the lymphomyeloid-producing HE clusters a1–a2 and the erythromyeloid-producing HE cluster b2–b3 closely correlated with the emergence of respectively cluster p1 and p2 (Fig. 2e), we investigated whether these two progenitor populations show signs of early LMP or EMP commitment based on previously published EMP and LMP gene signatures (Fig. 4g and Supplementary Table 1). Although these signatures could distinguish p1 (which resembled EMP) from p2 (which resembled LMP) the difference between the two progenitor populations was minimal (Fig. 4g). This prompted us to investigate if we could define a more powerful gene signature to identify early EMP and LMP potential during progenitor emergence. We used the intersection of pairwise differential gene expression analysis (LMP versus EMP and p1 versus p2) to extract prospective EMP fate and LMP fate signatures (Supplementary Table 1 and Extended Data Fig. 3f). These 'fate' signatures performed better at assigning p1 to an LMP and p2 to an EMP fate (Fig. 4h), suggesting that these gene signatures could be useful to determine whether early progenitors have an EMP or LMP fate.

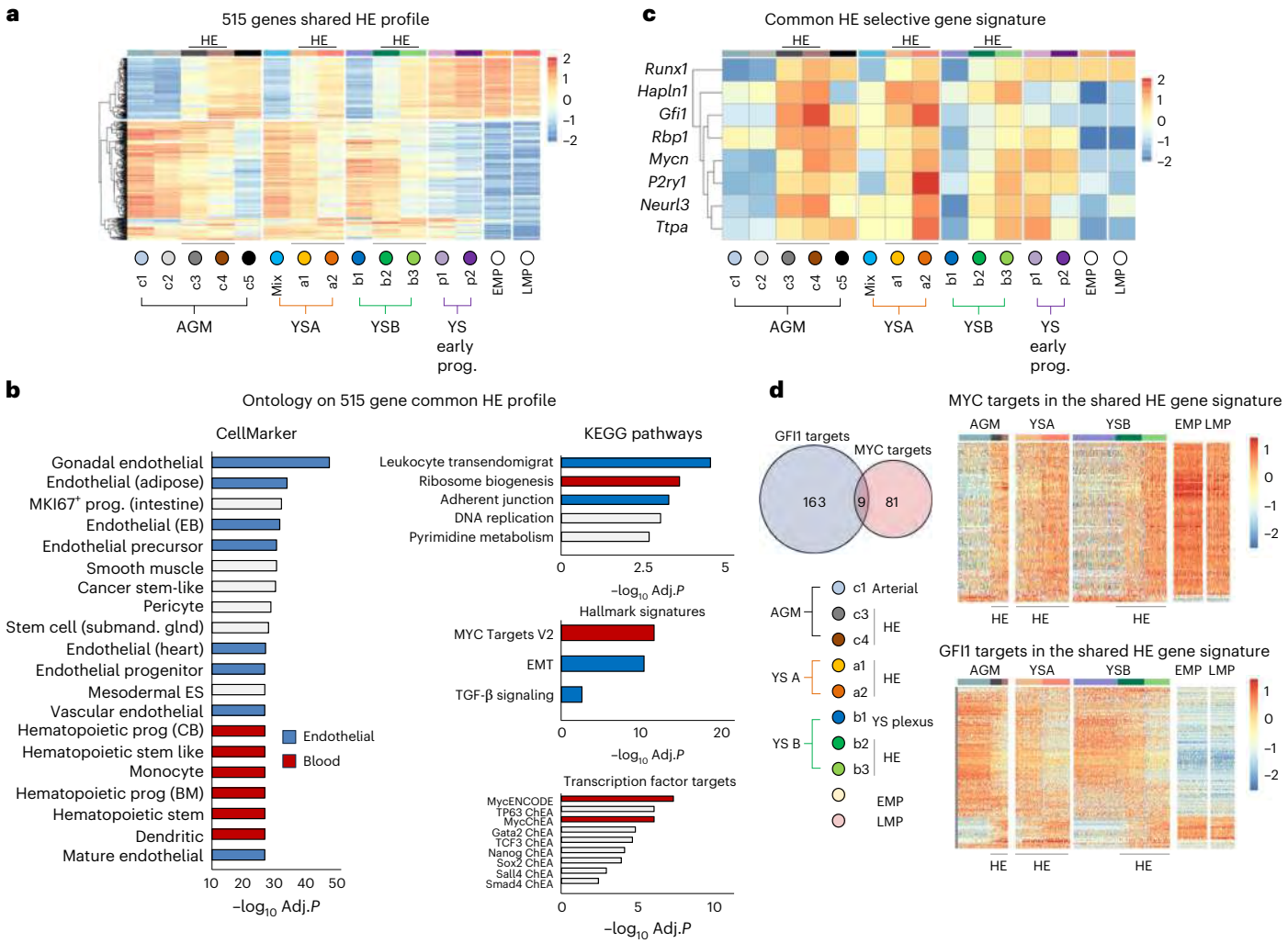
Overall, these results demonstrate that the two extra-embryonic HE populations associate with wave 2 EMP and LMP production, a finding consistent with their distinct endothelial identities and temporal abundance[32]. Given these findings, we hereafter named the three different HE populations based on their distinct localizations; HE^{AGM} (clusters c3–c4), HE^{YSA}/clusters a1–a2, which are found within the YS arteries, HE^{YSP}/clusters b2–b3 which are found within the YS endothelial plexus.

Identification of a shared common HE signature marked by *Gfi1* and *Mycn*

Next, we used the three HE transcriptomes to identify shared HE and EHT characteristics. Acknowledging the continuous nature of the EHT process and the hybrid endothelial–hematopoietic identity of the HE, we identified shared differentially expressed genes (DEGs) between the three HE populations and the extremities (non-HE endothelium and the EMP/LMP populations) of the EHT trajectory (Extended Data Fig. 4a and Supplementary Table 2a). The resulting 515 genes profile is a hybrid of genes expressed in endothelial cells (330 of 515) and genes expressed in hematopoietic cells (178 of 515) (Fig. 5a and Supplementary Table 2a). It contains many genes, including *Proc*, *Neurl3*, *Runx1* and *Gfi1* previously associated with an HE identity (Supplementary Table 2a). Ontology analyses revealed enrichments for categories typically associated with HE and EHT including EMT, TGF-β signaling[33] and ribosome biogenesis (Fig. 5b). Notably, just 7 out of the 515 genes displayed a distinct HE-restricted expression pattern: *Neurl3*, *Hapln1*, *Rbp1*, *Ttpa*, *P2ry1* and the transcription factors *Gfi1* and *Mycn* (Fig. 5c). Almost half (49%, 253 genes) of the genes within the shared profile could be identified as potential targets of these two transcription factors (Fig. 5d and Supplementary Table 2a,b). Most of the potential MYC

Fig. 5 | The shared common HE signature is marked by the transcription factors *Gfi1* and *Mycn*. a, Heatmap depicting the relative expression of all the 515 genes in the shared common profile across intra- and extra- embryonic EHT trajectories (as defined in Fig. 2b). All HE populations display a mixed expression of genes that are strongly expressed in either the endothelial or hematopoietic arms of the EHT trajectory. YS-derived EMPs and LMPs are included for reference (right). **b**, Gene Ontology analysis of the 515 gene universal HE profile. Top Gene Ontology hits (capped at 20) from the CellMarker, KEGG, Hallmark and ChEA/ENCODE databases are shown. Adjusted *P* values were calculated using Fisher's exact test with Benjamini–Hochberg correction **c**, Heatmap depicting the relative expression of all seven HE-selective genes within the shared HE profile and *Runx1* across all three EHT trajectories (as defined in Fig. 2b). YS-derived EMPs and LMPs are included for reference (right columns). **d**, The shared HE profile contains many GFI1 and MYCN target genes. Intersect of GFI and MYCN target genes in the universal HE profile (left). Single-cell heatmaps depicting the expression of shared HE profile MYCN target genes (top) and GFI1 target genes (bottom) across intra- and extra- embryonic HE populations (right). **e**, Violin plots demonstrating that the HE signature defined in Fig. 5c effectively identifies HE cells in all three EHT trajectories analyzed in this manuscript (top)[26] as well as in previously published AGM and YS datasets (bottom)[30,36–38]. Where appropriate the *y* axis of the plots shows the names of the population/cluster nomenclature used in the relevant publications. AE, arterial endothelium; HC, hematopoietic cell. Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5 × the interquartile range. Two-sided Wilcoxon rank-sum tests were used (with *P* values adjusted via the Benjamini–Hochberg procedure to control the FDR) to compare relevant populations.

https://doi.org/10.1038/s44161-025-00740-z



target genes displayed increased expression toward the hematopoietic end of the three HE. Conversely, most of the GFI1 targets[21,34,35] were downregulated toward the hematopoietic end (Fig. 5d and Supplementary Table 2a,b). Finally, we verified that the seven HE-restricted genes in combination with *Runx1* can be used as an eight-gene HE-selective gene signature to identify cells with HE characteristics across independent mouse intra-embryonic (AGM)[26,30,36] and extra-embryonic (YS)[37,38] scRNA-seq EHT datasets (Fig. 5e).

Overall, we established a shared HE profile that encompasses an eight-gene HE signature that is sufficient to identify cells with HE characteristics. Furthermore, *Gfi1* and *Mycn* are the only transcription factors with a HE-restricted expression pattern.

Chromatin modifiers and splicing machinery are differentially expressed between intra- and extra-embryonic HE

To identify differences between the three HE populations, we conducted pairwise differential expression analyses; HE^{AGM} versus HE^{YSA}, HE^{AGM} versus HE^{YSP} and HE^{YSA} versus HE^{YSP} (Fig. 6a, Extended Data Fig. 4b,c and Supplementary Table 3a). Genes significantly upregulated in HE^{YSP} exhibited a distinct (myeloid and erythroid) hematopoietic identity (Extended Data Fig. 5a) with some myeloid genes already expressed within the non-HE YS endothelium (Extended Data Fig. 5b,c and Supplementary Table 3b). Genes significantly upregulated in the other two HE populations did not display a similarly overt hematopoietic signature (Extended Data Fig. 5a and Supplementary Table 3b). HE^{YSA} most closely resembled HE^{AGM} (Extended Data Fig. 4c and Supplementary Table 3a) with the notable expression of Notch pathway components in both the HE^{AGM} and the HE^{YSA} consistent with their arterial identity (Extended Data Fig. 5d)[39].

Overall, HE^{AGM} contained a large group of genes that were more highly expressed compared to one or both extra-embryonic HE (Fig. 6a, Extended Data Fig. 4c and Supplementary Table 3a). Gene Ontology analysis identified two main functionalities within these HE^{AGM} selective genes; chromatin modification and RNA processing/splicing (Fig. 6b and Supplementary Table 3a–c). Specifically, 28 genes related to chromatin modification (Extended Data Fig. 6a) and 49 RNA processing genes (Fig. 6a) demonstrated a ≥ 1.5 $\log_2$ fold change (FC) in HE^{AGM} over at least one of the extra-embryonic HE populations (Supplementary Table 3a–c).

The HE^{AGM}-specific upregulation of RNA processing genes is in line with a recent study describing changes in RNA transcript diversity during AGM EHT[40]. Indeed, we observed HE^{AGM}-specific upregulation of genes encoding the splice site recognition proteins SRSF1, SRSF2 and SRSF9, implicated in changes in transcript diversity observed during EHT in the AGM[40] (Fig. 6c and Supplementary Table 3a). The most differentially expressed RNA processing factors included *Psip1*, encoding a SRSF1 interacting protein, *Hnrnpl*, an activator/repressor of exon inclusion, and *Casc3*, which functions in the non-sense-mediated decay pathway (Fig. 6c and Supplementary Table 3c).

Overall, these analyses reveal that the HE^{YSP} has a distinct hematopoietic profile. Furthermore, HE^{AGM} displays a unique gene expression profile, not observed in either YS HE, characterized by higher expression levels of chromatin modifiers and spliceosome components.

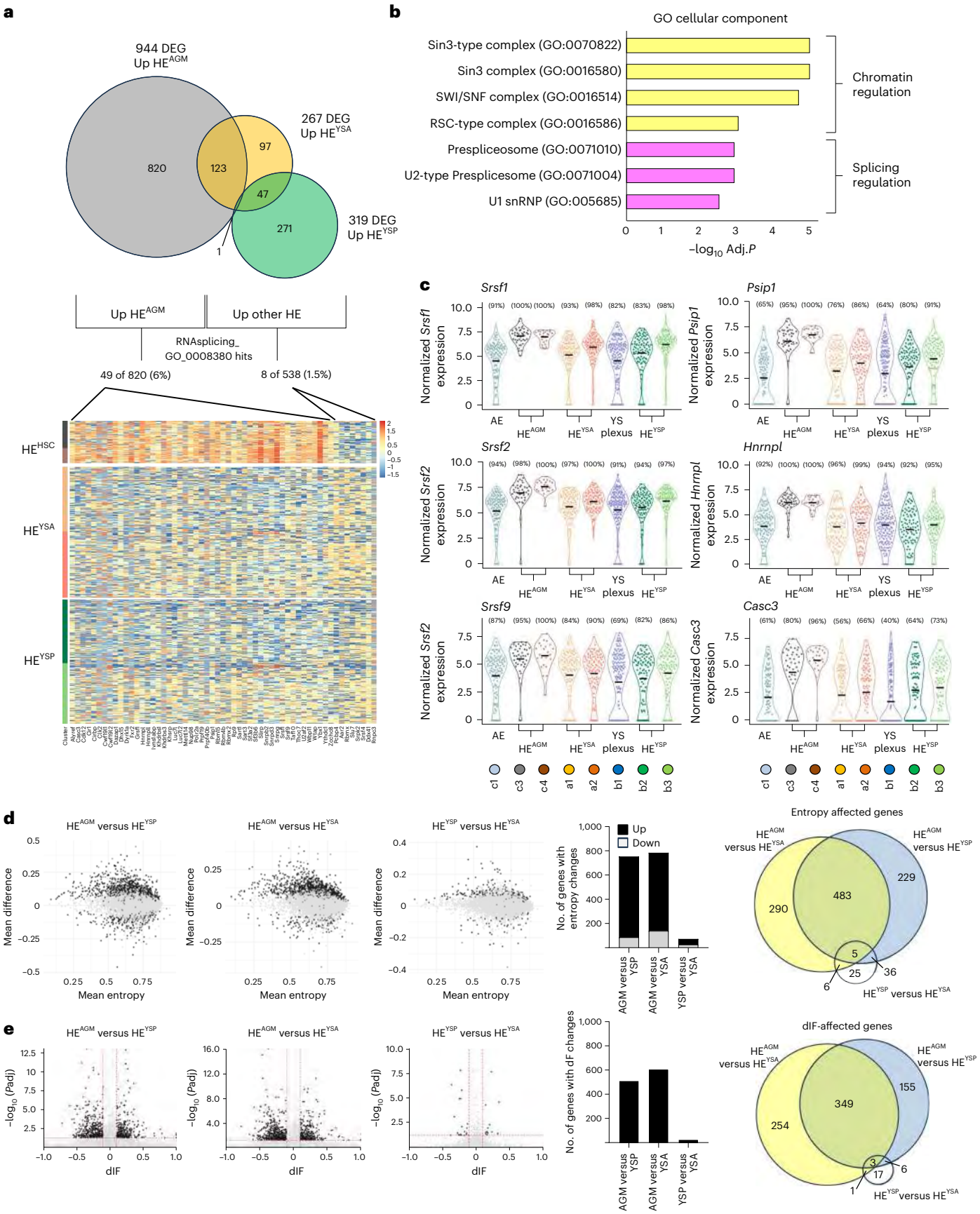
HE^{AGM} transcriptome displays a higher isoform complexity compared to extra-embryonic HE populations

The increased expression level of splicing-related genes in HE^{AGM} suggests that this HE has a distinct isoform expression landscape compared to the YS HE populations. To assess this, we queried our Smart-seq2 dataset at the isoform level. To interrogate differences in isoform expression patterns on a gene level we calculated changes in gene entropy (mean Laplace entropy difference) and dIF (difference in isoform fraction) between the three HE populations[41,42]. In this context increased entropy represents a shift toward a higher transcriptome complexity (more balanced expression of multiple isoforms), while dIF calculations detect shifts in the dominant isoform expressed from a given locus (Extended data Fig. 6b). Both these metrics highlighted a prominent difference in the isoform landscape in the AGM HE compared to both YS HE populations (Fig. 6d,e and Supplementary Table 4a–d). In contrast only minor differences were observed in between the two YS HE populations (Fig. 6d–e and Supplementary Table 4a–d). A total of 1,049 gene loci showed significant differences in entropy (Fig. 6d) compared to one or both YS HE populations. Furthermore, the vast majority (84%) of these gene loci showed increased entropy values in HE^{AGM} compared to the YS HE populations. In contrast, entropy differences were found in only 72 gene loci when the two YS HE were compared to each other, with 64% showing increased entropy within the HE^{YSP} when compared to HE^{YSA}. Analysis of dIF changes (Fig. 6e) gave similar results with a large set of genes (768) showing significant shifts in dominant isoform expression when comparing HE^{AGM} to one or both YS HE populations. 46% of these genes had different dIF values compared to both HE^{YSA} and HE^{YSP}. Only a small set of 27 genes displayed significant dIF differences between the YS HE populations.

Collectively, 1,597 genes exhibited isoform level differences between HE^{AGM} and one or both extra-embryonic HE populations. This gene set included a substantial fraction of genes not detected by standard differential gene expression analyses (Fig. 7a left and Supplementary Table 4a), highlighting that the two methods of analyses capture distinct subsets of potential effector genes. Ontology analyses (Fig. 7a right and Supplementary Table 4e) of the set of 1,597 revealed enrichment of genes involved in basal cellular machineries, including RNA (spliceosome), ribosome and cell cycle-related ontologies. Overall, we found that singular clear shifts from one specific isoform to another were rare, with often multiple different (sized)

Fig. 6 | Differential expression of chromatin modifiers and splicing machinery between intra- and extra-embryonic HE correlates with distinct isoform expression landscapes. a, Venn diagram depicting the result of pairwise DEG analysis on HE^{AGM}, HE^{YSA} and HE^{YSP} (top). 'Up' indicates a gene is upregulated ($\log_2FC > 1.5$) versus at least one other HE population. Single-cell heatmap depicting the genes from GO: RNAsplicing_GO_0008380 that are differentially expressed between the HE populations (bottom). **b**, Gene Ontology analysis (GO cellular components) on genes that are differentially expressed between HE^{AGM} and YS HE ($\log_2FC > 1.5$ higher expression). Adjusted *P* values were calculated using Fisher's exact test with Benjamini–Hochberg correction **c**, Violin plots depicting the expression of selected RNA processing genes across HE^{AGM}, HE^{YSA} and HE^{YSP}. Arterial endothelium and plexus endothelium are shown for reference. **d**, Analyses of isoform entropy difference between HE^{AGM} versus HE^{YSP}, HE^{AGM} versus HE^{YSA} and HE^{YSP} versus HE^{YSA}. Scatter-plots showing the genes having differential usage pattern for the indicated comparison (left). Black dots represent genes with significant mean entropy differences (mean difference > 0.1, FDR *P*adj < 0.05). Statistical test Wilcoxon signed-rank test, two-tailed. Gray dots represent genes with nonsignificant changes. Bar graphs depicting the number of genes with differential entropy values in the different comparisons (middle). The proportion of genes with increased and decreased entropy values are depicted in black and gray, respectively. HE^{AGM} is skewed toward genes with higher entropy values (chi-squared goodness of fit test, $P < 0.0001$). Venn diagram depicting the intersect of the different comparisons (right). Only a very small set of genes shows entropy differences between the HE^{YSB} versus HE^{YSA}. **e**, Differences in isoform fraction (dIF) analyses between HE^{AGM} versus HE^{YSP}, HE^{AGM} versus HE^{YSA} and HE^{YSP} versus HE^{YSA}. Scatter-plots showing the gene with dIF changes for the indicated comparison (left). Black dots represent genes with significant dIF changes (dIF > 0.1, FDR *P*adj < 0.05). Statistical test used was the IsoformSwitchAnalyzeR implementation of the differential transcript usage (DTU) test in the satuRn R package (https://f1000research.com/articles/10-374/v2) (a generalized linear model-based test). Gray dots represent genes with nonsignificant changes. Bar graphs depicting the number of genes with dIF changes in the different comparisons (middle). Venn diagram depicting the intersect of the different comparisons (right). Only a very small set of genes show dIF differences between the HE^{YSP} versus HE^{YSA}.

https://doi.org/10.1038/s44161-025-00740-z



transcripts (both coding and noncoding) showing subtle shifts in proportions (Supplementary Table 4). Examples (Extended Data Fig. 7a–d) were we observed discernible differential splicing events (defined as, partial or complete, exon inclusion/exclusion changes) between HE^{AGM} and YS HE are *Rpl34*, encoding part of the large 60s ribosomal subunit; *Arglu1* a splice modulator; *Ythdf2* a m^6A-dependent RNA degrader; *Pfn1*, a cytoskeleton modulating protein. Notably, the latter two have been implicated in HSPC biology[43,44]. Ontology analyses of the small number of genes affected on an isoform level between HE^{YSP} and HE^{YSM} did not yield robust Gene Ontology results (Extended Data Fig. 6c and Supplementary Table 4e).

Together, these data demonstrate that the increased expression of splicing-related genes in HE^{AGM} correlates with an isoform expression landscape that is distinct from both YS HE populations.

Exclusive expression of Runx1 Δ exon 6 transcripts negatively impacts HSC emergence

To focus on potential drivers of an HSC fate we restricted our analyses to transcription/chromatin factors within the HE^{AGM} isoform list (210/1579; Fig. 7b and Supplementary Table 4a). As with the full list, this sublist also contained a substantial number of candidates not found by differential gene expression or analyses of the shared HE profile (Fig. 7b left and Supplementary Table 4a). Cell-type ontology analyses revealed an enrichment for factors associated with stemness (Fig. 7b middle and Supplementary Table 4e). Most of these factors (24 of 32) displayed differential entropy values (Fig. 7b right and Supplementary Table 4a). The majority of the genes in the entropy category displayed increased entropy values in HE^{AGM} (23 of 24), which precludes the identification of a single dominant differentially expressed isoforms (Supplementary Table 4). A small subset of the genes ($n = 8$) demonstrated dIF changes. Further screening for isoform differences between HE^{AGM} and both YS HE populations, as well as differential splicing events, highlighted *Runx1* as a notable candidate.

Multiple annotated *Runx1* transcripts showed shifts in proportion between the three HE populations (Extended Data Fig. 7e). Overall, there is a mix of distal transcripts (Runx1c and short Runx1c 5′ transcripts) and proximal transcripts (Runx1b and short Runx1b 5′ transcripts). The pattern of expression suggests the three populations are at different stages of shifting from the earliest expressed isoform, *Runx1b*, toward the late expressed isoform *Runx1c*. The proportion of full Runx1c transcripts is at its highest in HE^{AGM}, is decreased HE^{YSA} and is at its lowest in HE^{YSP} (Extended Data Fig. 7e, middle). The opposite was observed for short Runx1c 5′ transcripts (Extended Data Fig. 7e middle and left). We also observed a small but significant twofold increase in the proportion of Runx1 isoforms lacking exon 6 (*Runx1* Δ6) in HE^{AGM} compared to HE^{YSP} (Extended Data Fig. 7e). The difference between HE^{AGM} and HE^{YSA} was much less obvious suggesting that the shift in *Runx1* Δ6 correlates with an arterial identity of the HE.

Although *Runx1c* is known to be preferentially expressed in HSCs[45], previous manipulations of Runx1 isoform expression, by enforcing the expression of only *Runx1b* transcripts, did not reveal striking effects on the HSC population[46]; however, the absence of Δ6 transcripts has been previously associated with reduced numbers of HSPCs, including long-term HSCs (LT HSCs), in the bone marrow of adult mice[47]. Together with our observations this suggest that the Δ6 isoform could potentially impact positively on HSC emergence; however, the specific exclusion of Runx1 exon 6 has not been evaluated in vivo. Therefore, we generated homozygote *Runx1*Δ6 embryos by CRISPR-Cas9 mediated deletion in mouse zygotes followed by implantation[48]. Analysis of E11.5 YS and AGM regions of these embryos demonstrated a significant reduction of emerging hematopoietic cells (defined as either $CD31^{pos}KIT^{pos}CD45^{pos}$ (Fig. 7c) or $KIT^{pos}CD45^{pos}$ (Extended Data Fig. 7f)) specifically in the AGM regions of Δ6 embryos, while the YS seemed largely unaffected (Fig. 7c). Furthermore, in Δ6 embryos we observed significantly less phenotypic LT HSCs within both the E11.5 AGM ($CD31^{pos}SCA^{pos}KIT^{pos}CD45^{pos}EPCR^{pos}$) as well as the E16.5 fetal liver (FL) ($LIN^{neg}CD48^{neg}SCA^{pos}KIT^{pos}CD150^{pos}$) (Fig. 7d). To functionally validate and evaluate the phenotypic Runx1Δ6 FL LT HSCs we performed transplantation experiments (Fig. 7e). Equal numbers of FACS-sorted E14 FL LT HSCs were transplanted (150 LT HSCs per mouse. Figure 7e) into sublethally irradiated mice. Both WT and Runx1Δ6 FL LT HSCs were able to reconstitute hematopoiesis in recipient mice; however, Runx1Δ6 LT HSCs exhibited signs of reduced capacity and/or fitness as the contribution to peripheral blood at 11 weeks was slightly lower ($P < 0.05$) (Fig. 7e). Although not statistically significant a similar trend was observed when analyzing week 12 donor contribution in the total bone marrow of the recipient mice, the bone-marrow lineage negative population ($Ter119^{neg}CD3^{neg}B220^{neg}GR1^{neg}MAC1^{neg}$) and the bone-marrow LSK ($Ter119^{neg}CD3^{neg}B220^{neg}GR1^{neg}MAC1^{neg}SCA1^{pos}KIT^{pos}$) population (Fig. 7f). Lineage commitment appeared unaffected, with

Fig. 7 | Loss of exon 6 containing Runx1 isoforms impacts HSC emergence. **a**, Venn diagram showing the intersect between isoform-based entropy and dIF level differences between HE^{AGM} and one or both extra-embryonic HE populations as well genes found to upregulated in the HE^{AGM} (as shown in Fig. 6a) (left). Gene Ontology analyses across Wikipathways 2024 Mouse, KEGG 2021 human and MSigDB Hallmark 2020 databases (right). The input gene lists consisted of the 1,579 genes that showed differential isoform expression (entropy and/or dIF) between HE^{AGM} and one or both YS HE. Gene lists can be interrogated in Supplementary Table 4. Adjusted *P* values were calculated using Fisher's exact test with Benjamini–Hochberg correction **b**, Venn diagram depicting the intersects between genes upregulated in the HE^{AGM} (as depicted in Fig. 6a), the shared HE profile (as depicted in Fig. 5a) and the 210 gene list of transcription and chromatin factors with distinct HE^{AGM} isoform expression profiles (left). Cell identity analysis performed on all transcription and chromatin factors present in the list of genes with potential HE^{HSC}-selective isoform expression (middle). Adjusted *P* values were calculated using Fisher's exact test with Benjamini–Hochberg correction. Venn diagram intersect for the stemness genes identified in the cell identity analyses (right). The diagram shows if the genes were selected based on changes in Entropy or dIF. **c**, Emerging hematopoietic cells ($CD31^{pos}KIT^{pos}CD45^{pos}$) in E11.5 WT and CRISPR-Cas9 Runx1 Δ exon 6 embryos identified by flow cytometry (Extended Data Fig. 1d). Percentage of emerging hematopoietic cells in E11.5 AGM regions (top). Each point represents a single AGM. Percentage of emerging hematopoietic cells in E11.5 YSs (bottom). Each point represents a single YS. WT $n = 11$, Δ exon 6 $n = 7$. Bars represent the average percentage of emerging hematopoietic cells ± s.e.m. Statistical test used was an unpaired two-tailed *t*-test. **d**, Phenotypic LT HSCs in E11.5 AGM and E16 FL identified by flow cytometry (Extended Data Fig. 1d). Left, percentage of LT HSCs in E11.5 AGMs. Each point represents a single AGM (WT $n = 10$, Δ exon 6 $n = 7$). Right, percentage of LT HSCs in E16 FLs. Each point represents a single FL (WT $n = 11$, Δ exon 6 $n = 11$). Bars represent the average percentage of LT HSC ± s.e.m. Statistical test used was an unpaired two-tailed *t*-test. **e**, Schematic of E14 FL LT HSCs transplantation experiments (top). A total of 150 phenotypic FL LT HSCs (CD45.2) were transplanted into sublethally irradiated NSGS mice (CD45.1). Donor contribution was followed for 12 weeks. Peripheral blood was analyzed by flow cytometry 4, 8 and 11 weeks post-transplant. At 12 weeks post-transplant the bone marrow (BM) was collected and analyzed by flow cytometry. Contribution of donor cells (CD45.2) to the peripheral blood of the transplanted mice at week 4 (WT $n = 4$, Δ exon 6 $n = 5$), 8 (WT $n = 3$, Δ exon 6 $n = 4$) and week 11 (WT $n = 3$, Δ exon 6 $n = 3$) (bottom). Bars represent the average percentage of donor derived blood cells ± s.e.m. Unpaired two-tailed *t*-test. **f**, Bar graphs presenting the percentage of donor derived cells in the BM of recipient mice 12 weeks post-transplant. Donor cell contribution to the total BM (left). Donor cell contribution to the lineage negative ($TER\text{-}119^{neg}CD3^{neg}B220^{neg}GR1^{neg}MAC1^{neg}$) BM population (middle). Donor cell contribution to the BM LSK ($TER\text{-}119^{neg}CD3^{neg}B220^{neg}GR1^{neg}MAC1^{neg}SCA1^{pos}KITt^{pos}$) population (right). $n = 3$. Bars represent the average percentage of donor derived blood cells ± s.e.m. Unpaired two-tailed *t*-test. **g**, Myeloid ($GR1^{pos}$ and/or $MAC1^{pos}$), B cell ($CD19^{pos}$) and T cell ($CD3^{pos}$) lineage output of donor cells in recipient BM 12 weeks post-transplant. $n = 3$. Bars represent the average percentage of lineage contribution of donor derived blood cells ± s.e.m. Unpaired two-tailed *t*-test.

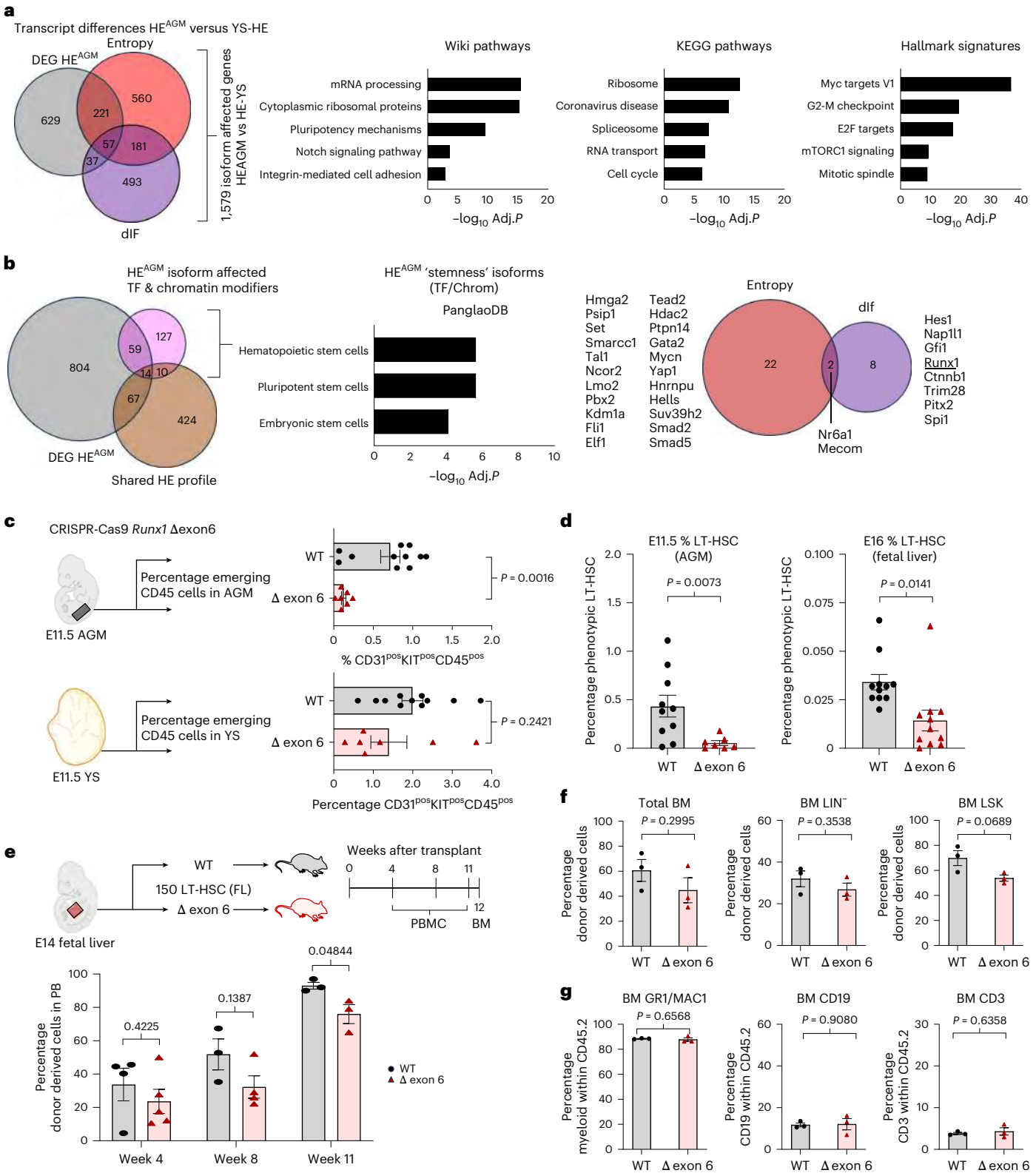
https://doi.org/10.1038/s44161-025-00740-z

the WT and Runx1Δ6 transplants showing comparable contributions to myeloid (GR1^{pos} and/or MAC1^{pos}), CD19^{pos} (B cell) and CD3^{pos} (T cell) populations (Fig. 7g).

Overall, these data indicate that loss of exon 6-containing Runx1 transcripts negatively affects early HSC emergence in the embryo and suggest a balance between exon 6 containing and exon 6 skipping *Runx1* transcripts is required for HSC emergence in the AGM.

Discussion

Hematopoietic cell therapies are potent treatment modalities for many blood diseases, including cancer. A major bottleneck for these treatments is sourcing sufficient patient compatible blood cells. Consequently, unraveling the molecular cues driving the generation of specific blood cell types, to reproduce these processes in vitro, is of great interest. HSCs and lineage-restricted EMP/LMPs are first established



from HE cells during embryogenesis. Intra-embryonic (dorsal aorta) HE, with HSC potential, has been the focus of multiple scRNA-seq studies. Single-cell transcriptomics of HE in the YS, the initial site of EMP and LMP generation, has so far garnered much less attention. Here we present a comprehensive full-length scRNA-seq dataset that covers three parallel embryonic EHT trajectories, each of which contains their own distinct HE population, skewed toward, respectively, HSPCs (intra-embryonic HE^{AGM}), LMPs (YS HE^{YSA}) and EMPs (YS HE^{YSP}).

In contrast to the exclusively arterial identity of intra-embryonic HE[27,49], extra-embryonic HE activity[50–52,32] has been reported throughout the YS endothelium, with an arterial identity being linked to LMP generation[32,50,53]. Combining functional assays, whole-mount imaging and scRNA-seq, we reveal the presence of two different YS HE populations with divergent trajectories, one with an arterial endothelial identity and localization (HE^{YSA}) and another with a plexus endothelial identity and localization (HE^{YSP}). In line with the sequential emergence of embryonic EMPs and LMPs[1], the two YS HE populations also arise sequentially. Recently, EHT events within the large intra- and extra-embryonic arteries have been linked to the generation of short-term fetal-restricted HSPCs[54]. The ability to enrich HE^{LMP}, using our newly identified CD24-LYVE1-MCAM antibody panel, warrants further characterization of the potential of this HE population.

Comparative analysis of all three EHT trajectories identified a shared HE signature composed of a small set of eight genes that can ascertain cells with HE characteristics regardless of their site of origin. Notably, transcriptional repressor *Gfi1* and transcriptional activator *Mycn* were the only two transcription factors with HE-selective expression patterns. *Gfi1* is an established player in EHT initiation via its ability to downregulate the endothelial program[20,21] as also reflected in our data. Conversely, many genes that were newly activated (upregulated versus endothelium) within the HE populations were MYC target genes. It has been recently shown that, within the context of the AGM, *Mycn* expression is required for efficient EHT[55]. Our data suggest that *Mycn*, like *Gfi1*, is a HE-specific core functional component of EHT progression independent of the embryonic localization. Overall, many genes previously only reported/characterized in the context of AGM EHT, and suggested to be associated with the acquisition of HSPC potential, are also part of our universal HE profile, highlighting the need for comparative analyses to identify specific cell fate regulators.

Probing differential gene expression between HE populations revealed a high degree of similarity between YS HE^{YSA} and intra-embryonic HE^{AGM}, likely reflecting their shared arterial endothelial identity. In contrast, YS HE^{YSP} showed a distinct transcriptional profile, characterized by prominent expression of myeloid/hematopoietic genes, which was also partially observed in non-HE endothelial cells of the YS plexus. In HE^{AGM}, we found a prominent enrichment of genes involved in chromatin modification and RNA processing. The latter is especially interesting as several recent studies have indicated that changes in the isoform landscape play a role in the emergence of the hematopoietic system and HSC biology[40,56]. One study by Wang et al. focusing on EHT in the AGM has shown distinct changes in isoform expression profiles in the transition toward HE and subsequently T1-preHSCs[40]. The affected genes were involved in RNA metabolic processes, including RNA splicing, RNA transport and ribonucleoprotein complex biogenesis[40]. Our findings are consistent with these observations and further highlight this phenomenon as specific to HE^{AGM}. In contrast, differences in the isoform landscapes between the two YS HE populations were minimal, suggesting that their identity and potential are predominantly driven by differential gene expression. Notably, we observed HE^{AGM} specific changes in isoform expression patterns for transcription/chromatin factors, including *Runx1*, associated with a stem cell identity. We experimentally demonstrated here that limiting the isoform diversity of Runx1, by introducing an in-frame deletion of exon 6 in all transcripts, negatively impacted immunophenotypic LT HSC detection in AGM and FL. Of note, the absence of Δ6 transcripts has previously been shown to negatively impact HSPCs, including LT HSCs, in the bone marrow of adult mice[47]. Indeed, it has been reported that Runx1Δ6 isoforms can enhance the transactivation ability of the exon 6 containing Runx1 isoforms in vitro[47]. More recently, the interaction between the ETS factor ELF1 and RUNX1 has been shown to enhance HSC self-renewal and prevent HSC differentiation[57]. Notably, the RUNX1 E26 transformation-specific (ETS) factor binding domain has been localized to the region encompassing *Runx1* exon 6 and exon 7 (ref. 58). Together, these data point toward a role for balanced expression of Runx1_exon6 and Runx1_Δexon 6 transcripts in lineage choice.

The dataset presented here provides a unique resource for further characterization of the three HE populations in the mouse embryo. A particularly intriguing observation is that HE^{AGM} exhibits a distinct isoform landscape compared to the YS HE populations; however, identifying isoform combinations that directly determine cell fate remains a substantial challenge. This not only due to the underlying biology, such as the higher isoform entropy observed in HE^{AGM}, but also due to technical limitations in isoform resolution from short-read Smart-seq2 data. To attempt to address this, we performed long-read nanopore sequencing on a subset of 220 (160 HE and 60 early progenitors) cells from our Smart-seq2 dataset. While this approach confirmed a global shift toward higher isoform entropy in HE^{AGM} (Extended Data Fig. 7g), the coverage achieved was lower than that of the Smart-seq2 data and insufficient for robust isoform-level analysis. Further advancements in the sequencing depth and accuracy of long-read single-cell technologies will be particularly beneficial for isoform quantification, transcript coverage and the discovery of novel isoforms. Another limitation of our study is the difficulty in reliably predicting the fate of individual transient HE cells at the single-cell level. For example, within the HE^{AGM} population, we cannot tell which cells will become HSC versus other progenitors. Likewise, in HE^{YSP} and HE^{YSA} populations, we cannot predict erythroid versus myeloid or lymphoid versus myeloid outcomes. Current transcriptomic comparisons using predefined signatures (HSCs, EMPs and LMPs) lack the resolution to detect lineage commitment this early in hematopoietic emergence[26]. These early transient cell states, which may disappear before cells become committed progeny, likely influence fate decisions. This underscores the need to compare HE populations with different developmental outcomes, not just committed populations. As discussed above, many 'AGM-specific' EHT genes associated with HSPC potential are expressed in multiple HE subsets, indicating a role in EHT rather than in lineage commitment. Identifying and validating these transient states will require improved perturbation screens, lineage tracing, and novel analytical approaches.

To conclude, our results reveal three distinct EHT trajectories and suggest that hematopoietic fate decisions in HE^{AGM}, including those toward an HSC cell fate, could at least in part be governed on an isoform level. Overall, our scRNA-seq dataset capturing three distinct EHT trajectories, giving rise to EMPs, LMPs and HSPCs, represents a powerful and unique resource for future investigations of cell fate decision in different HE.

Methods

Mouse embryo generation and processing

Mouse work was performed in accordance with the United Kingdom Animal Scientific Procedures Act (ASPA) 1986. Animal experiments performed at the Cancer Research United Kingdom Manchester Institute (CRUK-MI) were approved by the Animal Welfare and Ethics Review Body of the CRUK-MI. Experiments performed at the University of Oxford were approved by the Oxford Clinical Medicine Ethical Review Committee. Mice were housed in individually ventilated cages under standard conditions, including a 12-h light–dark cycle, ambient temperature of 19–23 °C and relative humidity of 45–65%, in accordance with UK Home Office guidelines and institutional protocols. The transgenic reporter mouse lines (strain C57BL/6JOlaHsd) *Gfi1* (refs. 20,26,29)

and *Runx1b*^{RFP} (refs. 26,28) and *Vwf*^{eGFP} (ref. 31) have been described previously. Vaginal plug detection was considered as E0.5 and staging was confirmed for each embryo at the time of collection by visual inspection. For experiments using the *Gfi1*^{GFP} and *Runx1b*^{RFP} reporters, WT Hsd:ICR (CD-1) females were used to set up breeding pairs ensuring reporter sorted cells were exclusively obtained from embryos. The following primers (custom DNA Oligos Merck) were used to genotype embryos. Gfi1GFP, forward1_5′-CCCTTCTCTCAGAACTCAGAG-3′, forward2_5′-GGAAACGAGGTGGCTTGGAG-3′, reverse_5′-GTCTTGT AGTTGCCGTCGTC-3′ (WT: 245 bp, KI:390 bp). Runx1bRFP, forward1_ 5′-ATGGTGATACAAGGGACATCTTCCC-3′, forward2_5′-ACTTGTAT GTTGGTCTCCCG-3′, reverse_5′-ACCAGAGACTTCTACTACAGGC-3′ (WT, 550 bp; KI, 200 bp).

For the single cell RNA-seq and in vitro functional assays, dissected YSs were digested in a mix of Collagenase IV (2 mg ml^{-1}, Worthington) and DNase I (200 U ml^{-1}, Calbiochem) at 37 °C for 15 min. The dissociated cells were pelleted (300*g* for 5 min at 4 °C) and resuspended in phosphate buffered saline (PBS) containing 10% fetal bovine serum (FBS) and further processed for FACS analyses/sorting.

For YS preparation for whole-mount immunofluorescence staining[25,59,60], embryos were dissected in calcium and magnesium-free PBS, 10% FBS and 0.1 mM EDTA. Embryos were fixed in PBS 4% paraformaldehyde for 1 h, rinsed with PBS (3×, 5 min at RT) and incubated in 50% methanol in PBS (4 °C for 10 min). Samples were stored at −20 °C in 100% methanol until further use.

CRISPR-Cas9 *Runx1* Δ exon 6 embryos

One-cell-stage embryos were electroporated with guides targeting the exon 6 of the *Runx1* gene and Cas9 protein, then reimplanted into surrogate mothers[48]. The guides targeted the following sequences flanking exon 6 of *Runx1* (PAM sequences are underscored) (custom DNA Oligos Merck): 5′-CCTCCCGGTCCCTACACTAGGAC−3′ and 5′-CCCACGGAGCCCACTACCCTCTG-3′ At E11.5, embryos were collected and genotyped using primer pairs flanking exon 6: forward1_ 5′- AGTGGGCTGAAGGAACCT -3′, reverse1_5′-ACGGATTACAGTCTCCA GGA -3′ (WT 779 bp, ko 539 bp) and forward2_5′ CAAGGGGCAAT-GTCCAACAA -3′, reverse2 5′- ACCTGGAACCGATAACTGCA -3′ (WT 637 bp, ko 397 bp). The AGMs of these embryos were subsequently dissected and processed and analyzed by flow cytometry to identify any defects in blood cell development[26]. For E16.5 FLs, dissected livers were crushed with the end of a 1-ml syringe through a 40-µm cell strainer into IMDM + 10% FBS.

Transplantation assay of Runx1 Δ exon 6 embryos

Female NSGS (NOD.Cg-Prkdcscid Il2rgtm1Wjl Tg(CMV-IL3,CSF2, KITLG)1Eav/MloySzJ) (CD45.1) mice, aged 8–12 weeks, were used as recipients after two rounds of irradiation at 200 cGy, 3 h apart. Runx1 Δ exon 6 heterozygote males and females (CD45.2) between the ages of 2–6 months were mated, and vaginal plug detection was considered as day 0.5. E14.5 FLs were genotyped and processed for FACS isolation of LT HSCs (TER-119^{neg}CD3^{neg}B220^{neg}GR1^{neg}CD48^{neg}SCA1^{pos}KIT^{pos}CD150^{pos}) as described above (Supplementary Table 5 lists the antibodies used). Each recipient received 150 LT HSCs intravenously in 200 µl of PBS, along with 20,000 nucleated bone- marrow cells from NSGS donors as a support. Peripheral blood was taken in weeks 4, 8 and 11 after transplantation and terminal samples were collected in week 12.

Flow cytometry

Flow cytometry analyses were performed on a BD LSRFortessa X-20 Cell Analyzer (BD Biosciences) and a Novocyte Quanteon (Agilent). All cell sorting was performed on a BD FACSAria III Cell Sorter (BD Biosciences). Antibodies used for FACS are listed in Supplementary Table 5. For scRNA-seq cells were directly sorted into lysis buffer and snap-frozen before further processing. FlowJo software (BD Biosciences) was used to analyze all FACS data.

In vitro single-cell assays

All single cell assays on YS-derived cells were performed using co-culture with OP9 stromal cells (mouse bone-marrow stromal cell line; ATCC CRL-2749 obtained from the American Type Culture Collection)[21]. In brief, hematopoietic activity assays were performed by FACS sorting single YS cells onto OP9 cells in 96-well plates (one cell per well). The cells were cultured in IMDM (Invitrogen), 10% fetal calf serum, L-glutamine (4 mM), penicillin–streptomycin (50 U ml^{-1}), α-monothioglycerol (15 mM), ascorbic acid (50 ng ml^{-1}), transferrin (180 µg ml^{-1}), IL-11 (5 ng ml^{-1}), EPO (2 U ml^{-1}), oncostatin M (10 ng ml^{-1}), IL-6 (20 ng ml^{-1}), bFGF (10 ng ml^{-1}), IL-3 (100 ng ml^{-1}), SCF (100 ng ml^{-1}), Flt3L (100 ng ml^{-1}) and 2% leukemia inhibitory factor (LIF) supernatant for 10 days before microscopically scoring wells that showed signs of hematopoietic proliferation.

Hematopoietic lineage potential assays were performed similarly but with a different media composition: αMEM (Invitrogen), 10% fetal calf serum, L-glutamine (4 mM), penicillin–streptomycin (50 U ml^{-1}), 2-mercaptoethanol (100 mM), SCF (5 ng ml^{-1}), IL-7 (2 ng ml^{-1}) and Flt3L (5 ng ml^{-1}). After 7 days, wells containing proliferating cells were passaged onto fresh OP9 cells and culture for 7 additional days. Lineages of the hematopoietic cells were defined based on FACS analyses of CD19, CD11b and TER119 cell surface expression (Supplementary Table 5).

Hematopoietic colony-forming unit assays

FACS-sorted YS and FL populations were examined by culturing cells, with (YS) or without (YS and FL) previous co-culture on OP9 for 48 h, in a semi-solid methylcellulose matrix (MethoCult GF M3434, Stem Cell Technologies). Colony output was determined after 7–10 days of culture by colony morphology. Where applicable, OP9 co-culture was performed in IMDM (Invitrogen), 10% fetal calf serum, L-glutamine (4 mM), penicillin–streptomycin (50 U ml^{-1}), α-monothioglycerol (15 mM), ascorbic acid (50 ng ml^{-1}), transferrin (180 µg ml^{-1}), IL-11 (5 ng ml^{-1}), EPO (2 U ml^{-1}), oncostatin M (10 ng ml^{-1}), IL-6 (20 ng ml^{-1}), bFGF (10 ng ml^{-1}), IL-3 (100 ng ml^{-1}), SCF (100 ng ml^{-1}), Flt3L (100 ng ml^{-1}) and 2% LIF supernatant.

Whole-mount Immunofluorescence staining and analyses

Whole-mount staining and analyses[25,59,60]. All primary and secondary antibodies used for immunofluorescence are listed in Supplementary Table 5. Fixed (4% paraformaldehyde) samples were routinely stored at −20 °C in 100% methanol (see 'mouse embryo generation and processing' section). Following rehydration, YS samples were treated with a permeabilizing blocking solution (0.2% Triton X-100, 2% donkey serum and 2% FBS) and incubated overnight with primary antibodies. The next day a second step with secondary antibodies was carried out. After staining, YSs were cleared overnight in a 50% solution of glycerol in PBS at 4 °C and then flat-mounted on Superfrost glass slides. Samples were imaged using a Zeiss 710 confocal microscope equipped with a LD LCI Plan-Apochromat ×25/0.8 Imm Corr DIC M27 objective or an EC Plan-Neofluar ×40/1.30 Oil DIC M27 objective. Confocal image acquisition was carried out using Zeiss Zen software v.2.3 SP1; image processing and analysis was carried out using IMARIS Viewer software v.9.7.2 (Bitplane), ImageJ/Fiji (v.2.3.5–2.9.0) and Adobe Photoshop CC 2021. vWF-associated and Lyve1-associated mean of fluorescence intensity (MFI) was measured by ImageJ/Fiji as mean of gray value in a selected area (an example is shown in Fig. 3a) and expressed in arbitrary units. The ratio of vWF-associated MFI to Lyve1-associated MFI was calculated for the same area. Cell counts were performed using Fiji/ImageJ Cell Counter tool.

scRNA-seq and data processing

Single cells were sorted into wells of a 384-well plate containing lysis buffer and snap-frozen. Libraries were prepared using a modified Smart-seq2 protocol[61]. Paired-end 38 bp or 75 bp sequencing was carried out on the NextSeq500 or NovaSeq 6000 platform (Illumina).

Following sequencing, the raw fastq files were obtained by bcltoFastq conversion (v.2.20.0.422) and were subsequently aligned to the mm10 reference genome using STAR aligner (v.2.7.9a) with the argument 'STARsolo'. This argument allowed simultaneous mapping of reads and quantification of gene expression. The reference genome and gene transfer format file were downloaded from 10x Genomics webpage at https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-mm10-2020-A.tar.gz. The output of 'STARsolo' was loaded into R (v.4.1.0) using the Bioconducter package DropletUtils (v.1.12.1). Downstream analyses were conducted in R using SingleCellExperiment (v.1.14.1) and Seurat (v.4.0.6). A total of 2,365 cells was sequenced (795 on the NextSeq500 and 1570 on the NovaSeq 6000). Next, cells with <2,000 detected genes, >15% mitochondrial content and >10% hemoglobin percentage were excluded, leaving 2,255 high quality cells (705 on the NextSeq500 and 1,550 on the NovaSeq 6000). This filtering process was adopted as previously described[26].

Analysis of YS EHT scRNA-seq datasets

A total of 1,469 scRNA-seq cells were considered high quality YS cells (115 on the NextSeq5000 and 1,354 on the NovaSeq). During the sequencing, 225 technical replicates (the same cells sequenced twice) were introduced. Duplicated technical replicates were removed, retaining cells with that yielded the highest number of genes leaving a total of 1,214 YS cells. Following read quantification and filtering, single cell analysis was performed using the scater (v.1.20.1) package. Raw counts were log-normalized (logNormCounts), gene variance was modeled (modelGeneVar) and the top 2,000 highly variable genes (HVGs) were identified (getTopHVGs). Following normalization, cells were subsequently clustered using graph-based clustering (buildSNNGraph, parameters: $k = 10$, use.dimred = 'PCA'). Based on graph-based clustering, we noted groups of outlier cells that (1) contained high percentage of ribosomal genes with the lowest genes detected; (2) contained high expression of hemoglobin gene (*Hbb-y*); (3) were potential mesenchymal cells with high expression of mesenchymal genes (*Dlk1* and *Ptn*); (4) were a distinct cluster of cells expressing the marker *Folr1*; and (5) were matured megakaryocyte or platelet-contaminating cells with high expression Pf4, Gp5 and *Gp5*. These outlier cells ($n = 139$) were excluded leaving 1,075 YS cells. We next used unsupervised hierarchical clustering (hclust) utilizing the 'ward.D2' distance measure to cluster the cells. The number of clusters were determined based on the dynamic tree cut functionally (cutreeDynamic) yielded six clusters.

Integration of AGM and YS EHT scRNA-seq datasets

AGM datasets were obtained from the Gene Expression Omnibus (GEO) (GSE150412)[26]. From the raw fastq sequencing files, we used the same processing pipeline as was used in the YS EHT scRNA-seq (as described above) to obtain sequencing counts in the AGM dataset. The raw counts of the AGM and the raw counts of YS data were jointly analyzed as a single AnnData object using the scanpy workflow (v.1.6.1). Low-quality cells were removed as previously described[26] and using the same criteria described above. As the previous AGM scRNA-seq cells were sequenced on the NextSeq500 platform and the YS scRNA-seq cells were sequenced on the NovaSeq, a number of AGM FACS-ENDO ($n = 21$) were concurrently isolated, processed and sequenced on the NovaSeq platform with the YS FACS population. Two strategies were employed to determine and subsequently mitigate batch effects. First, differential expression was performed between the FACS-ENDO populations sequenced across the two platforms. Genes with greater than $\log_2$FC of 1.5 and adjusted P value < 0.01 were considered as genes associated with experimental batches. Second, gene that showed variation in detection rates (>50%) between the sequencing platform were identified. These genes were excluded from further analysis.

To focus on the similarity and differences during EHT in the AGM and YS, the non-EHT related populations (AGM venous endothelial and AGM mesenchymal) and the YS FACS-HE$^{KIT\text{-}Neg}$ cells, YS clusters that have progressed beyond the early progenitor stage (YS EMP and YS LMP) were computationally excluded. The raw counts of the remaining cells were log-normalized (sc.pp.normalize_total) and HVGs identified (sc.pp.highly_variable_genes). To generate low dimensional representation, principal-component analysis (PCA) (sc.tl.pca) was conducted on the scaled expression values (sc.pp.scale). The top 50 principal components were used to determine a k-nearest neighbor graph (sc.pp.neighbors(n_neighbors = 20)). Two rounds of semi-supervised Leiden clustering were carried out to identify clusters. Initially, an unsupervised Leiden clustering (sc.tl.leiden(resolution = 2)) was used followed by a semi-supervised merging of clusters with <40 cells. Next, to generate a simplified graph representation of the data, partition-based graph abstraction (PAGA) (sc.pl.paga) was used based on the Leiden groupings. The final UMAP representation was generated using PAGA-initialized positions. The scanpy results were imported into R, where the final representations of the data were generated.

Differential expression analysis and construction of a common HE signature

Differential expression between two groups was performed using the 'limma' package (v.3.54.2) and the 'voom' function. Before differential expression, genes with more than 90% dropout were excluded. Additionally, to mitigate skewing of differential gene expression analyses between HE groups in the HE^{AGM}, HE^{YSA} and HE^{YSP} trajectories, HE clusters were downsampled to the cluster with the lowest cell number in each individual HE. This resulted in three normalized HE populations, each encompassing two clusters with equal representation. The normalized HE was used to generate a universal HE gene expression profile by performing differential expression analyses versus the closest endothelial population and versus the most progressed hematopoietic cells in our dataset (EMPs and LMPs) as illustrated in Extended Data Fig. 3a. Only genes expressed in at least 33% of the cells (for each type of HE) that displayed a $\log_2$FC > 1 and false discovery rate (FDR) < 0.05 versus either the endothelial or hematopoietic ends of the trajectory were taken forward. For inter-HE differential gene expression, the following cut-offs were used: $\log_2$FC > 1.5, FDR < 0.05 and percent of gene-expressing cells (in the upregulated population) >50%. Differentially expressed gene lists were interrogated for enrichment of biological features using the online Enrichr tool[62]. Only results with an adjusted P value < 0.05, an odds ratio >2 and >5 gene hits were taken forward.

Calculation of gene signature scores

To determine a collective gene signature enrichment, the UCell package[63] (v.2.2.0) was used. Based on a given gene list, the UCell signature score (ScoreSignatures_UCell) was calculated for each cell. Gene list for the different signatures used are listed in Supplementary Table 1. EMP-fate and LMP-fate signature were constructed by intersecting the DEGs between EMPs and LMPs (adjusted P value < 0.05) and YS clusters p1 and p2 (adjusted P value < 0.05).

Statistical comparison of UCell scores. To compare UCell signature enrichment between cell clusters, pairwise statistical comparisons were performed using the Wilcoxon rank-sum test (Mann–Whitney U-test). This nonparametric test was chosen due to the non-normal distribution of UCell scores, which typically exhibit right-skewed distributions with a high proportion of zero values. Statistical significance was assessed at $P = 0.05$, and comparisons were visualized using the ggsignif package. For analyses involving multiple comparisons, P values were adjusted using the Benjamini–Hochberg method to control the FDR.

Analysis of publicly available scRNA-seq data. We analyzed the published scRNA-seq data of Fadlullah et al.[26], Zhu et al.[30], Hou et al.[36], Wang et al.[38] and Li et al.[37]. In the scRNA-seq data from Fadlullah et al.[26], we reprocessed the data from raw fastq files using the STARsolo workflow described above. We extracted cluster information

and retained the following AGM EHT population: Arterial endothelial, Pre-HE, HE-HSC, and IAHC. In the scRNA-seq data from Zhu et al.[30], we directly downloaded the count matrix files and the cell annotations from GEO (GSE137116). The Zhu et al. data were filtered to retain cells from E10.5 embryos. Furthermore, only cells from the populations related to EHT were kept: 'Endo (other)', 'Endo (Wnt_low) [AE]', 'Endo (Wnt_high) [AE]', 'Conflux endo [AE]', 'Pre-HE [AE]', 'HE' and 'IAC'. In the scRNA-seq data from Hou et al. (GSE139389), we downloaded the count matrix files from GEO and extracted cluster annotations from the supplementary data (sheet 8, 41422_2020_300_MOESM5_ESM.xls). We retained AGM E10.0-E10.5 endothelial cells corresponding to the following populations: 'vECs' (venous endothelial cells), 'earlyAEC' (early arterial endothelial cells), 'lateAEC' (late arterial endothelial cells), 'Neurl3-EGFP^{+}' (Neurl3-positive cells), 'tif-HEC' (transcriptomic and immunophenotypic and functional HEC) and 'HC' (hematopoietic cells). In the scRNA-seq data from Wang et al. (GSE167588), we downloaded count matrix files from GEO and extracted cluster annotations from the supplementary data (11427_2021_1935_MOESM7_ESM.xls). The dataset included both YS and caudal region populations. We retained the YS: 'YS_Aplnr^{+} EC' (YS Aplnr-positive endothelial cells), 'YS_aEC' (YS arterial endothelial cells), 'YS_HE' (YS HE), 'YS_Ery' (YS erythroid cells). In the scRNA-seq data from Li et al. (GSE173833), we downloaded count matrix files from GEO (GSM5281418) for YS PK44 (CD41^{-}CD43^{-}CD45^{-}CD31^{+}CD201^{+}Kit^{+}CD44^{+}) cells from E10.0 embryos. As cluster annotations were not provided, we performed hierarchical clustering using DEGs between endothelial-biased and hematopoietic-biased populations as described in the original publication (Supplementary Table 1 of the publication). We used Ward's linkage method with Euclidean distance and dynamic tree cutting to identify three distinct clusters (PK44-endo, PK44-mix and PK44-hematopoetic) representing different stages of YS cell progression.

Publicly available GFI genomic binding data

The following three GFI binding datasets (GEO accession codes: GSE57251, GSE22178, GSE69101) obtained from early hematopoietic populations were used: (1) GFI1 and GFI1b binding data from DamID of HE from embryonic stem cells, GEO_GSE57251: GSM1377856, GSM1377857 and GSM1377858 (ref. 21); (2) GFI1b binding data from ChIP–seq of HPC7 cell line (downloaded from Supplementary Table 1 of the online version of the manuscript). Also available at GEO_GSE22178: GSM552235 and GSM552236 (ref. 35); (3) GFI1 and GFI1b binding data from ChIP–seq of mES-derived early hematopoietic progenitors, GEO_GSE69101: GSM1692809, GSM1692853 and GSM1692854 (ref. 34). In each of the studies mentioned, the BED files were downloaded and were annotated with ChIPpeakAnno (v.3.20.1). Peaks were filtered to retain regions within 3 kb of transcription start site for ChIP–seq data and 5 kb of the gene body for DamID data. Genes were identified as potential GFI targets if binding was observed in at least one GFI1 and one GFI1b dataset.

Single-cell isoform transcript analyses

SMARTseq. Raw reads were aligned and quantified with Salmon[64] (v.1.10.2; --libtype OU) against GENCODE transcripts (release 37). Quantification files were read into R using either tximport (v.1.28.0) or importIsoformExpression (IsoformSwitchAnalyzeR v.2.0.1)[42,65]. Isoform switching analysis was perfomed using IsoformSwitchAnalyzeR with isoformSwitchTestSatuRn[42]; significant isoform switches were defined as those with an adjusted $P \leq 0.05$ and $|dIF| > 0.1$. Splicing entropy was assessed using SplicingFactory (v.1.8.0)[41]; significant entropy changes were defined with adjusted $P \leq 0.05$. To detect a skew in entropy changes a chi-squared goodness of fit test was applied.

Nanopore sequencing. Data were basecalled using Dorado (v.0.9.1) (Oxford Nanopore Technologies Dorado, 2025; https://github.com/nanoporetech/dorado) using the high-accuracy model (dna_r10.4.1_e8.2_400bps_hac@v5.0.0). Raw reads were trimmed for SMARTseq adapters using bbduk.sh (v.39.08)[66]. Chimeric reads were filtered from the data using YACRD (v.1.0.0) and porechop (v.0.2.4)[67,68]. Cleaned reads were aligned to GENCODE transcripts (release 37) using Minimap2 (v.2.26)[69], retaining a maximum of ten alignments per read. Alignments were quantified using NanoCount (v.1.0.0.post6)[70]. Abundance files were read into R (v.4.3.0) filtered for low library size samples using findOutliers (scuttle v.1.10.3; type = 'lower', nmads = 1) and normalized using DESeq2 (v.1.40.2)[71]. Estimated and normalized counts were subject to analysis using IsoformSwitchAnalyzeR and SplicingFactory as previously described[41,42].

Reporting summary

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

Data availability

Gene expression data can be queried at https://shiny.cruk.manchester.ac.uk/AGM_YS_dataset_final/. Raw data are deposited in the GEO under accession codes GSE274544 and GSE309071. Three Source data files accompany this manuscript (for the main figures, extended figures and tables).

Code availability

Code is available at https://github.com/zakiF/PublishedPapers/tree/master/YolkSac_AGM and https://github.com/RASellers-CRUK/GL_IsoformAnalysis

References

1. Dzierzak, E. & Bigas, A. Blood development: hematopoietic stem cell dependence and independence. *Cell Stem Cell* **22**, 639–651 (2018).
2. Neo, W. H., Lie-A-Ling, M., Fadlullah, M. Z. H. & Lacaud, G. Contributions of embryonic HSC-independent hematopoiesis to organogenesis and the adult hematopoietic system. *Front. Cell Dev. Biol.* **9**, 631699 (2021).
3. Bertrand, J. Y. et al. Three pathways to mature macrophages in the early mouse yolk sac. *Blood* **106**, 3004–3011 (2005).
4. Palis, J., Robertson, S., Kennedy, M., Wall, C. & Keller, G. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* **126**, 5073–5084 (1999).
5. McGrath, K. E. et al. Distinct sources of hematopoietic progenitors emerge before hscs and provide functional blood cells in the mammalian embryo. *Cell Rep.* **11**, 1892–1904 (2015).
6. Boiers, C. et al. Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell* **13**, 535–548 (2013).
7. Lin, Y., Yoder, M. C. & Yoshimoto, M. Lymphoid progenitor emergence in the murine embryo and yolk sac precedes stem cell detection. *Stem Cells Dev.* **23**, 1168–1177 (2014).
8. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897–906 (1996).
9. Taoudi, S. et al. Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell* **3**, 99–108 (2008).
10. Kobayashi, M. et al. HSC-independent definitive hematopoiesis persists into adult life. *Cell Rep.* **42**, 112239 (2023).
11. Patel, S. H. et al. Lifelong multilineage contribution by embryonic-born blood progenitors. *Nature* **606**, 747–753 (2022).
12. Boisset, J. C. et al. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120 (2010).
13. Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891 (2009).

14. Jaffredo, T., Gautier, R., Eichmann, A. & Dieterlen-Lievre, F. Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny. *Development* **125**, 4575–4583 (1998).
15. Lacaud, G. & Kouskoff, V. Hemangioblast, hemogenic endothelium, and primitive versus definitive hematopoiesis. *Exp. Hematol.* **49**, 19–24 (2017).
16. Lancrin, C. et al. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* **457**, 892–895 (2009).
17. Lie-A-Ling, M. et al. Regulation of RUNX1 dosage is crucial for efficient blood formation from hemogenic endothelium. *Development* https://doi.org/10.1242/dev.149419 (2018).
18. Okuda, T., van Deursen, J., Hiebert, S. W., Grosveld, G. & Downing, J. R. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321–330 (1996).
19. Wang, Q. et al. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl Acad. Sci. USA* **93**, 3444–3449 (1996).
20. Lancrin, C. et al. GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood* **120**, 314–322 (2012).
21. Thambyrajah, R. et al. GFI1 proteins orchestrate the emergence of haematopoietic stem cells through recruitment of LSD1. *Nat. Cell Biol.* **18**, 21–32 (2016).
22. Hadland, B. K. et al. A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood* **104**, 3097–3105 (2004).
23. Thambyrajah, R. & Bigas, A. Notch signaling in HSC emergence: when, why and how. *Cells* https://doi.org/10.3390/cells11030358 (2022).
24. Neo, W. H. et al. Cell-extrinsic hematopoietic impact of Ezh2 inactivation in fetal liver endothelial cells. *Blood* **131**, 2223–2234 (2018).
25. Neo, W. H. et al. Ezh2 is essential for the generation of functional yolk sac derived erythro-myeloid progenitors. *Nat. Commun.* **12**, 7019 (2021).
26. Fadlullah, M. Z. H. et al. Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation marked by ACE. *Blood* **139**, 343–356 (2022).
27. Yokomizo, T. & Dzierzak, E. Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development* **137**, 3651–3661 (2010).
28. Sroczynska, P., Lancrin, C., Kouskoff, V. & Lacaud, G. The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood* **114**, 5279–5289 (2009).
29. Yucel, R., Kosan, C., Heyd, F. & Moroy, T. Gfi1:green fluorescent protein knock-in mutant reveals differential expression and autoregulation of the growth factor independence 1 (Gfi1) gene during lymphocyte development. *J. Biol. Chem.* **279**, 40906–40917 (2004).
30. Zhu, Q. et al. Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood* **136**, 845–856 (2020).
31. Sanjuan-Pla, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
32. Yzaguirre, A. D. & Speck, N. A. Insights into blood cell formation from hemogenic endothelium in lesser-known anatomic sites. *Dev. Dyn.* **245**, 1011–1028 (2016).
33. Thambyrajah, R. et al. HDAC1 and HDAC2 modulate TGF-β signaling during endothelial-to-hematopoietic transition. *Stem Cell Rep.* **10**, 1369–1383 (2018).
34. Goode, D. K. et al. Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev. Cell* **36**, 572–587 (2016).
35. Wilson, N. K. et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* **7**, 532–544 (2010).
36. Hou, S. et al. Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses. *Cell Res* **30**, 376–392 (2020).
37. Li, Y. Q. et al. Spatiotemporal and functional heterogeneity of hematopoietic stem cell-competent hemogenic endothelial cells in mouse embryos. *Front. Cell Dev. Biol.* **9**, 699263 (2021).
38. Wang, C. et al. Adult-repopulating lymphoid potential of yolk sac blood vessels is not confined to arterial endothelial cells. *Sci. China Life Sci.* **64**, 2073–2087 (2021).
39. Thambyrajah, R. et al. Cis inhibition of NOTCH1 through JAGGED1 sustains embryonic hematopoietic stem cell fate. *Nat. Commun.* **15**, 1604 (2024).
40. Wang, F. et al. Single-cell architecture and functional requirement of alternative splicing during hematopoietic stem cell formation. *Sci. Adv.* **8**, eabg5369 (2022).
41. Danko, B., Szikora, P., Por, T., Szeifert, A. & Sebestyen, E. SplicingFactory-splicing diversity analysis for transcriptome data. *Bioinformatics* **38**, 384–390 (2022).
42. Vitting-Seerup, K. & Sandelin, A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics* **35**, 4469–4471 (2019).
43. Lv, J. et al. Endothelial-specific m(6)A modulates mouse hematopoietic stem and progenitor cell development via Notch signaling. *Cell Res.* **28**, 249–252 (2018).
44. Zheng, J. et al. Profilin 1 is essential for retention and metabolism of mouse hematopoietic stem cells in bone marrow. *Blood* **123**, 992–1001 (2014).
45. Draper, J. E. et al. RUNX1B expression is highly heterogeneous and distinguishes megakaryocytic and erythroid lineage fate in adult mouse hematopoiesis. *PLoS Genet.* **12**, e1005814 (2016).
46. Draper, J. E. et al. Mouse RUNX1C regulates premegakaryocytic/erythroid output and maintains survival of megakaryocyte progenitors. *Blood* **130**, 271–284 (2017).
47. Komeno, Y. et al. Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood* **123**, 3760–3769 (2014).
48. Modzelewski, A. J. et al. Efficient mouse genome engineering by CRISPR-EZ technology. *Nat. Protoc.* **13**, 1253–1274 (2018).
49. de Bruijn, M. F., Speck, N. A., Peeters, M. C. & Dzierzak, E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J.* **19**, 2465–2474 (2000).
50. Frame, J. M., Fegan, K. H., Conway, S. J., McGrath, K. E. & Palis, J. Definitive hematopoiesis in the yolk sac emerges from Wnt-responsive hemogenic endothelium independently of circulation and arterial identity. *Stem Cells* **34**, 431–444 (2016).
51. Goldie, L. C., Lucitti, J. L., Dickinson, M. E. & Hirschi, K. K. Cell signaling directing the formation and function of hemogenic endothelium during murine embryogenesis. *Blood* **112**, 3194–3204 (2008).
52. Li, W., Ferkowicz, M. J., Johnson, S. A., Shelley, W. C. & Yoder, M. C. Endothelial cells in the early murine yolk sac give rise to CD41-expressing hematopoietic cells. *Stem. Cells Dev.* **14**, 44–54 (2005).
53. Ning, X. et al. Divergent expression of Neurl3 from hemogenic endothelial cells to hematopoietic stem progenitor cells during development. *J. Genet. Genomics* **50**, 661–675 (2023).
54. Barone, C. et al. Hemogenic endothelium of the vitelline and umbilical arteries is the major contributor to mouse fetal lympho-myelopoiesis. Preprint at *bioRxiv* https://doi.org/10.1101/2024.07.11.603050 (2024).

55. Laruy, B., Garcia-Gonzalez, I., Casquero-Garcia, V. & Benedito, R. Endothelial-to-hematopoietic transition is induced by Notch glycosylation and upregulation of Mycn. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.13.295238 (2020).
56. Ramabadran, R. et al. DNMT3A-coordinated splicing governs the stem state switch towards differentiation in embryonic and haematopoietic stem cells. *Nat. Cell Biol.* **25**, 528–539 (2023).
57. Robertson, A. L. et al. Hematopoietic stem cell division is governed by distinct RUNX1 binding partners. Preprint at *bioRxiv* https://doi.org/10.1101/2024.06.07.596542 (2024).
58. Yu, M. et al. The unique functions of Runx1 in skeletal muscle maintenance and regeneration are facilitated by an ETS interaction domain. *Development* **151** (2024).
59. Azzoni, E. et al. The onset of circulation triggers a metabolic switch required for endothelial to hematopoietic transition. *Cell Rep.* **37**, 110103 (2021).
60. Azzoni, E. et al. Kit ligand has a critical role in mouse yolk sac and aorta-gonad-mesonephros hematopoiesis. *EMBO Rep.* https://doi.org/10.15252/embr.201745477 (2018).
61. Draper, J. E. et al. A novel prospective isolation of murine fetal liver progenitors to study in utero hematopoietic defects. *PLoS Genet.* **14**, e1007127 (2018).
62. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
63. Andreatta, M. & Carmona, S. J. UCell: robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
64. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
65. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
66. Bushnell, B. BBMap: A fast, accurate, splice-aware aligner. In *Proc. 9th Annual Genomics of Energy & Environment Meeting* (U.S. Department of Energy Joint Genome Institute, 2014).
67. Marijon, P., Chikhi, R. & Varre, J. S. yacrd and fpa: upstream tools for long-read genome assembly. *Bioinformatics* **36**, 3894–3896 (2020).
68. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* **3**, e000132 (2017).
69. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
70. Gleeson, J. et al. Accurate expression quantification from nanopore direct RNA sequencing with NanoCount. *Nucleic Acids Res.* **50**, e19 (2022).
71. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

We thank the following facilities of the Cancer Research UK Manchester Institute for technical support: Advanced Imaging, Biological Resources Unit, Genome Editing and Mouse Models, Flow Cytometry and the Molecular Biology Core Facility. The Lacaud group is supported by Cancer Research UK, C5759/A20971. E.A. was supported by a Fondazione Cariplo 'Biomedical Research conducted by young researchers' grant, reference no. 2018-0102, a Leukemia Research Foundation 'New Investigator Blood Cancer Research Grant Program', award ID 831382, a Cariplo-Telethon Alliance grant, reference no. GJC22013 and a WorldWide Cancer Research grant, reference no. 24-0083. C.B. was supported by Fondazione Umberto Veronesi. The Jacobsen group was supported by The Swedish Research Council (538-2013-8995) and the Medical Research Council (MC_UU_12009/5). We are grateful to A. Porter (CRUK-MI Research Integrity and Training Adviser) for reviewing the manuscript and providing valuable feedback. We apologize to the authors whose work we were unable to cite due to space limitations.

Author contributions

W.H.N., M.Z.H.F., H.B., R.T., M.L. and G.L. wrote the manuscript. W.H.N., M.Z.H.F., C.B., R.T., E.A., M.L. and G.L. designed experiments. W.H.N., M.Z.H.F., M.L., C.B., G.Q., F.T.F., J.C., G.S., N.M., R.T. and E.A. performed experiments. M.Z.H.F., H.B. and R.S. performed bioinformatic analyses. M.I. and S.M.B. supervised H.B.'s bioinformatic analyses. Whole-mount imaging experiments were designed performed and analyzed by C.B., G.Q., F.T.F., G.S. and E.A. VWF-reporter embryos were processed and provided by J.C. and S.E.W.J. Sequencing and library preparation was performed by J.W. and supervised by W.B. N.M. generated the Runx1 Δ exon 6 transgenic mice. M.L. and G.L. conceived and supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s44161-025-00740-z.

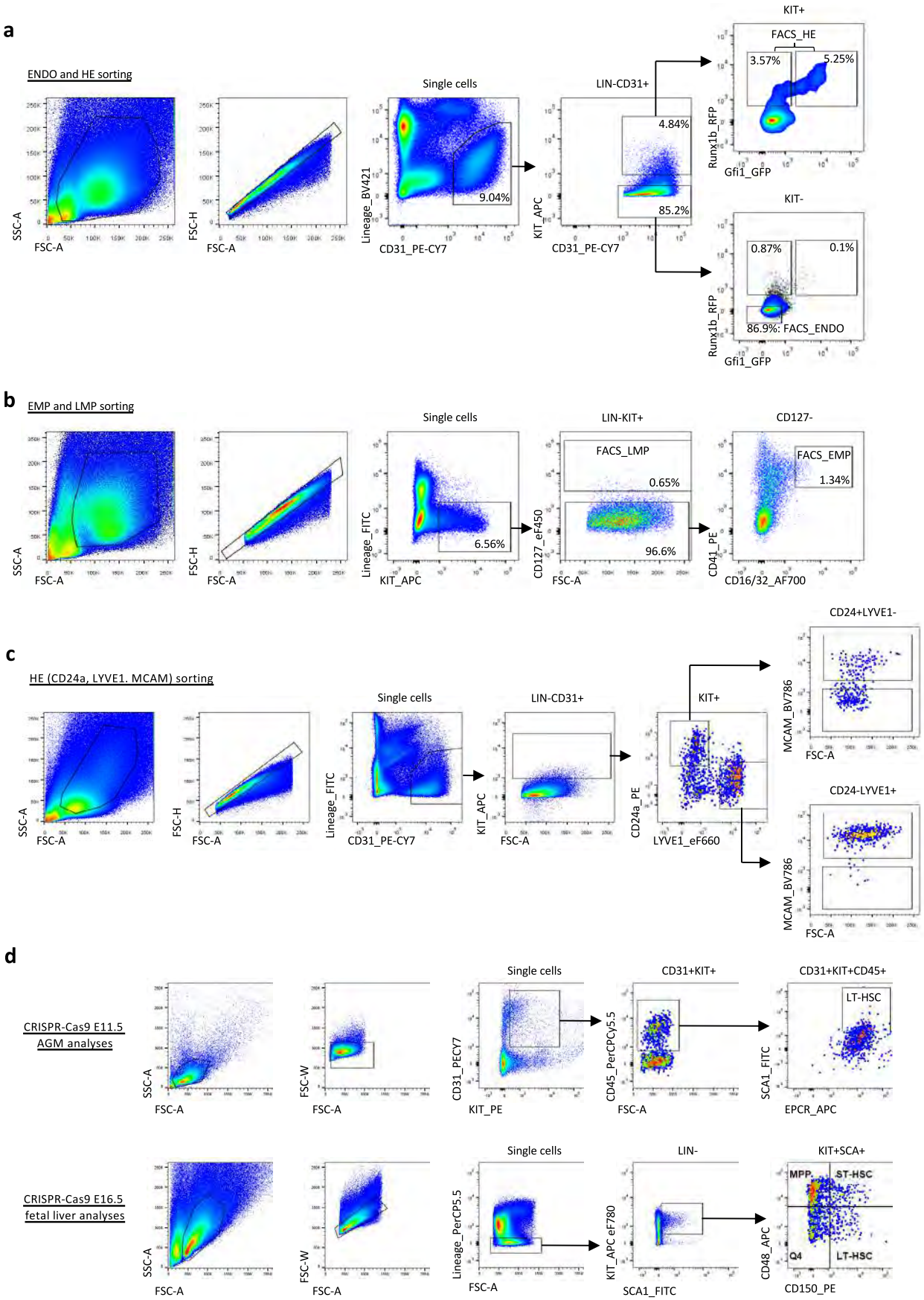
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44161-025-00740-z.

Correspondence and requests for materials should be addressed to Michael Lie-A-Ling or Georges Lacaud.

Peer review information *Nature Cardiovascular Research* thanks Andrea Ditadi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

[1]Stem Cell Biology Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. [2]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA. [3]Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK. [4]School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy. [5]Haematopoietic Stem Cell Biology Laboratory, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. [6]Department of Immunology and Inflammation, Imperial College London, London, UK. [7]Computational Biology Support, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. [8]Molecular Biology Core Facility, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. [9]Genome Editing and Mouse Models, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK. [10]Departamento de Biología Molecular, Facultad de Medicina, Universitat Cantabria, Santander, Spain. [11]Center for Hematology and Regenerative Medicine, Department of Medicine Huddinge & Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden. [12]Department of Hematology, Karolinska University Hospital, Stockholm, Sweden. [13]Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy. [14]These authors contributed equally: Wen Hao Neo, Muhammad Zaki Hidayatullah Fadlullah. ✉e-mail: Michael.Lie-a-ling@cruk.manchester.ac.uk; Georges.Lacaud@cruk.manchester.ac.uk

https://doi.org/10.1038/s44161-025-00740-z

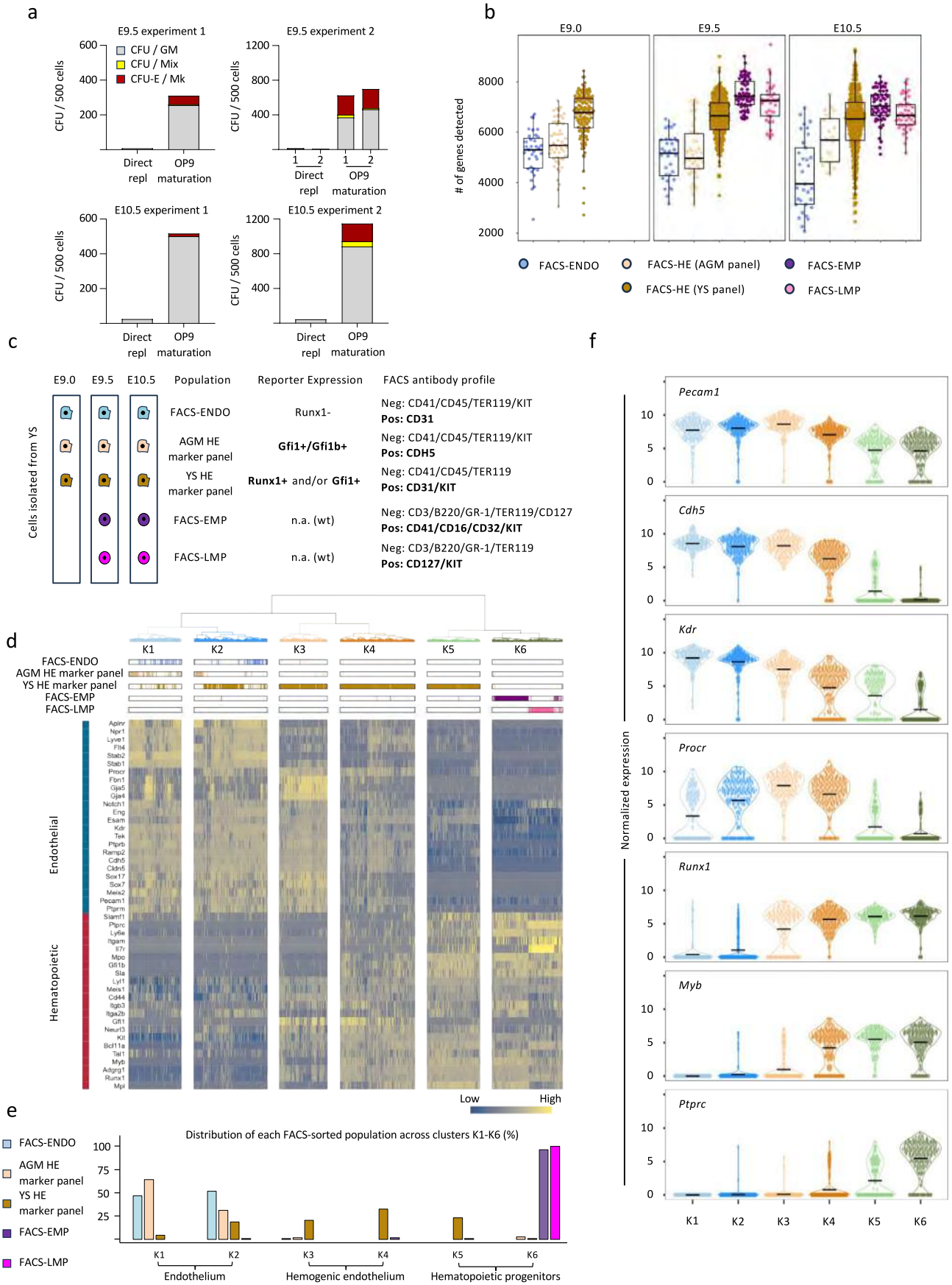


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s44161-025-00740-z

Extended Data Fig. 1 | FACS sorting profiles used in this study. a, Gating strategy for endothelial and hemogenic endothelial sorts from yolk sac (YS). To determine which populations in the YS contain hemogenic potential, the following populations were sorted and analyzed for hematopoietic activity *in vitro:* $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{neg}Runx1b{:}RFP^{pos}$ $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{neg}Runx1b{:}RFP^{pos}Gfi1{:}GFP^{pos}$ $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}Runx1b{:}RFP^{pos}$ $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}Runx1b{:}RFP^{pos}Gfi1{:}GFP^{pos}$ Non-hemogenic endothelial cells (FACS-ENDO) for scRNA-seq were sorted as: $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{neg}Runx1b{:}RFP^{neg}$ Hemogenic endothelium enriched (FACS-HE) populations for scRNA-seq were sorted as: $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}$**R**$unx1b{:}RFP^{pos}$**G**$fi1{:}GFP^{pos}$ $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}$**G**$fi1{:}GFP^{pos}$ $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}$**R**$unx1b{:}RFP^{pos}Gfi1{:}GFP^{neg}$ **b**, Gating strategy for erythromyeloid progenitors (EMP) and lymphomyeloid progenitors (LMP) sorts from YS. FACS-EMP: $CD3^{neg}B220^{neg}GR1^{neg}TER119^{neg}KIT^{pos}CD127^{neg}CD41^{pos}CD16/32^{pos}$. Bertrand, J.Y. et al. Blood 106, 3004-3011 (2005). Frame, J.M., Fegan, K.H., Conway, S.J., McGrath, K.E. & Palis, J. Stem Cells 34, 431-444, 2016. FACS-LMP: $CD3^{neg}B220^{neg}Gr1^{neg}TER119^{neg}KIT^{pos}CD127^{pos}$ Boiers, C. et al. Cell Stem Cell 13, 535-548, 2013. Yoshimoto, M. et al. Proc Natl Acad Sci U S A 108, 1468-1473, 2011. **c**, Gating strategy to enrich for HE with LMP potential and HE with EMP potential from YS. FACS-$HE^{LMP:}$ $CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}CD24a^{pos}LYVE1^{neg}MCAM^{neg}$ FACS-HE^{EMP}:$CD41^{neg}CD45^{neg}TER119^{neg}CD31^{pos}KIT^{pos}CD24a^{neg}LYVE1^{pos}MCAM^{neg}$ **d**, Gating strategy to analyze phenotypic long-term HSC (LT-HSC) and emerging hematopoietic cells ($CD31^{pos}KIT^{pos}CD45^{pos}$) in E11.5 AGM regions and long-term HSC (LT-HSC) in E16.6 fetal livers. LT-HSC (AGM): $CD31^{pos}SCA^{pos}KIT^{pos}CD45^{pos}EPCR^{pos}$. LT-HSC (fetal liver): $B220^{neg}CD3^{neg}GR1^{neg}TER119^{neg}CD48^{neg}SCA^{pos}KIT^{pos}CD150^{pos}$.

https://doi.org/10.1038/s44161-025-00740-z

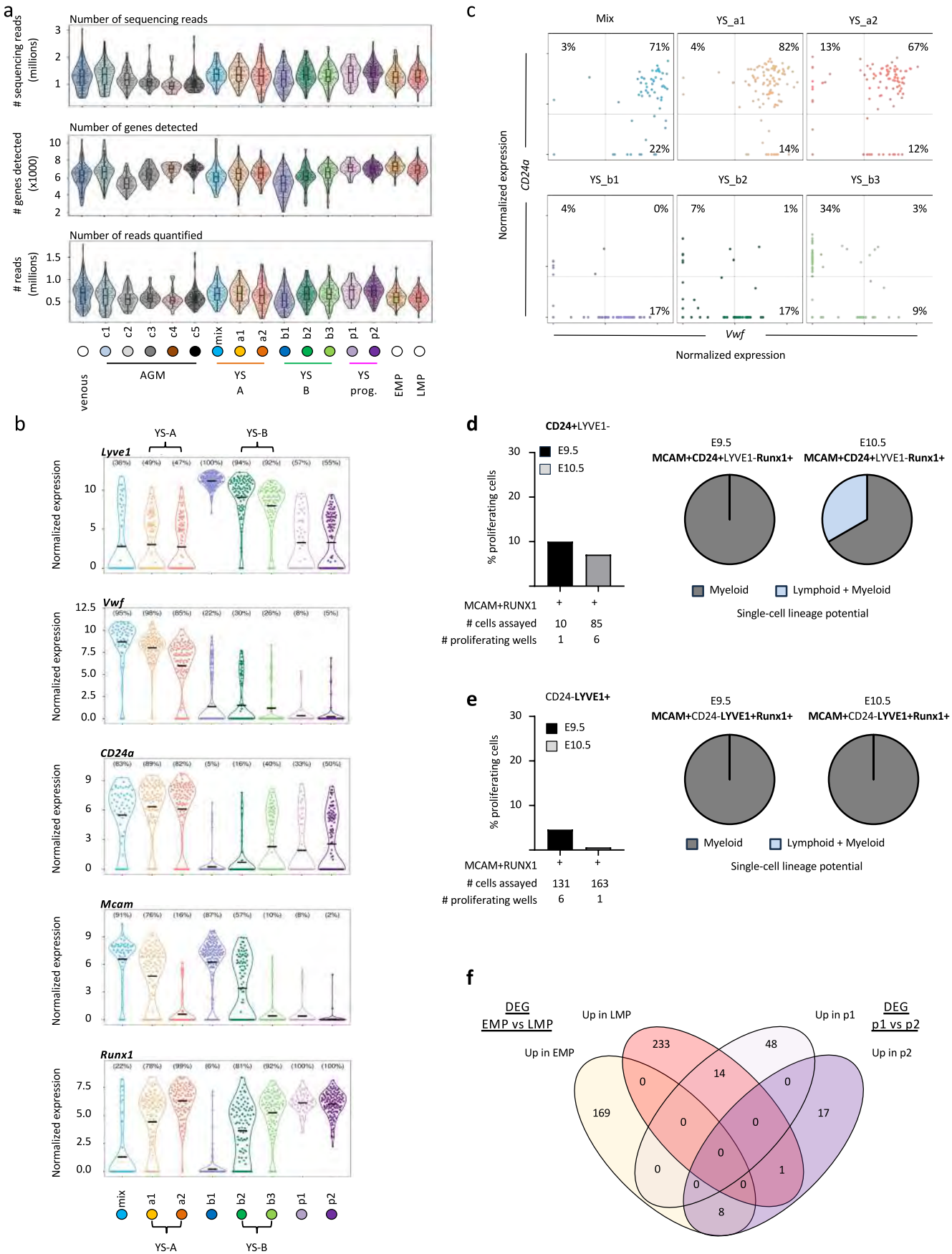


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s44161-025-00740-z

Extended Data Fig. 2 | Identification of HE activity in KIT^{pos}Runx1^{pos} CD31^{pos}LIN^{neg} extra-embryonic cells for single-cell profiling of extra-embryonic EHT. a, Hematopoietic colony forming unit (CFU) assay on KIT^{pos}Runx1^{pos}CD31^{pos}LIN^{neg} extra-embryonic cells FACS-isolated from Runx1^{RFP} reporter mice. Cells were either directly replated, or co-cultured with OP9 feeder cells for 48 h before replating. Hematopoietic colonies were quantified after 10 days. Individual CFU assays are shown. N = 2 biological experiments. E/Mk = Erythrocyte / Megakaryocyte, Mix = Granulocyte / Erythrocyte / Macrophage / Megakaryocyte, G/M = Granulocyte / Macrophage. **b**, Numbers of genes detected and the number of cells that passed QC within each FACS-sorted population from dissected E9.0, E9.5 and E10.5 YS (full-length scRNA-seq, Smart-seq2). Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5x the inter-quartile range. **c**, Schematic of the cell populations FACS-sorted from dissected E9, E9.5 and E10.5 YS and processed for full-length single-cell Smart-seq2 RNA sequencing. Endo: endothelium, AGM HE marker panel (CD41^{neg}CD45^{neg}TER119^{neg}CDH5^{pos}KIT^{neg}Gfi1/Gfi1b^{pos}), YS HE marker panel (CD41^{neg} CD45^{neg} TER119^{neg}CD31^{pos}KIT^{pos}Runx1/Gfi1^{pos}), EMP: erythro-myeloid progenitor, LMP: lympho-myeloid progenitor. **d**, Tree dendrogram generated by hierarchical clustering of the sorted populations in (c). Below the dendrogram, the contribution of the different FACS-sorted populations to each cluster is shown. All cells sorted from the YS using the AGM HE marker panel cluster together with YS FACS-ENDO cells. Bottom: heatmap depicting the expression of endothelial (top) and hematopoietic (bottom) genes across clusters K1-K6 **e**, Bar graph depicting the distribution (as percentage) of each FACS-sorted population listed in (c) across clusters K1-K6. **f**, Violin plots depicting the expression of selected endothelial (*Pecam1*, *Cdh5*, *Kdr*, *Procr*) and hematopoietic genes (*Runx1*, *Myb*, *Ptprc*) across clusters K1-K6. Black bars represent the mean expression level.

https://doi.org/10.1038/s44161-025-00740-z

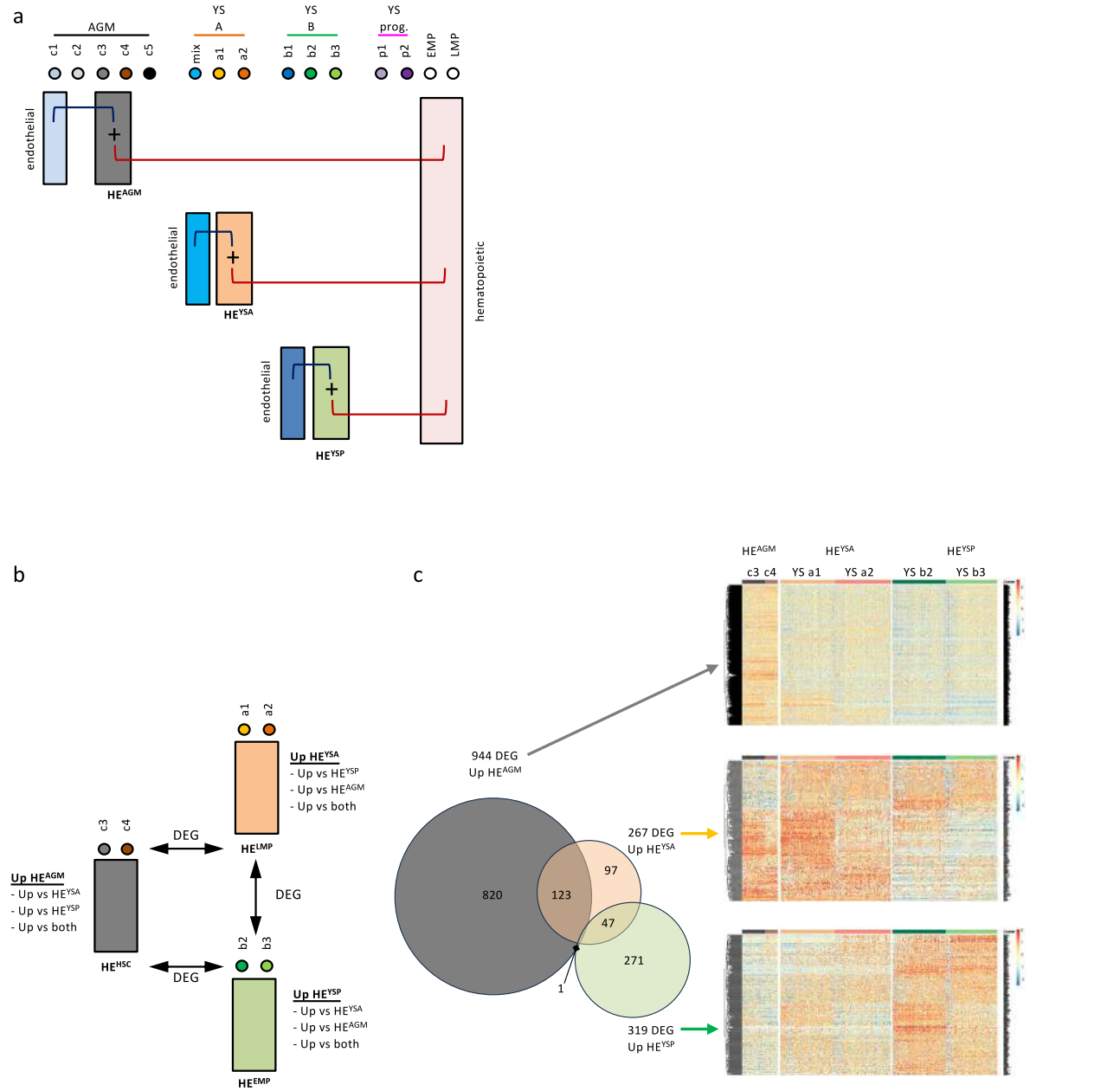


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s44161-025-00740-z

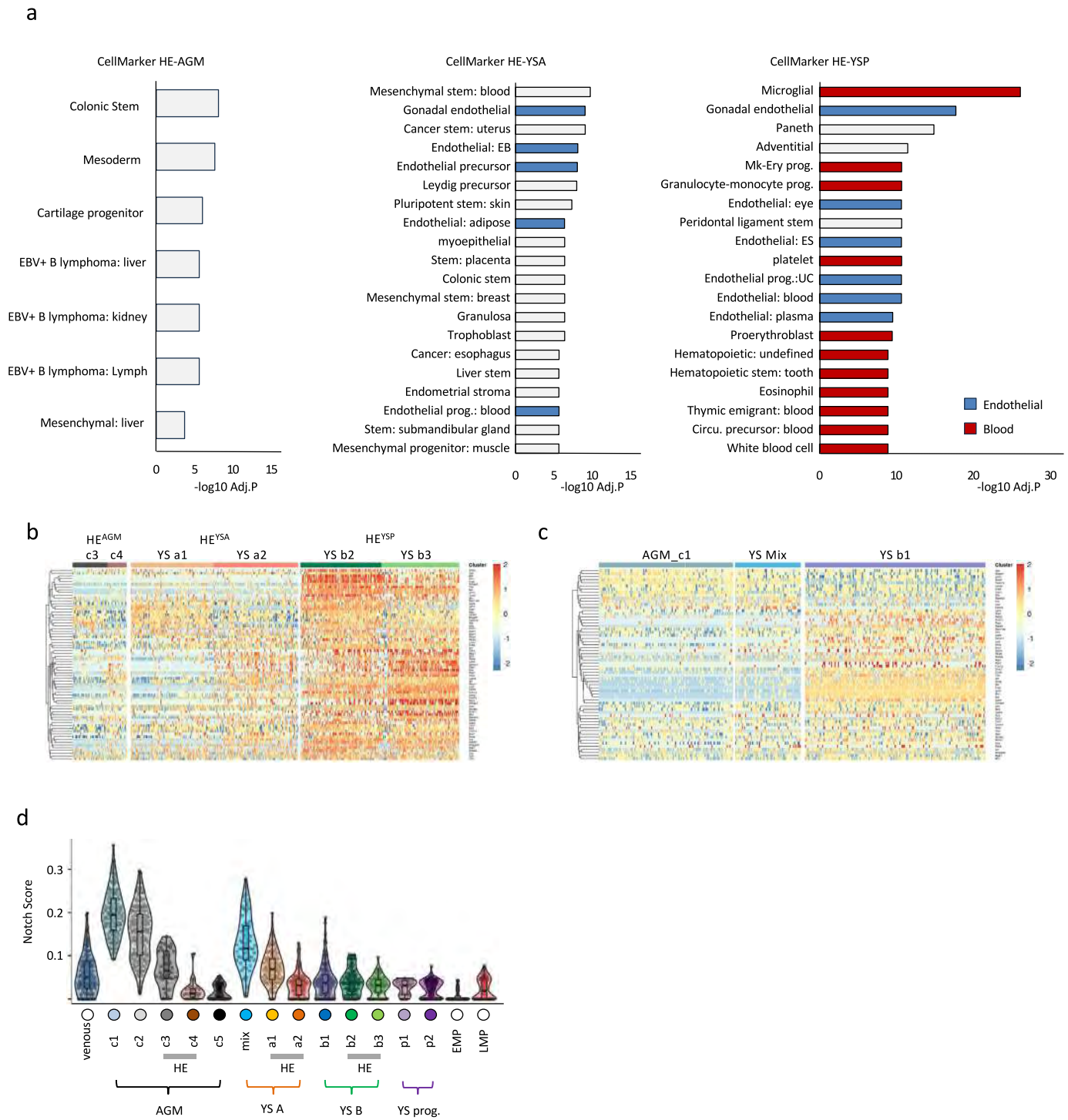
Extended Data Fig. 3 | Two distinct extra-embryonic EHT trajectories. **a**, Violin plots depicting the # of sequencing reads, # of genes detected, # of reads quantified for all scRNA-seq clusters/populations analyzed in this study. Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75^{th} and 25^{th} percentile and whiskers extend to 1.5x the inter-quartile range. The AGM-derived part of the dataset is described in depth in Fadlullah, M.Z.H. et al. *Blood* 139, 343-356, 2022. **b**, Violin plots depicting the expression of candidate genes for immunofluorescence- based analyses of the YS-middle and YS-bottom clusters. **c**, Top: Correlation between transcript expression of *CD24a* and *Vwf* in YS-middle trajectory clusters (mix-m1-m2) and YS-bottom trajectory clusters (b1-b2-b3). Both *Vwf* and *CD24a* are selective for clusters mix, m1 and m2. **d**, Left: Single-cell hematopoietic assays of early YS-middle cells (likely cluster m1) ($KIT^{pos}CD31^{pos}LIN^{neg}LYVE1^{neg}CD24^{pos}MCAM^{pos}Runx1{:}RFP^{pos}$) cultured on OP9 feeder cells for 14 days. The percentage of wells containing proliferating hematopoietic cells is shown. Right: Lineage distribution of the hematopoietic cells shown in the left panel, as determined by flow cytometry for myeloid (GR1, MAC1/CD11b*)*, erythroid (TER119) and lymphoid (CD19) markers. No erythropoiesis was observed. **e**, Left: Single-cell hematopoietic assays of early YS-bottom cells (likely cluster b2) ($KIT^{pos}CD31^{pos}LIN^{neg}LYVE1^{pos}CD24^{neg}MCAM^{pos}Runx1{:}RFP^{pos}$) cultured on OP9 feeder cells for 14 days. The percentage of wells containing proliferating hematopoietic cells is shown. Right: Lineage distribution of the hematopoietic cells shown in the left panel, as determined by flow cytometry for myeloid (GR1, MAC1/CD11b*)*, erythroid (TER119) and lymphoid (CD19) markers. No erythropoiesis was observed. **f**, EMP-fate (8 genes) and LMP-fate (14 genes) signatures were extracted by intersecting pairwise differential gene expression results (EMP vs LMP and P1 vs P2). Fate signatures are listed in Supplementary Table 1.

https://doi.org/10.1038/s44161-025-00740-z



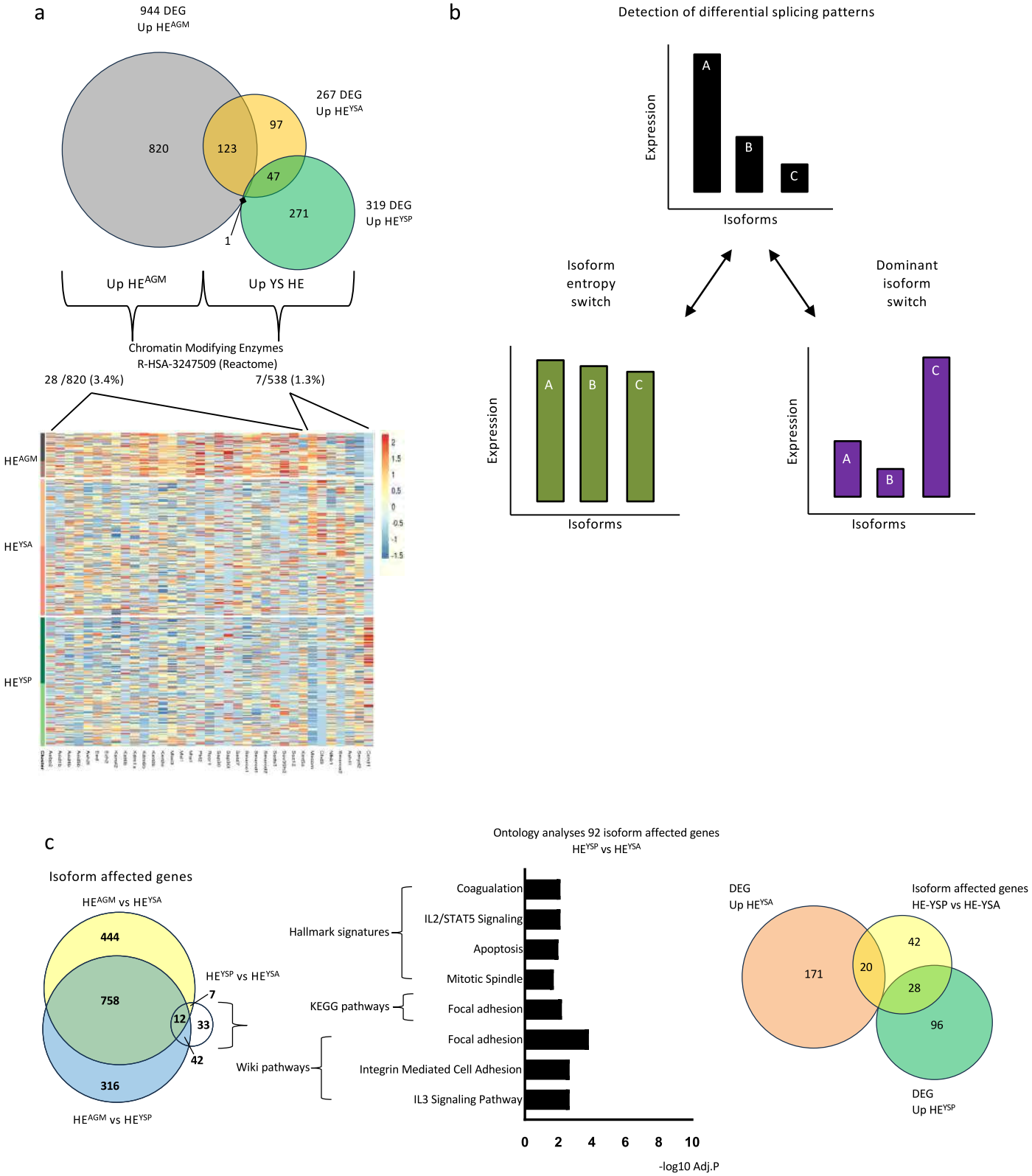
Extended Data Fig. 4 | Shared and Unique HE profiles. a, Schematic depicting which differential gene expression (DEG) lists (log2FC > 1) were used to establish the shared HE profile. **b**, Schematic depicting how unique HE profiles were established using pairwise comparison between the three HE populations. HE^{AGM} (clusters c3 and c4 from the AGM), HE^{YSA} (clusters a1 and a2 from the YS), HE^{YSP} (clusters b2 and b3 from the YS). **c**, Left: Venn diagram depicting the result of pairwise DEG analysis on HE^{AGM}, HE^{YSA} and HE^{YSP}. "Up" indicates a gene is upregulated (log2FC > 1.5) vs at least one other HE. The genes present in each group are depicted in the single-cell heatmaps on the right. Genes up in HE^{YSP} show a high overlap with HE^{AGM} gene hits.

https://doi.org/10.1038/s44161-025-00740-z



Extended Data Fig. 5 | unique HE profiles. a, CellMarker analysis on the DEG between the three HE populations. Top hits in the CellMarker database are shown (capped at 20). Genes upregulated in HE^{YSP} have a strong hematopoietic identity. This is not observed in any of the other HE populations. Adjusted p-values were calculated using Fisher's exact test with Benjamini–Hochberg correction. **b**, Single-cell heatmap of hematopoietic genes, extracted from the HE^{YSP} Cellmarker analysis in (a), across three HE populations. **c**, Single-cell heatmap of hematopoietic gene expression, extracted from the HE^{YSP} CellMarker analysis in (a), across non-HE endothelial populations. Plexus endothelium expresses relatively high levels of hematopoietic genes compared to the other endothelial populations. **d**, Violin plots depicting Notch signature scores across all clusters defined in main Fig. 2b. For reference, AGM-derived venous endothelial cells (left column) and extra-embryonic-derived EMP and LMP populations (right columns) are also included. Embedded boxplots indicate the median (horizontal line), the upper and lower hinges represent the 75th and 25th percentile and whiskers extend to 1.5x the inter-quartile range.

https://doi.org/10.1038/s44161-025-00740-z

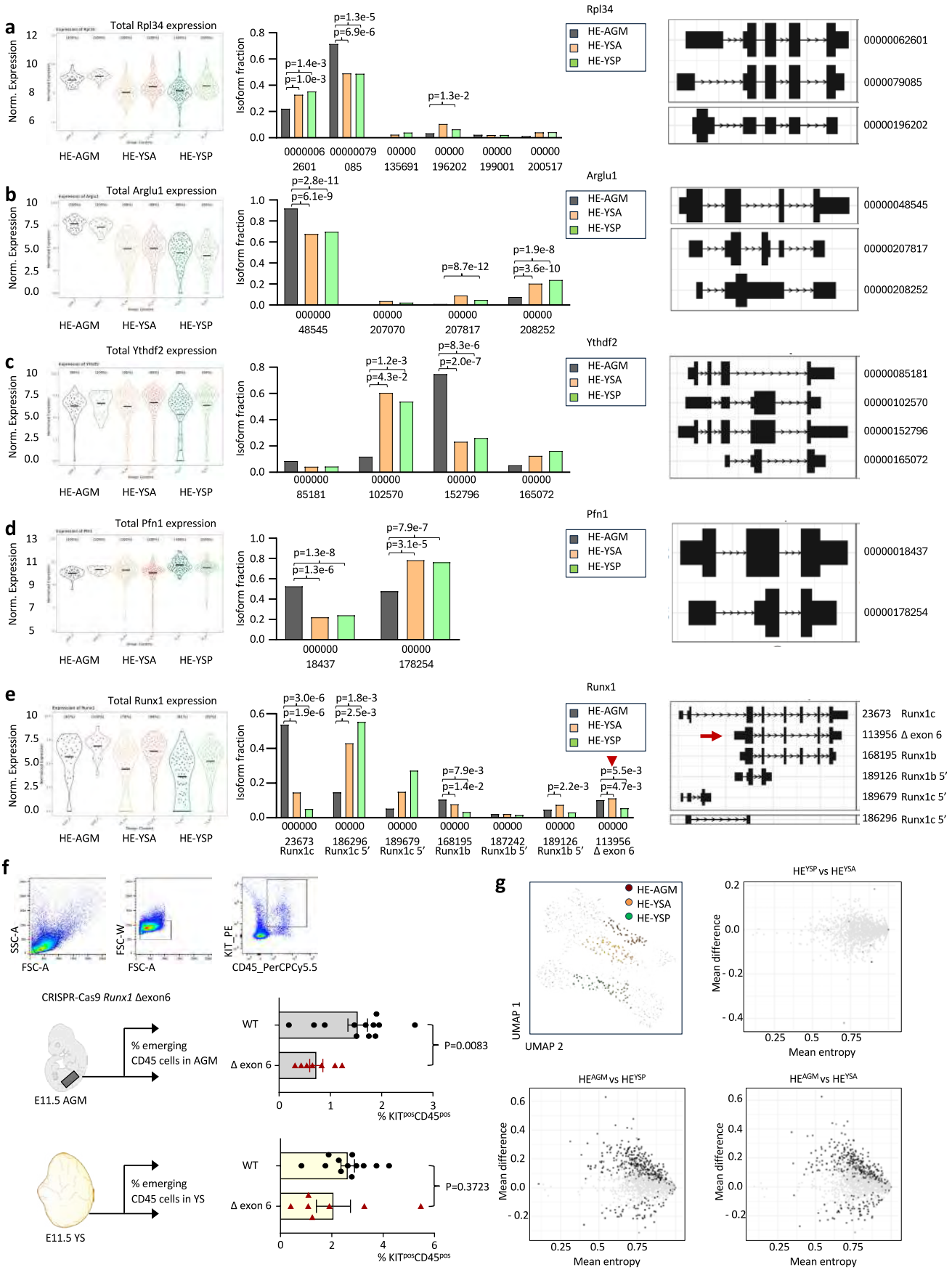


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s44161-025-00740-z

Extended Data Fig. 6 | Enrichment of Chromatin modifiers in HE^{AGM} and isoform level analyses of HE populations. a, Top: Venn diagram depicting the result of pairwise DEG analysis on HE^{AGM}, HE^{YSA} and HE^{YSP}. "Up" indicates a gene is upregulated (> logfc1.5) vs at least one other HE. Bottom single-cell heatmap depicting the genes from Chromatin Modifying Enzymes R-HSA-3247509 (Reactome) that are differentially expressed between the HE populations. **b**, Schematic representation of different isoform usage patterns. Left: schematic representation of isoform Mean Entropy Differences (MED). Right: Schematic representation of dIF (difference in Isoform Fraction) changes **c**, Analyses of isoform differences (entropy and/or dIF) between HE^{AGM} vs HE^{YSP}, HE^{AGM} vs HE^{YSA} and HE^{YSP} vs HE^{YSA}. Left: Ven diagram depicting the intersect of the different comparisons. Only a small set of 94 genes show isoform level differences between the HE^{YSA} vs HE^{YSP}. Middle: Gene ontology analyses across Wikipathways 2024 Mouse, KEGG 2021 human and MSigDB Hallmark 2020 databases. The input gene lists consisted of the 92 genes that showed differential isoform expression (entropy and/or dIF) between HE^{YSA} and HE^{YSP}. All significant hits are shown. Gene lists can be interrogated in Supplementary Table 4. Adjusted p-values were calculated using Fisher's exact test with Benjamini–Hochberg correction. Right: Ven diagram showing the overlap of the isoform affected genes and genes that are differentially expressed between the two YS HE populations (HE^{YSA} and HE^{YSP}).

https://doi.org/10.1038/s44161-025-00740-z



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

https://doi.org/10.1038/s44161-025-00740-z

Extended Data Fig. 7 | The unique isoform landscape of HE^{AGM}. a-e, gene and isoform expression for *Rpl34* (**a**), *Arglu1* (**b**), *Ythdf2* (**c**), *Pfn1* (**d**) and *Runx1* (**e**). Left: violin plot of single cell normalized (total) gene expression in HE^{AGM} (clusters c3-c4), HE^{YSA} (clusters a1-a2) and HE^{YSP} (cluster b2-b3), Middle: isoform usage bar graph depicting all detected isoforms (ENSMUST). Statistical test used was the IsoformSwitchAnalyzeR implementation of the DTU test in the satuRn R package [https://f1000research.com/articles/10-374/v2] (a generalized linear model-based test). Right: schematic of relevant isoforms. The red arrow in (**e**) highlights the *Runx1* isoform that lacks exon 6. **f**, Emerging hematopoietic cells (KIT^{pos}CD45^{pos}) in E11.5 wildtype and CRISPR-Cas9 Runx1 Δ exon 6 embryos identified by flow cytometry. Top, representative analysis flow cytometry plots. Middle, the percentage of KIT^{pos}CD45^{pos} cells in E11.5 AGM regions. Each point represents a single AGM. Bottom, the percentage of KIT^{pos}CD45^{pos} cells in E11.5 Yolk sacs. Each point represents a single yolk sac. WT N = 12, Δ exon 6 N = 7. Bars represent the average + s.e.m. Statistical test: unpaired two-tailed t-test. **g**, Nanopore long-read sequencing of embryonic HE populations. Top right: single cell UMAP depicting three embryonic EHT trajectories as depicted and described in main Fig. 2a-b. The dots indicate 160 cells that have been re-sequenced on the nanopore long-read platform. Right and bottom: Analyses of isoform entropy difference between HE^{YSP} vs HE^{YSA}, HE^{AGM} vs HE^{YSP} and HE^{AGM} vs HE^{YSA} and. Left: scatter plots showing the genes having differential usage pattern for the indicated comparison. Black dots represent genes with significant mean entropy differences (mean difference > 0.1, FDR adjP < 0.05). Statistical test: Wilcoxon signed-rank test, two-tailed. Gray dots represent genes with non-significant changes.

nature portfolio

Corresponding author(s): Michael Lie-A-Ling, Georges Lacaud

Last updated by author(s): Sep 18, 2025

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Zeiss Zen software version 2.3 SP1; IMARIS Viewer software version 9.7.2 (Bitplane), ImageJ/Fiji (versions 2.3.5-2.9.0) and Adobe Photoshop 504 CC 2021. Fiji/ImageJ Cell Counter tool FlowJo (v10.1)
Data analysis	Zeiss Zen software version 2.3 SP1; IMARIS Viewer software version 9.7.2 (Bitplane), ImageJ/Fiji (versions 2.3.5-2.9.0) and Adobe Photoshop 504 CC 2021. Fiji/ImageJ Cell Counter tool FlowJo (v10.1) bcltoFastq conversion (version 2.20.0.422) STAR aligner (version 2.7.9a) R (version 4.1.0) Bioconducter package DropletUtils (version 1.12.1). SingleCellExperiment (version 1.14.1) Seurat (version 4.0.6). Scater (version 1.20.1) buildSNNGraph scanpy workflow (version 1.6.1) 'limma' package (version 3.54.2)

UCell package128 (version 2.2.0)
splice aware aligner Salmon 1.4.0
IsoformSwitchAnalyzeR 2.01.07
SplicingFactory 1.8.0
DESeq2 1.40.2
MinionQC.R 1.4.2131
MinKNOW 23.04.6
Guppy 6.5.7
Minimap2 2.26
IGV Desktop App 2.17.0
ChIPpeakAnno version 3.20.1
https://github.com/zakiF/PublishedPapers/tree/master/YolkSac_AGM
https://github.com/RASellers-CRUK/GL_IsoformAnalysis

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The following published datasets were used in this study:
GEO
GSE137116
GSE139389
GSE167588
GSE173833
GSM5281418
GSE57251
GSM1377856
GSM1377857
GSM1377858
GSE22178
GSE69101
GSM1692809
GSM1692853
GSM1692854
GRCm39 M33 mouse reference
GRCm39 M32 mouse reference
mm10 reference

For the data generated in this study:
Gene expression data can be queried at https://shiny.cruk.manchester.ac.uk/AGM_YS_dataset_final/. Raw data is deposited in GEO accession GSE274544 and GSE309071. Three source data files accompany this manuscript
NCVR_2024_09_0957A_extended_source.xlsx, NCVR_2024_09_0957A_main_source.xlsx and NCVR_2024_09_0957A supplemental_tables_1.xlsx

Code used:
https://github.com/zakiF/PublishedPapers/tree/master/YolkSac_AGM and https://github.com/RASellers-CRUK/GL_IsoformAnalysis
https://github.com/RASellers-CRUK/GL_IsoformAnalysis

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. For in vivo transplantation experiments, five mice per group were used.
Data exclusions	All replicate data were used for statistical analysis and no data were excluded.
Replication	All data were replicated at least one time (N=2 biological experiments)
Randomization	Samples were randomly allocated into experimental groups and processed in no particular order to minimize batch effects.
Blinding	The investigators were not blinded to group allocation during data collection and analysis.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used

Antibody Clone Manufacturer Application dilution Experiment
B220 APC RA3-6B2 eBioscience FACS 1:200 Co-culture readout
B220 eF450 RA3-6B2 eBioscience FACS 1:400 Co-culture readout
B220 FITC RA3-6B2 Invitrogen FACS 1:400 YS lineage staining
B220 Biotin RA3-6B2 eBioscience FACS 1:200 FL HSC staining, FACS sort LT-HSC
CD16/CD32 93 Invitrogen FACS 1:100 Fc blocking
CD16/CD32 AF700 93 eBioscience FACS 1:200 YS EMP/LMP staining
CD19 APC_eF780 eBio1D3 eBioscience FACS 1:200 Co-culture readout
CD19 PE-Cy7 eBio1D3 eBioscience FACS 1:200 terminal BM samples
CD24a PE 30-F1 eBioscience FACS 1:200 YS EHT staining
CD24a PerCPCy5.5 M1/69 BD FACS 1:200 YS EHT staining
CD25 Biotin eBio7D4 eBioscience FACS 1:200 Co-culture readout
CD3 APC 145-2C11 eBioscience FACS 1:200 terminal BM samples
CD3 FITC 145-2C11 eBioscience FACS 1:400 YS lineage staining
CD3 biotin 145-2C11 eBioscience FACS 1:200 FL HSC staining, FACS sort LT-HSC
CD31 PECy7 390 BioLegend FACS 1:200 YS EHT, AGM-HSC staining
CD41 APC MWReg30 eBioscience FACS 1:200 AGM /YS lineage staining
CD41 Biotin MWReg30 eBioscience FACS 1:400 AGM /YS lineage staining
CD41 FITC MWReg30 eBioscience FACS 1:400 AGM /Ys lineage staining
CD41 PE MWReg30 eBioscience FACS 1:200 YS EHT staining
CD41 PECy7 MWReg30 eBioscience FACS 1:200 YS EMP/LMP staining
CD45 Biotin 30-F11 Invitrogen FACS 1:400 AGM /YS lineage staining
CD45 FITC 30-F11 Invitrogen FACS 1:400 AGM /YS lineage staining
CD45 PerCPCy5.5 30-F11 eBioscience FACS 1:400 Co-culture readout, AGM-HSC staining

CD45.1 APC- Efl 780 A20 eBioscience FACS 1:400 peripheral blood analysis
CD45.2 PerCp 5.5 104 eBioscience FACS 1:400 peripheral blood analysis
CD127 eF450 A7R34 eBioscience FACS 1:100 YS EMP/LMP staining
CD127 PE A7R34 Invitrogen FACS 1:100 YS EMP/LMP staining
CD48 APC, FACS sort LT-HSC HM48-1 eBioscience FACS 1:200 FL HSC staining
CD150 PE-Cy7 TC15-12F12.2 BioLegend FACS 1:400 FACS sort LT-HSC
CD150 PE TC15-12F12.2 BioLegend FACS 1:400 FL HSC staining
CD201 (EPCR,ProcR) APC eBio1560 eBioscience FACS 1:200 AGM-HSC
c-Kit APC 2B8 eBioscience FACS 1:400 AGM/YS EHT staining
c-KIT PE 2B8 eBioscience FACS 1:400 AGM-HSC / AGM staining
c-Kit APC eF780 2B8 eBioscience FACS 1:200 AGM/YS EHT, AGM-HSC and FL HSC staining
c-Kit SB436 2B8 eBioscience FACS 1:600 AGM/YS EHT staining
Ly-6G / Gr1 APC RB6-8C5 eBioscience FACS 1:400 Co-culture readout
Ly-6G / Gr1 FITC RB6-8C5 eBioscience FACS 1:400 YS lineage staining
Ly-6G / Gr1 PECy7 RB6-8C5 eBioscience FACS 1:200 Co-culture readout
LYVE1 eF660 ALY7 eBioscience FACS 1:10000 YS EHT staining
Ly-6G / Gr1 biotin RB6-8C5 eBioscience FACS 1:200 Fetal liver HSC, FACS sort LT-HSC
CD11b / Mac1 APC M1/70 eBioscience FACS 1:400 Co-culture readout
CD11b / Mac1 PE M1/70 eBioscience FACS 1:200 Co-culture readout
CD146 / MCAM Biotin ME-9FI BD FACS 1:200 YS EHT staining
CD146 / MCAM BV786 ME-9FI BD FACS 1:200 YS EHT staining
NK1.1 PerCpCy5.5 PK136 eBioscience FACS 1:200 Co-culture readout
Ly-6A/E (Sca-1)-FITC E13-161.7 BD FACS 1:200 AGM-HSC, Fetal liver HSC
Streptavidin BV421 Biolegend FACS 1:400 AGM/YS EHT staining
Streptavidin PerCP5.5 eBioscience FACS 1:200 Fetal liver HSC
SYTOX Green Invitrogen FACS 1:1000 Viability dye
TER119 Biotin TER-119 Biolegend FACS 1:400 AGM /YS lineage staining, FACS sort LT-HSC
TER119 BV650 TER-119 Biolegend FACS 1:100 Co-culture readout
TER119 FITC TER-119 Biolegend FACS 1:400 AGM /YS lineage staining
TER119 PE TER-119 Biolegend FACS 1:200 Co-culture readout
Thy1.2 (CD90.2) 50600 53-2.1 Invitrogen FACS 1:200 Co-culture readout
Lyve1 AF2125 R&D Systems Whole mount IF 1:200 detection of HE-EMP
Runx1+Runx2+Runx3 EPR3099 Abcam Whole mount IF 1:400 detection of HE
GFP A10262 Invitrogen Whole mount IF 1:500 detection of Vwf:GFP reporter
anti goat Alexa Fluor Plus 647 A32849 Invitrogen Whole mount IF 1:500 detection of LYVE1
anti rabbit Alexa Fluor 555 A32794 Invitrogen Whole mount IF 1:500 detection of RUNX
anti chicken Alexa Fluor 488 A78948 Invitrogen Whole mount IF 1:500 detection of GFP

Validation

All antibodies used have been validated by the manufacturers for their respective applications.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	OP9 cells (mouse bone marrow stromal cell line; ATCC® CRL-2749™) were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA).
Authentication	not done
Mycoplasma contamination	not tested
Commonly misidentified lines (See ICLAC register)	NA

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	mouse embryos E9 - E16, were genereated from crosses between reporter male (C57BL/6JOlaHsd) mice and wt female Hsd:ICR (CD-1®) mice. For transplantations NSGS (NOD.Cg-Prkdcscid Il2rgtm1Wjl Tg(CMV-IL3,CSF2,KITLG)1Eav/MloySzJ) mice (CD45.1) aged 8-12 weeks were used.
Wild animals	No
Reporting on sex	Embryos were not sexed. adult transplant recipient were female
Field-collected samples	na
Ethics oversight	Mouse work was performed in accordance with the United Kingdom Animal Scientific Procedures Act (ASPA) 1986. Animal experiments performed at the Cancer Research United Kingdom Manchester Institute (CRUK-MI) were approved by the Animal

Welfare and Ethics Review Body (AWERB) of the CRUK-MI. Experiments performed at the University of Oxford were approved by the Oxford Clinical Medicine Ethical Review Committee.

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

Plants

Seed stocks	na
Novel plant genotypes	na
Authentication	na

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	For the single cell RNA-seq and in vitro functional assays, dissected yolk-sacs were digested in a mix of Collagenase IV (2 mg/ml, Worthington) and DNase I (200 U/ml, Calbiochem) at 37 °C for 15 min. The dissociated cells were pelleted (300 g for 5 min at 4°C) and resuspended in phosphate buffered saline containing 10% fetal bovine serum (10% FBS in PBS) and further processed for FACS-analyses/sorting. E16.5 Fetal livers, dissected livers were crushed with the end of a 1ml syringe through a 40um cell strainer into IMDM+10% FBS. Dissected AGM tissues at E10.5 were finely chopped and the obtained fragments were digested in a mix of Collagenases IV (2 mg/ml, Worthington) and DNase I (200 U/ml, Calbiochem) at 37 °C with gentle agitation for 15 min. The dissociated cells were centrifuged at 300 g for 6 min and resuspended in PBS supplemented with 10% FBS for subsequent processing. All flow cytometry analyses was performed on a BD LSRFortessa™ X-20 Cell Analyzer (BD Biosciences) or Novocyte Quanteon (Agilent). All cell sorting was performed on a BD FACSAria™ III Cell Sorter (BD Biosciences). Antibodies used for FACS are listed in table 5. For scRNA-seq cells were directly sorted into lysis buffer and snap-frozen before further processing. FlowJo software (BD Biosciences) was used to analyze all FACS data.
Instrument	BD LSRFortessa™ X-20 Cell Analyzer (BD Biosciences). BD FACSAria™ III Cell Sorter (BD Biosciences). Novocyte Quanteon (Agilent).
Software	BD FACSDiva™ Software, FlowJo v10
Cell population abundance	see manuscript
Gating strategy	see manuscript. extended figure 1

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

Fetal-restricted hematopoietic progenitors arise from hemogenic endothelium in vitelline and umbilical arteries

Cristiana Barone[1,*], Giulia Quattrini[1,*], Alessandro Muratore[1], Giorgio Anselmi[2], Yurim Park[2], Naeema T. Mehmood[2], Elena Morganti[3], Roberto Orsenigo[1,†], Filipa Timóteo-Ferreira[1], Anna Cazzola[1,$], Arianna Patelli[1], Thea Milanesi[1], Bianca Nesti[3,4], Francisca Soares-da-Silva[5-7], Matthew Nicholls[2], Gloria Zambelli[1], Mario Mauri[1], Silvia Bombelli[1,8], Sofia De Marco[1], Deborah D'Aliberti[1], Silvia Spinelli[1], Veronica Bonalume[9], Alison Domingues[9,#], Mahdieh Naghavi Alhosseini[1], Gianluca Sala[1], Arianna Colonna[1], Elisabetta D'Errico[1], Cristina D'Orlando[1], Cristina Bianchi[1], Roberto A. Perego[1], Raffaella Meneveri[1], Ana Cumano[5-7], Silvia Brunelli[1,10], Marella F.T.R. De Bruijn[2], Andrea Ditadi[3], Alessandro Fantin[9], Rocco Piazza[1,10], Emanuele Azzoni[1,10*]

[1]*School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy*

[2]*MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK*

[3]*San Raffaele Telethon Institute for Gene Therapy, IRCCS San Raffaele Scientific Institute, Milan, Italy*

[4]*Vita-Salute San Raffaele University, Milan, Italy.*

[5]*Unit of Lymphocytes and Immunity, Immunology Department, Institut Pasteur, Paris, France*

[6]*INSERM U1223, Paris, France*

[7]*Université Paris Cité, Cellule Pasteur, Paris, France*

[8]*Fondazione Human Technopole, Milan, Italy.*

[9]*Department of Biosciences, University of Milan, Via G. Celoria 26, 20133, Milan, Italy*

[10]*Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy*

** equal contribution.*

[†] Current address: Biomedical Research in Melanoma-Animal Models and Cancer Laboratory, Vall d'Hebron Research Institute (VHIR), Vall d'Hebron Hospital Barcelona-UAB, Barcelona, Spain

[$]Current address: Department of Biology, University of Copenhagen, Copenhagen, Denmark

[#]Current address: Université Paris Cité, INSERM, Innovations thérapeutiques en hémostase, F-75006 Paris, France

*corresponding author. E-mail: emanuele.azzoni@unimib.it

BiorXiv, 2024

Nature Cardiovascular Research, under revision

Summary

Embryonic hematopoiesis involves successive waves of progenitors from distinct anatomical sites, but the origins and contributions of early hematopoietic stem/progenitor cells (HSPCs) remain incompletely defined. We employed genetic fate mapping in mice to temporally label hemogenic endothelium (HE) subsets and track their progeny. Here we show that a wave of fetal-restricted HSPCs arise from HE in the vitelline and umbilical arteries between embryonic days 8.5 and 9.5, preceding definitive hematopoietic stem cell (HSC) emergence. Lineage tracing coupled with single-cell transcriptomics and functional assays revealed these progenitors are transient, distinct from erythro-myeloid progenitors, and contribute extensively to fetal lympho-myelopoiesis but decline postnatally. Our findings reveal a previously unrecognized early HE wave as a key source of fetal-restricted HSPCs, refining the spatial-temporal understanding of layered hematopoiesis and informing developmental origins of blood cell diversity.

Keywords:

Fetal Hematopoiesis; Hematopoietic Stem Cells; Embryo; Yolk Sac; Aorta-Gonad-Mesonephros; Single-cell RNA sequencing

Introduction

The vertebrate embryonic hematopoietic system develops through a series of overlapping waves of blood progenitors, each with progressively broader lineage potential [1, 2]. Although the phenomenon of 'layered hematopoiesis'—the sequential emergence of distinct blood cell populations—is highly conserved across species, its detailed analysis in model organisms has been challenging due to the temporal and spatial overlap of these waves and their multiple anatomical sources. Furthermore, the identification of embryonic blood progenitor cells has been complicated by the extensive sharing of surface markers among these populations, and only recently has their heterogeneity begun to be unraveled through advances in single-cell methodologies [3].

Adult repopulating Hematopoietic Stem Cells (HSCs) are firstly and autonomously generated in the aorta-gonad-mesonephros (AGM) region starting from E10.5 in the mouse [4]. HSCs originate from a specialized population of endothelial cells termed hemogenic endothelium (HE) in the major embryonic arteries [5-8] and mature through a hierarchy of pro- and pre-HSC intermediates [9, 10]. Starting from E12, HSCs colonize the fetal liver (FL) where they are thought to expand in numbers [11], and toward the end of gestation relocate to the bone marrow (BM) where they will reside throughout adult life. The BM niche, however, does not acquire robust HSC support capability until after birth [12].

Prior to HSC generation and following the early onset of primitive hematopoiesis, several waves of oligopotent progenitors begin emerging at E8.25 [2, 13], initially from HE in the yolk sac (YS) [14]. Among HSC-independent progenitors, Erythro-Myeloid Progenitors (EMPs) generate tissue resident macrophages persisting until adult life [15, 16]. EMPs were also reported to significantly contribute to fetal erythropoiesis [17] and fetal innate lymphoid cells [18]. Although EMPs can generate multiple myeloid lineages [19], their physiological contribution to fetal and postnatal hemopoiesis is still unclear. Immune-restricted Lympho-Myeloid Progenitors (LMPs) emerge in the YS slightly later than EMPs and were shown to take part in fetal lympho-myelopoiesis, even though for a limited time window and contributing less than 20% of myeloid cells at E14.5 [20]. Despite conclusive evidence that YS-derived HSC-independent B and T cell progenitors exist and persist to adulthood [21-23], some controversy still remains regarding the identity of the first progenitors responsible for the colonization of the fetal thymus, and in particular whether they originate from HSCs or not [24, 25].

HSC-independent progenitors are necessary and sufficient to sustain fetal life until the end of gestation [26]. Indeed, recent work showed that HSCs exert a limited contribution to pre-natal hematopoiesis [27, 28]. Lineage tracing identified embryonic multipotent progenitors (eMPPs) appearing concomitantly to definitive HSCs, that significantly contribute to fetal and adult multi-lineage hematopoiesis [28-30]. Although the origin of eMPPs appears to be HSC-independent [28, 31], it is not clear

when and where they emerge during development. The existence of fetal HSCs with characteristics distinct from adult HSCs has also been suggested [32, 33], however, as for eMPPs, a prospective identification of these progenitors, which would allow localization in their niche of emergence, is currently not possible [34]. Moreover, the true extent of their contribution to fetal and adult hematopoiesis needs further clarification.

To genetically label and trace discrete subsets of HE, we took advantage of well-established conditional fate-mapping strategies in mouse. We found that a wave of fetal-restricted hematopoietic stem/progenitor cells (HSPCs) emerges from HE between E8.5 and E9.5, before the onset of adult-type HSCs, and acts as a major driver of fetal lympho-myelopoiesis. Through a combination of whole-mount imaging and single cell transcriptomics, we localized the initial emergence of HSPCs belonging to this wave to the hematopoietic clusters of vitelline and umbilical arteries (VU). Moreover, we show that fetal-restricted HSPCs are not endowed with long-term multilineage engraftment potential but instead represent a heterogeneous subset of hematopoietic progenitors poised for differentiation, likely including eMPPs.

Results

HE lineage tracing identifies a population of fetal-restricted hematopoietic progenitors

All hematopoietic cells, with the possible exception of some primitive erythrocytes, originate from *Cdh5*+ HE [1]. To differentially label embryonic hematopoietic waves and systematically trace their contribution during fetal and adult hematopoiesis, we employed a well validated pulse-chase approach using tamoxifen-inducible *Cdh5-CreER^{T2}* mice [35], together with reporter lines selected for their suitability to different applications (*R26^{tdTomato}* / *R26^{zsGreen}* or *R26^{EYFP}*) (**Figure S1A**). This strategy benefits from the use of a single Cre line, thereby avoiding bias from cell type-specific promoters. To achieve precise temporal control, we used (Z)-4-hydroxytamoxifen (4-OHT), which has a short *in vivo* half-life (<3 hours), reaches peak serum levels rapidly after administration and is cleared to undetectable within 12-hours, as shown by mass spectrometry [36]. In contrast, tamoxifen requires hepatic metabolization into 4-hydroxytamoxifen (4-OHT), a process that takes approximately 6–12 hours in vivo; tamoxifen itself remains detectable in serum for up to 48 hours post-administration.[37].

We first evaluated the labeling of YS EMPs (Ter119- Kit+ CD41^{low} CD16/32+)[19]. In *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos, E9.5 and E10.5 YS EMPs were found labeled at high efficiency with 4-OHT activation at both E7.5 and E8.5 (**Figure S1B,C**). EMP labeling was also confirmed by whole-mount imaging of E9.5 YS (**Figure S1D**), which identified no significant differences in the number of labeled Kit+ clusters when traced at the two activation time points (**Figure S1E**). Consistent with this, brain microglia, which originates from YS EMPs [38], was found highly labeled with both 4-OHT at E7.5 and E8.5, in the E16.5 fetus and in the adult (**Figure S1F,G**). In contrast, as expected, 4-OHT at E10.5 did not label microglia (**Figure S1G**).

We analyzed the labeling of LMPs (CD31+ Kit+ CD45+) in the E10.5 AGM and in YS including VU connecting the yolk sac to the embryo proper (YS+VU). While in the AGM LMP labeling was partial with both activation time points, 4-OHT at E8.5 traced the majority of LMPs in the YS+VU (**Figure S2A,B**) , consistent with LMP emergence at/around E9.5 in the YS [20]. In the E11.5 FL, LMP (Lin- Kit+ CD45+ Flt3+ IL7Rα+) recombination frequency was low with 4-OHT at E7.5 and increased with activation at E8.5 (**Figure S2C,D**). Interestingly, the percentage of traced CD45+ Kit+ hematopoietic progenitors in the E11.5 FL doubled when 4-OHT was delivered at E8.5 compared to E7.5 (**Figure S2E**), suggesting that part of them originate independently from EMPs.

Next, we investigated the extent of labeling of immunophenotypic type I (CD31+ Kit+ CD41^{low} CD45- CD43+ CD201+) and type II pre-HSCs (CD31+ Kit+ CD41^{low} CD45+ CD43+ CD201+)[9, 39]

in AGM and YS+VU of *Cdh5-CreER^{T2}::R26$^{tdTomato/zsGreen}$* E11.5 embryos (**Figure 1A**). 4-OHT at E7.5 labeled a minority of type I and type II pre-HSCs in both AGM (**Figure 1B,C**) and YS+VU (**Figure 1D**). Type I pre-HSC labeling was also low at E10.5 at this activation time (**Figure S2A,B**). In contrast, both E8.5 and E10.5 activations yielded significant pre-HSC labeling at E11.5. While type I pre-HSCs were captured at a higher frequency with 4-OHT at E10.5 in the AGM, labeling of more mature type II pre-HSCs was significantly higher (~60%) in both AGM and YS+VU with 4-OHT at E8.5, as opposed to an apparently complementary subset (~40%) labeled with 4-OHT at E10.5 (**Figure 1B-D**). Strikingly, hematopoietic progenitors other than pre-HSCs (CD31+ Kit+ CD41^{low} CD45+/- CD43+ CD201-) exhibited a much more selective labeling pattern and were almost exclusively labeled by the E8.5 activation in AGM and YS+VU (**Figure 1B-D**). When dissected and analyzed separately from each other, we did not detect any difference in the labeling frequencies of any progenitor subset in the YS or the VU (**Figure S2F**); therefore, we kept these tissues together for the remainder of this study.

We then tested the labeling of fetal and adult phenotypic HSCs (Lin- Kit+ Sca1+ CD150+ CD48-) (**Figure 1E**). E14.5, E16.5 FL and adult bone marrow (BM) HSCs were extensively labeled with 4-OHT at E10.5, but not at E7.5 (**Figure 1F-H**), confirming a previous report [40]. Interestingly, 4-OHT at E8.5 labeled E14.5 FL HSCs with variable efficiency (58% on average; **Figure 1F**) not dependent on the specific reporter line (**Figure S2G**), but decreasing to an average of 25% at E16.5 (**Figure 1G**) and 10% in the adult BM (**Figure 1H**), raising the possibility of the existence of a wave of "fetal-restricted" HSCs, as previously suggested [32]. A similar labeling pattern, however, was also observed in Lin- Kit+ Sca1+ (LSK) progenitors (**Figure S2H-J**). Non-HSC hematopoietic progenitors (LK, Lin- Kit+ Sca1-), comprising granulocyte-macrophage progenitors (GMP), common myeloid progenitors (CMP) and megakaryocyte-erythroid progenitors (MEP) labeling was at the highest with 4-OHT at E8.5 in E14.5 FL (**Figure S2K**), whereas at E16.5 and in adult BM they were mostly labeled with 4-OHT at E10.5 (**Figure S2L,M**).

Taken together, these data suggest that HE lineage tracing in the *Cdh5-CreER^{T2}* model (4-OHT at E8.5) may capture a putative population of HSPCs largely restricted to fetal life. In contrast, 4-OHT pulses at E7.5 or E10.5 respectively label either EMPs or adult-type HSCs. Therefore, the same genetic model can be used to study the relative fetal and adult contributions of three sequential waves of HE in an unbiased way.

Fetal lympho-myelopoiesis is largely contributed from hematopoietic progenitors originating from E8.5-E9.5 HE

To examine the fetal lympho-myeloid contribution of the three HE waves, we analyzed those lineages in *Cdh5-CreER^{T2}::R26^{tdTomato}* E16.5 FL and thymus. 4-OHT activation at E7.5 yielded an average labeling of only 35% of F4/80+ CD11b^{low} macrophages, less than 10% B cells, 10-15% T cells and 5% F4/80$^{low/-}$ CD11b+ myeloid cells in the E16.5 FL and thymus (**Figure 2A,B**). E10.5 activation resulted in labeling of 30% of B and myeloid cells (**Figure 2A**), a similar contribution to T cells, with higher labeling (30-40%) in less differentiated thymocytes (DN1 and DN2) (**Figure 2B**) and negligible labeling in macrophages. Strikingly, the highest labeling frequencies in all observed lineages were detected with 4-OHT activation at E8.5, yielding on average 70% of labeled B and myeloid cells in the E16.5 FL (**Figure 2A; Figure S3A**) and 50-60% of labeled E16.5 fetal thymocytes (**Figure 2B; Figure S3B**). In contrast to E10.5 activation, double-positive (CD4+ CD8+ DP) T cells displayed the highest labeling frequency, and embryonic γδ T cells were preferentially labeled by 4-OHT administration at E8.5, showing an approximately four-fold higher labeling efficiency compared to E10.5 (**Figure 2B**). Thus, our data is consistent with the previously reported model suggesting the existence of two separate waves of thymus-settling progenitors (TSP) [41].

Post-natal analysis confirmed absence of peripheral blood (PB) labeling (<5%) in all lineages with activation at E7.5, both at 21 days (**Figure S3C**) and 2 months (**Figure 2C,D**), while E10.5 activation resulted in high levels of recombination in all lineages (**Figure 2C,D; Figure S3C**). Conversely, activation at E8.5 showed highly variable labeling at 21 days (average 30-35%) (**Figure S3C**), which exhibited a decreasing trend at 2 months (20-25%; **Figure 2C,D; Figure S3D**). Notably, comparison of fetal and postnatal stages revealed divergent dynamics for the two labeling windows. Labeling obtained with 4-OHT activation at E8.5, which was highest at E16.5, declined markedly after birth, indicating that progeny of this HE wave progressively lose representation in postnatal hematopoiesis (**Figure 2D**). In contrast, labeling following 4-OHT activation at E10.5 increased from E16.5 to postnatal stages (**Figure 2D**), consistent with the gradual takeover of hematopoiesis by adult-type definitive HSCs. Together, these reciprocal trends indicate a developmental handover between a predominantly fetal hematopoietic program emerging from E8.5–E9.5 HE and a later-arising program specified from E10.5 HE that sustains postnatal blood production.

These data show that the hematopoietic wave that emerges from HE between E8.5 and E9.5 is a major contributor to fetal, but not adult, lympho-myelopoiesis. As HE labeling at E8.5 marks phenotypic progenitors and pre-HSCs (**Figure 1**) and similar labeling dynamics were seen for LK, LSK and HSC (**Figure 2A-D, Figure S2H-M**), this wave of HE likely contains the precursors that generate a pre-constituted hierarchy of fetal-restricted hematopoietic progenitors, including fetal HSCs, HSC-independent eMPPs and other progenitors [28-30, 34, 42]. Therefore, we will collectively refer to these as "fetal-restricted HSPCs".

Fetal-restricted HSPCs first emerge from HE of the vitelline and umbilical arteries

To obtain insight into the dynamics of fetal-restricted HSPC generation, we investigated sites of hematopoietic emergence using whole-mount confocal imaging. As mentioned, Kit+ CD31+ hematopoietic clusters in the E9.5 YS, corresponding to EMPs, were equally labeled by 4-OHT activation at E7.5 or E8.5 (**Figure S1D,E**). CD31+ Kit+ hematopoietic clusters in the dorsal aorta (DA), thought to contain pre-HSCs [43], peak at E10.5 [44]. However, the first intra-embryonic Kit+ hematopoietic clusters appear within the portion of the vitelline artery (VA) most proximal to the DA [9, 45, 46]. Although the majority of these clusters are thought to contain progenitor cells other than pro-HSC [9], vitelline and umbilical arteries are known to represent sites of pre-HSC emergence [7, 47]. Importantly, because of the lack of specific ways to trace these early progenitors, their contribution has never been analyzed in an unperturbed system. Remarkably, few Kit+ hematopoietic clusters in the *Cdh5-CreER^{T2}::R26^{EYFP}* E9.5 VA were labeled by 4-OHT at E7.5, but, in contrast, their majority was labeled by the E8.5 activation (**Figure 3A,B**), correlating with the emergence of fetal-restricted HSPCs. These observations indicate that such progenitors first emerge within the VA. Flow cytometry analysis of E9.5 hematopoietic progenitors other than EMP (non-EMPs, Kit+ CD41^{low} CD16/32-) showed absence of differential labeling in the YS, but significantly higher labeling within the caudal part of the embryo (CP) when traced at E8.5 (**Figure S4A**), consistent with imaging data and an identity of VA clusters independent from EMPs. Next, we evaluated labeling of hematopoietic clusters in the VA and umbilical artery (UA) of E10.5 embryos (CD31+ Kit+ or CD31+ Runx1+ round-shaped cells) and compared it to the DA. 4-OHT activation at E7.5 yielded low labeling frequency of both VU and aortic clusters (**Figure 3C-F**). Conversely, 4-OHT at E8.5 resulted in high levels of recombination in VU clusters, but significantly lower in DA clusters (**Figure 3C-F**). In line with these results, differential labeling of non-EMPs by flow cytometry was also observed at E10.5 in the AGM (**Figure S4B**). Interestingly, HE labeling was low in the E10.5-E11.5 AGM at all activation time points, as assessed by whole mount imaging (CD31+ Runx1+ flat-shaped cells; **Figure 3E,F**) and flow cytometry (CD31+ Kit+ CD41- CD45- CD43-) (**Figure 3G, Figure S4C**). Notably, neither E7.5 or E8.5 activation resulted in appreciable recombination within the ventral aortic endothelium (**Figure 3E**), consistent with the lack of tracing of adult BM HSCs (**Figure 2D**) arising in this location [48]. AGM non-hemogenic endothelium (CD31+ Kit- CD41- CD45- CD43-) was not labeled (4-OHT at E7.5) or labeled at low frequency (<10%; 4-OHT at E7.5 or E8.5) (**Figure 3G**).

Previous *ex vivo* culture assays revealed the presence of pre-HSCs in the YS at E10.5, but not at E9.5 [9, 45, 49]; from E11.5, YS pre-HSCs seem to decline [10, 11, 49]. Whole-mount analysis of E10.5 YS identified the presence of large Kit+ hematopoietic clusters, which were absent at E9.5. Within the YS, these clusters exclusively localized in the VA and its ramifications (which form a continuous connection with the intra-embryonic portion of the VA), and were previously documented to express Ly6a [46] and Hlf [50], markers associated with HSC activity. In *Cdh5-CreER^{T2}::R26^{zsGreen}* YS, the majority of cells in these clusters were not labeled with activation at E7.5 but, instead, were consistently labeled with 4-OHT at E8.5 (**Figure S4D,E**). In line with this and with EMP labeling in our system, while E7.5 activation identified hematopoietic clusters only in the YS vascular plexus, E8.5 activation labeled clusters in both plexus and YS arteries (**Figure S4F,G**). Similar to what already observed in the AGM, the levels of labeling of E10.5 and E11.5 YS HE were undetectable to low at all activation time points (**Figure S4G,H**).

Overall, these results suggest that fetal-restricted HSPCs first emerge from the HE of the VA and the UA. Whereas at E9.5 the location of the hematopoietic clusters likely to contain these progenitors is solely intra-embryonic, from E10.5 these clusters are also found within the major arteries of the YS. The lack of labeled HE already 24-hour post-activation suggests that endothelial-to-hematopoietic transition (EHT) events take place *in vivo* within a short time window of <12 hours, similar to what observed *ex vivo* [6] and that the entirety of labeled HE undergoes EHT, consistently with it being hematopoietic-committed and devoid of contribution to structural endothelium [51].

EMPs do not significantly contribute to hematopoietic clusters in intra- and extra-embryonic arteries

In order to gain insight on the origin of hematopoietic clusters in different sites of emergence, we performed lineage tracing using the *Csf1r-iCre* transgenic mouse line, which targets EMPs and their progeny [15, 16].

Kit+ clusters in the E9.5 VA of *Csf1r-iCre::R26^{tdTomato}* embryos showed near complete absence of labeling (**Figure 4A,D**), despite the expected highly efficient labeling of E9.5 EMPs (**Figure 4B**). Non-EMPs were not labeled in the E9.5 CP, whereas ~50% recombined in the E9.5 YS, possibly representing immature EMPs and/or Csf1r+ progenitors other than EMPs (**Figure 4B**). Interestingly, LMPs were previously shown to emerge in the YS vascular plexus starting from E9.5 and to express Csf1r [20, 25]. Hematopoietic clusters in the E10.5 DA, VA, UA and large arterial clusters in the E10.5 YS showed very low labeling, confirming that they do not originate from EMPs (**Figure 4C,D,**

Figure S5A). In contrast, extensive labeling was detected in the E10.5 FL, in agreement with its early colonization by EMPs (**Figure 4C,D**). Accordingly, at E10.5 EMP labeling remained high (>90%), while less than 25% non-EMPs were labeled in both YS+VU and embryo proper (**Figure 4E**).

These data show that hematopoietic clusters emerging in the major intra- and extra-embryonic arteries, including the E10.5 YS, largely contain cells that develop independently of EMPs.

A late wave of Csf1r+ progenitors exert a limited contribution to fetal lympho-myelopoiesis

Since 4-OHT activations at E7.5-E8.5 in the *Cdh5-CreER^{T2}* system both target EMPs and non-EMPs, we wanted to get a better assessment of the contribution of EMPs to fetal lympho-myelopoiesis. To this aim, we took advantage of the Tamoxifen-inducible *Csf1r^{MerCreMer}* line, previously used to trace EMPs in the absence of HSC labeling[15-17, 52]. As expected, in *Csf1r^{MerCreMer}::R26^{tdTomato}* E9.5-E10.5 embryos, both 4-OHT at E8.5 and E9.5 label EMPs but few non-EMPs (**Figure S5B-D**), as also confirmed by microglia labeling in the E16.5 brain **(Figure 4F).**

In the E16.5 FL and thymus, E8.5 activation yielded no detectable labeling, except FL macrophages (23%) (**Figure 4F**). By contrast, E9.5 activation labeled a subset of myeloid and B cells (<20%) and 30-40% T cells in the E16.5 FL and thymus, with the highest labeling detected in more differentiated thymocytes (CD4+ CD8+ DP, 50% labeling) (**Figure 4F**). LSK, including HSCs and multipotent hematopoietic progenitors (MPPs; CD48+ CD150-) were not labeled with either activation (**Figure 4F**).

These data show that Csf1r+ progenitors contribute *in vivo* to fetal lymphoid cells, confirming a previous report [52]. Lympho-myeloid contribution of Csf1r+ progenitors is limited in the FL but, surprisingly, more pronounced in the thymus. Notably, in Csf1r+ progenitors, lymphoid potential appears only at E9.5. Our data provide support for the existence of at least two waves of Csf1r+ progenitors emerging outside of the main arteries, with only the second being endowed with lympho-myeloid potential.

B- and T-lymphoid potential appears in intra- and extra-embryonic HE between E8.5 and E9.5

To determine whether the *ex vivo* potential of fetal-restricted HSPCs mirrored their *in vivo* fate, we isolated traced and untraced cells from *Cdh5-CreER^{T2}::R26$^{tdTomato/zsGreen}$* E9.5 YS or CP and E10.5 YS+VU or AGM (**Figure 5A,B**). We performed colony-forming unit-culture (CFU-C) assays able to

detect single or combined erythroid, myeloid and Mk potential of mature progenitors. CFU-Cs almost completely segregated with labeled cells at both activations in either YS or CP/AGM (**Figure 5C**), in agreement with the absence of differential labeling of EMPs at these two time points (**Figure S1B-E**).

We next tested the frequency of B and T cell progenitors within labeled and unlabeled hemato-endothelial cells (CD31+ and/or Kit+ ; 4-OHT at E8.5) **(Figure S6A,B)**. To this end, we performed a limiting dilution assay (LDA) in OP9/OP9-DL1 co-cultures. While B and T cell progenitors were almost exclusively detected in the labeled fractions (**Figure 5D,E; Figure S6C,D**), T progenitors were significantly more frequent than B progenitors in both AGM (1:81 versus 1:204, p=0.016) and YS+VU (1:123 versus 1:314, p=0.012) (**Figure 5D,E**).

These results show that while EMP emergence takes place from E7.5 onwards, progenitors displaying B- and T-lymphoid potential first emerge from intra- and extra-embryonic HE between E8.5 and E9.5. The different frequencies of B and T lymphocyte progenitors observed in AGM and YS+VU are consistent with a model in which developmental uncoupling may already take place in the tissues of emergence of these precursors.

Single cell transcriptomics coupled with HE lineage tracing identify distinct subsets of pre-HSPCs in AGM and YS+VU

To determine the transcriptional identity of fetal-restricted HSPCs and evaluate the relationship between their origin and molecular signature, we performed single-cell RNA sequencing (scRNA-Seq). We pooled hemato-endothelial CD31+ and/or Kit+ cells from *Cdh5-CreER^{T2}::R26^{tdTomato}* AGM or YS+VU (4-OHT at E8.5) and separated them into labeled (tdT+) and unlabeled (tdT-) cells (**Figure S6A,B**). A total of 44 clusters were identified (**Figure S7A**). Pre-HSPCs distributed in two distinct clusters and were recognized by their transcriptional signature, characterized by the combined expression of the hematopoietic genes *Myb, Kit, Runx1, Adgrg1, Flt3* together with genes reported to identify HSCs or their immediate precursors, including *Hlf* [28, 50, 53], *Cd27* [54, 55], *Mecom* [28], *Cd93* [56], *Hoxa7*, *Hoxa9* [53] among others (**Figure 6A,B**). Pre-HSPC 1 were largely contributed by labeled AGM and unlabeled YS+VU cells; in contrast, tdT+ YS+VU cells made up >75% of pre-HSPC 2 (**Figure 6C,D**). Both clusters expressed *Cd27* (**Figure 6B, Figure S7B**). Differential gene expression (DGE) analysis between the two clusters highlighted genes involved in hematopoietic differentiation as well as metabolic and inflammatory genes (**Figure S7C,D**).

Mapping to existing datasets [29, 55] confirmed a stem/progenitor cell identity of the pre-HSPC 1 cluster, while a lymphoid progenitor signature was more enriched in pre-HSPC 2 (**Figure 6E,F**). Separate evaluation of the signature expression levels in each subset revealed that uncommitted progenitor signatures were particularly expressed in AGM cells of both clusters, while the lymphoid progenitor signature was expressed at the highest level in tdT+ cells of the YS+VU, found almost exclusively in the pre-HSPC 2 cluster (**Figure S7E,F**).

To further dissect pre-HSPC heterogeneity at the transcriptional level, we restricted our analysis to pre-HSPC clusters (**Figure 6G**). Relative gene set comparison evidenced that the pre-HSC transcriptional signature was expressed at the highest level by labeled cells of the AGM (**Figure 6H**), that our whole-mount imaging preferentially localizes within the intra-embryonic portion of the VU (**Figure 3**). In contrast, expression of genes characteristic of hematopoietic differentiation was higher in tdT+ cells of the YS+VU (**Figure 6I**). AGM and YS+VU pre-HSPC exhibited a distinct metabolic signature, with glycolysis prominent in AGM and OXPHOS genes expressed at a higher level in YS+VU (**Figure S8A**). Pro-inflammatory genes, previously suggested to play a role in AGM HSC development [57-60], showed heterogeneous expression. Interestingly, interferon response genes were highest in labeled cells of AGM, while *Tnf, Nfkb, Tlr4* and innate immune response genes were more expressed in labeled cells of the YS+VU (**Figure S8B**). Heterogeneity between tdT+ AGM and YS+VU pre-HSPC was confirmed by DGE analysis, that highlighted higher expression of ribosomal genes in the latter, indicative of higher metabolic activity (**Figure S8C,D**)[61]. Genes involved in signaling pathways implied in HSC specification such as Notch, Shh and TGF-β [62] were highly expressed in tdT- AGM pre-HSPCs, probably the most immature subset (**Figure S8E**). Taken together, our data shows evidence of molecular heterogeneity within pre-HSPCs at the time of HSC emergence and confirms at the transcriptomic level that 4-OHT at E8.5 in *Cdh5-CreER^{T2}* mice labels the majority of E10.5 pre-HSPCs, likely representing the precursors of fetal-restricted HSPCs. Our data also suggest that pre-HSPCs in AGM and YS+VU include cells at distinct stages of maturation, with those located in the AGM representing the most uncommitted subset, while the ones in the YS+VU being primed for differentiation.

Differentially labeled CD27+ Kit+ hematopoietic clusters selectively localize to vitelline and umbilical arteries

Cd27 was one of the highly enriched genes in pre-HSPCs (**Figure 6B, Figure S7B**). It was recently shown to be expressed in HSCs, type II pre-HSCs and lymphoid progenitors in E10.5-E11.5 AGM [54, 55]. To assess whether CD27 would allow a more specific localization of labeled pre-HSPCs, we

performed whole-mount confocal imaging of E10.5 *Cdh5-CreER^{T2}* embryos (**Figure 7A**). At this stage, we noticed that a higher percentage of cells within VU Kit+ cluster cells expressed CD27 as compared to DA or YS arteries (**Figure 7B**), while clusters in the E9.5 YS did not express CD27 (**Figure S9A**), in agreement with the reported lack of CD27 expression in EMPs [54]. Evaluation of labeling of E10.5 Kit+ CD27+ cluster cells showed that E7.5 activation labeled a minority of those clusters in both AGM and YS (**Figure 7A,C,D; Figure S9B**). In contrast, 4-OHT at E8.5 yielded significant labeling in the AGM (**Figure 7A,C**) and in YS arteries (**Figure 7D; Figure S9B**). Strikingly, within the AGM, labeled Kit+ CD27+ clusters were detected in the VU but not in the DA (**Figure 7A,C**). These results strongly suggest that CD27+ pre-HSPCs with lymphoid potential other than EMPs, identified with 4-OHT at E8.5 in *Cdh5-CreER^{T2}* embryos, specifically localize to VU in both intra- and extra-embryonic regions.

Fetal-restricted HSPCs and adult-type HSCs exhibit distinct dynamics of engraftment.

To establish whether fetal-restricted HSPCs could yield multi-lineage engraftment, we directly transplanted E11.5 AGM and YS+VU from *Cdh5-CreER^{T2}::R26$^{tdTomato/zsGreen}$* pulsed with 4-OHT at E8.5 (fetal-restricted HSPCs) or E10.5 (adult-type HSCs) into lethally irradiated syngeneic CD45.1 recipients (**Figure 8A,B**). We did not detect any engraftment (defined as >5% donor chimerism) from YS+VU. Remarkably, all recipient mice repopulated with cells traced at E8.5, but none of those with cells traced at E10.5, displayed high levels of donor-derived labeling in myeloid and lymphoid cells of the PB (**Figure 8C**; **Figure S10A**). Analysis of BM at 16 weeks post-transplantation confirmed PB data, with differentiated cells labeled only with E8.5 activation (**Figure 8D**). Within the BM progenitor cell compartment, LSK, MPP and LK were only labeled in mice transplanted with AGMs traced at E8.5 (**Figure 8D**; **Figure S10B**). Notably, immunophenotypic HSCs were not found labeled with either activation time point (**Figure 8D**; **Figure S10B**). Within transplanted E11.5 AGM pools, a similar percentage of type I and type II pre-HSCs were labeled with both activations, while CD31+ Kit+ CD41^{low} CD45+/- CD43+ CD201- progenitors were highly labeled with the E8.5 activation at and not labeled with activation at E10.5 (**Figure S10C**), as previously shown (**Figure 1C**). These data suggest that, in the E11.5 AGM, adult repopulation potential may be confined to fetal-restricted HSPCs and is likely to reside in the CD201- subfraction, while adult-type HSCs labeled at E10.5 are not yet competent for engraftment and most probably require further maturation.

Next, we performed competitive transplants in which unfractionated cells from E14.5 *Cdh5-CreER^{T2}::R26^{tdTomato} / R26^{EYFP}* FL pulsed with 4-OHT at E8.5 or E10.5 were transplanted into lethally irradiated recipient mice (**Figure 8A**). Chimerism levels and labeling frequencies within donor-

derived fractions were followed over time in PB. To account for variability in labeling efficiency, donor-derived labeling was normalized to the percentage of labeled LSK in each donor sample prior to transplantation (**Figure S10D**). Both E14.5 FL cells labeled at E8.5 and E10.5 exhibited multi-lineage engraftment potential in primary recipients (**Figure 8E, Figure S10E**). However, analysis of the progenitor compartment in BM of primary recipients at 16 weeks showed that only labeled cells pulsed at E10.5 could expand in the host BM niche (**Figure 8F, Figure S10F**). We then performed secondary transplantations (**Figure 8A**). In secondary recipients, the PB contribution of fetal-restricted HSPCs decreased, whereas that of adult-type HSCs significantly increased with time (**Figure 8E**). Upon terminal analysis, no labeled phenotypic HSCs from donor cells pulsed at E8.5 were detected in the BM of secondary recipients, while labeled progenitors originally pulsed at E10.5, including phenotypic HSCs, were still detected (**Figure 8F**).

Taken together, these results suggest that, during development, adult multilineage hematopoietic engraftment potential first appears in fetal-restricted HSPCs; however, these cells are biased for differentiation and devoid of long-term self-renewal potential. Our data are consistent with adult-repopulating HSCs being still largely immature in the E11.5 AGM; therefore, as recently proposed [33, 63], they may require the FL microenvironment in order to complete their maturation.

Discussion

The first intra-embryonic hematopoietic clusters are generated in the portion of the VA most proximal to the DA at E9.5 [9, 45, 46]. Although vitelline and umbilical arteries (VU) are known to harbor HSC precursors, this knowledge relied on primary transplantation experiments [7, 47] in which post-transplant contribution to HSC was not assessed. Due to the lack of specific ways to identify and label VU-derived hematopoietic progenitors, their physiological role during development was unclear. Here we demonstrate for the first time in an unperturbed *in vivo* context that this vascular site is a source of multipotent hematopoietic progenitors that substantially contribute to fetal lympho-myelopoiesis. Using temporally resolved hemogenic endothelium (HE) lineage tracing, whole-mount imaging, and functional transplantation assays, we reveal that the HE of the VU originates HSPCs acting as major contributors to fetal lympho-myelopoiesis (**Figure 8G**). Thus, the HE wave we identified contains the precursors of the recently described pre-constituted FL hematopoietic hierarchy [28, 42], including fetal HSCs [32, 33] and eMPPs [29, 30], whereas adult-type HSCs are mainly generated by the HE of the DA. Given that the EHT output is heterogeneous, this raises the question of whether this diversification takes place post-EHT, or it is already imprinted at the HE stage. Recent evidence supports the latter [64, 65]. The hematopoietic contribution of fetal-restricted HSPCs rapidly declines after birth, marking them as a primarily transient population.

Our lineage tracing strategy cannot discriminate between the separate contributions of individual progenitors (i.e. fetal HSCs or eMPPs) but rather identifies a temporally and spatially distinct wave of HE as the source of fetal-restricted HSPCs. Although the existence of fetal-restricted HSCs and eMPPs had been demonstrated previously, the precise time and anatomical location of their emergence has not been investigated before. The first report to identify developmentally restricted HSCs was based on *Flt3* lineage tracing [32]; however, this study primarily relied on FL transplantation assays to probe their contribution, and thus offered little information on their origin and physiological role. Patel et al. also employed *Flt3* lineage tracing, but adopted an inducible strategy coupled with clonal barcoding to identify a population of eMPPs that arise early in embryogenesis and contribute to postnatal multi-lineage output [29]. This and another recent study[30] mainly focused on postnatal contribution and did not investigate in depth the emergence of eMPPs. Another report showed that hematopoietic progenitors in the FL are generated independently of adult-type HSCs and that *Mecom* levels, normally higher in the intra-embryonic region, can play a functional role in HSPC specification [28, 42]. We found *Mecom* to be more expressed in AGM than YS+VU pre-HSPCs (**Figure 6H**), in line with an identity of the latter distinct from adult-type HSCs. The high expression of *Mecom* in labeled AGM pre-HSPCs (**Figure 6H**) confirms that *Mecom*-expressing cells are predominantly located in the intra-embryonic portion of the main arteries, including VU [28]. Thus, *Mecom* expression

levels may not only be important to specify adult HSCs, consistent with *EVI1*^{creERT2} labeling phenotypic HSCs in the E14.5 FL [28], which our data suggest to be in part fetal-restricted (**Figure 1F**).

Our pulse-chase approach labels *Cdh5*-expressing endothelial and HE cells within a 12-hours *in vivo* window [36, 37]. Surprisingly, already 24 hours after labeling this resulted in near-exclusive recombination of hematopoietic clusters, while much lower tracing frequencies were detected in the HE or the adjacent vascular endothelium. HE is a transient hematopoietic-committed cell population, and it is not bipotent [51]. Additionally, HE was shown to represent a lineage distinct from arterial endothelium [66]. Thus, our data suggest that labeled HE undergoes EHT in a rapid time window of a few hours *in vivo*, consistent to what was observed *ex vivo* in time lapse imaging studies [6] and does not contribute to structural endothelium. The specificity of the labeling of distinct hematopoietic waves arises because of the sequential emergence of each HE, captured with our alternative activation modes. Neither 4-OHT activation at E7.5 or E8.5 label (or label very few) endothelial cells in the ventral side of the DA, providing a strong rationale for the lack of labeling of adult-type HSC (see **Figure 3E** and **Figure 7A**). In addition, 4-OHT at E7.5 labels a minority of cells in the VU, supporting that this activation mode does not mark fetal-restricted HSPCs arising in this location. Moreover, the fact that we detected very low labeling of non-hemogenic endothelium in hematopoietic sites even with the activation models that label hematopoietic cells in the same locations additionally raises the intriguing possibility that a significant portion of the early embryonic endothelium of these regions may in fact be hemogenic, as recently proposed [67].

Csf1r-based lineage tracing clarified the relationship between EMPs and fetal-restricted HSPCs. Both intra- and extra-embryonic hematopoietic clusters in the VU arise independently of *Csf1r*+ progenitors, confirming that VU clusters do not contain EMPs or their progeny and suggesting that most of the *Csf1r*-expressing progenitors likely arise in the YS vascular plexus. We made the interesting observation that late-emerging *Csf1r*+ progenitors contributed modestly (<20%) to fetal myeloid and B lymphoid compartments and more significantly to TSP in the fetal thymus. These data support the existence of at least two *Csf1r*+ progenitor waves: an early EMP wave restricted to erythro-myeloid output and a later wave with broader lympho-myeloid potential, likely overlapping with LMPs [20, 25]. The limited contribution to B and myeloid cells in the E16.5 FL implies that, at this stage, the majority of these cells originate from *Csf1r*-independent progenitors, i.e. fetal-restricted HSPCs. Nevertheless, the higher labeling frequencies observed in the thymus emphasizes the functional uncoupling between T- and B-myeloid-producing progenitors during development, further supported by our LDA assay identifying different frequencies of B and T cell progenitors in both AGM and YS+VU [41, 68, 69].

Single-cell RNA-seq revealed two pre-HSPC subsets in the E10.5 embryo, prominently originated by HE specified between E8.5 and E9.5, with shared progenitor signatures (*Kit*, *Runx1*, *Hlf*, *Cd27*, *Mecom*, *Hoxa7*, *Hoxa9*) but overall distinct molecular programs. AGM-derived pre-HSPCs displayed an uncommitted profile with a strong glycolytic signature, whereas YS+VU-derived cells exhibited increased oxidative metabolism, ribosomal activity, and differentiation-priming transcripts. These transcriptional differences suggest that fetal-restricted HSPCs represent a heterogeneous mixture of progenitors, with intra-embryonic cells poised to mature into fetal HSCs and extra-embryonic cells already primed for differentiation. Functional heterogeneity between intra- and extra- embryonic HE has been shown[26] and recent work characterized their transcriptomic heterogeneity [31, 65]. Interestingly, intra-embryonic HE shows elevated chromatin and RNA splicing gene expression and greater isoform complexity compared to YS HE [65].

Our transplantation experiments demonstrate that fetal-restricted HSPCs possess multilineage potential yet lack durable self-renewal. Given the labeling patterns observed in source tissues and in repopulated recipients, the progenitor subsets which could yield engraftment in our E11.5 AGM transplants were CD43+ CD201- progenitors (likely the most differentiated fraction) and/or phenotypic CD201+ pre-HSCs and HSCs labelled at E8.5, but not those labelled at E10.5. Further experiments will be needed to validate this finding. Competitive FL transplants further confirmed that HSPCs labeled at E8.5 can yield multilineage engraftment in primary recipients but fail to persist upon secondary transplantation, in contrast to the durable reconstitution by E10.5-labeled adult-type HSCs. These results underscore the intrinsic differentiation bias of fetal-restricted HSPCs and their functional distinction from definitive HSCs that may require the FL microenvironment for full maturation, as recently suggested [33, 63]. Although we cannot formally exclude that part of the observed dynamics reflects differences in the timing of pre-HSC emergence and maturation, the combined evidence from lineage output, transplantation, postnatal contribution, and spatial origin argues that our labeling strategy captures functionally heterogeneous HSPCs rather than a single homogeneous HSC pool.

Our data reinforce and enrich the concept that embryonic hematopoiesis is layered into multiple spatially and temporally distinct programs. While HSPC emergence is initially segregated in space, multi-lineage progenitors eventually converge in the FL [1]. For a long time, the FL was thought to be an intermediate reservoir for HSC expansion after their initial generation in the AGM, in preparation to their lifelong residency in the BM. Recent work revolutionized this concept and showed that FL is a complex hub in which distinct waves of progenitors differentiate, expand and mature with mechanisms still largely unknown [28, 29, 33, 56]. Indeed, ours and others' data imply that fetal-restricted HSPC and adult-type HSCs coexist in the FL, where they undergo different processes leading

515 respectively to differentiation and self-renewal. Testing to what extent HSPC fate is intrinsically pre-
516 determined, and what exactly is the role of the microenvironment will offer crucial insight. Based on
517 our findings, it is tempting to speculate that before FL colonization, the emergence and transiting of
518 fetal-restricted HSPC in the VU niche may influence their identity and fate. Metabolic [45, 70] and
519 inflammatory factors [57-60] were shown to promote HSC development. Indeed, here we observed that
520 these genes exhibit variable levels of expression in subsets of pre-HSCs of different origins (**Figures**
521 **S7 and S8**). Interferon signaling, more active in AGM-derived pre-HSPCs (**Figure S8B**), is required
522 for the switch between fetal and adult HSCs, beginning prior to birth [71]. Therefore, it is possible that
523 inflammatory cues are one of the main extrinsic factors required to specify adult-type HSC identity.
524 Thus, it will be interesting to investigate what are the specific signals that regulate fetal-restricted
525 HSPCs.

526 Collectively, our findings provide mechanistic and spatial resolution to the long-standing question of
527 how fetal-restricted and adult-type hematopoietic programs are organized in the embryo. The
528 discovery that a distinct subset of hemogenic endothelium in the vitelline and umbilical arteries gives
529 rise to a transient wave of multipotent progenitors that dominate fetal lympho-myelopoiesis highlights
530 a previously overlooked site and mechanism of hematopoietic diversification. These insights refine
531 our understanding of layered hematopoiesis and can inform studies of pediatric malignancies with
532 prenatal origins, where leukemic transformation may occur within these developmentally restricted
533 progenitor pools.

534

Acknowledgements

We kindly acknowledge the contributions of Riccardo Gamberale, Anna Giovenzana, Giulia Di Meo and the personnel of the animal facilities at San Raffaele Scientific Institute and the University of Milan for support and technical help. We thank Matteo Iannacone for providing *R26*^{zsGreen} mice. EA was supported by Fondazione Cariplo (Grant n. 2018-0102), Leukemia Research Foundation (Award ID: 831382), Cariplo Telethon Alliance (Grant n. GJC22013) and Worldwide Cancer Research (Grant ref. 24-0083). EA and AF were supported by Italian Ministry of University and Research (Grants n. P20223HEZC and 2022TT8Z3M). CBa was supported by Fondazione Umberto Veronesi. EA and CBa acknowledge support from the COST (European Cooperation in Science and Technology) Action IMMUNO-model, CA21135. GA and MDB were supported by program MC_UU_00029/5 in the MRC Molecular Haematology Unit Core award, as were experiments performed by YP and NM. The flow cytometry facility at the MRC WIMM is supported by the MRC TIDU; MRC MHU (MC_UU_12009); NIHR Oxford BRC; Kay Kendall Leukaemia Fund (KKL1057), John Fell Fund (131/030 and 101/517), the EPA fund (CF182 and CF170) and by the MRC WIMM Strategic Alliance awards G0902418 and MC_UU_12025. Work in the ADi lab is supported by the Italian Telethon Foundation (SR-Tiget grant award TGT21007). AF was supported by the Fondazione Cariplo (2018-0298) and the Associazione Italiana per la Ricerca sul Cancro (AIRC) (22905). SBr was supported by the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860034. FTF was supported by the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860034 and Cariplo Telethon Alliance grant n. GJC22013. RP was supported by Associazione Italiana Ricerca sul Cancro (AIRC) IG-29341, Italian MUR Dipartimenti di Eccellenza 2023-2027 (l. 232/2016, art. 1, commi 314 - 337) and European Union - NextGenerationEU through the Italian Ministry of University and Research under PNRR - M4C2-I1.3 Project PE_00000019 "HEAL ITALIA". Schematics were created with BioRender.com.

Author Contributions

CBa conceived, designed and performed experiments, analyzed and interpreted data, made figures and wrote the manuscript; GQ, RO, FTF, AM performed experiments, analyzed and interpreted data, edited the manuscript; GA analyzed and interpreted data and performed bioinformatic analysis; YP, NM, EM, ACa, AP, MN performed experiments and analyzed data; TM, BN, FSdS, GZ, VB, ADo, EDE, GS, ACo performed experiments; MM, DdA, SS made scRNA-Seq libraries; SBo, SDM performed cell sorting; CDO, MNA provided technical support; CBi, RAP, RM, MFTRDB provided

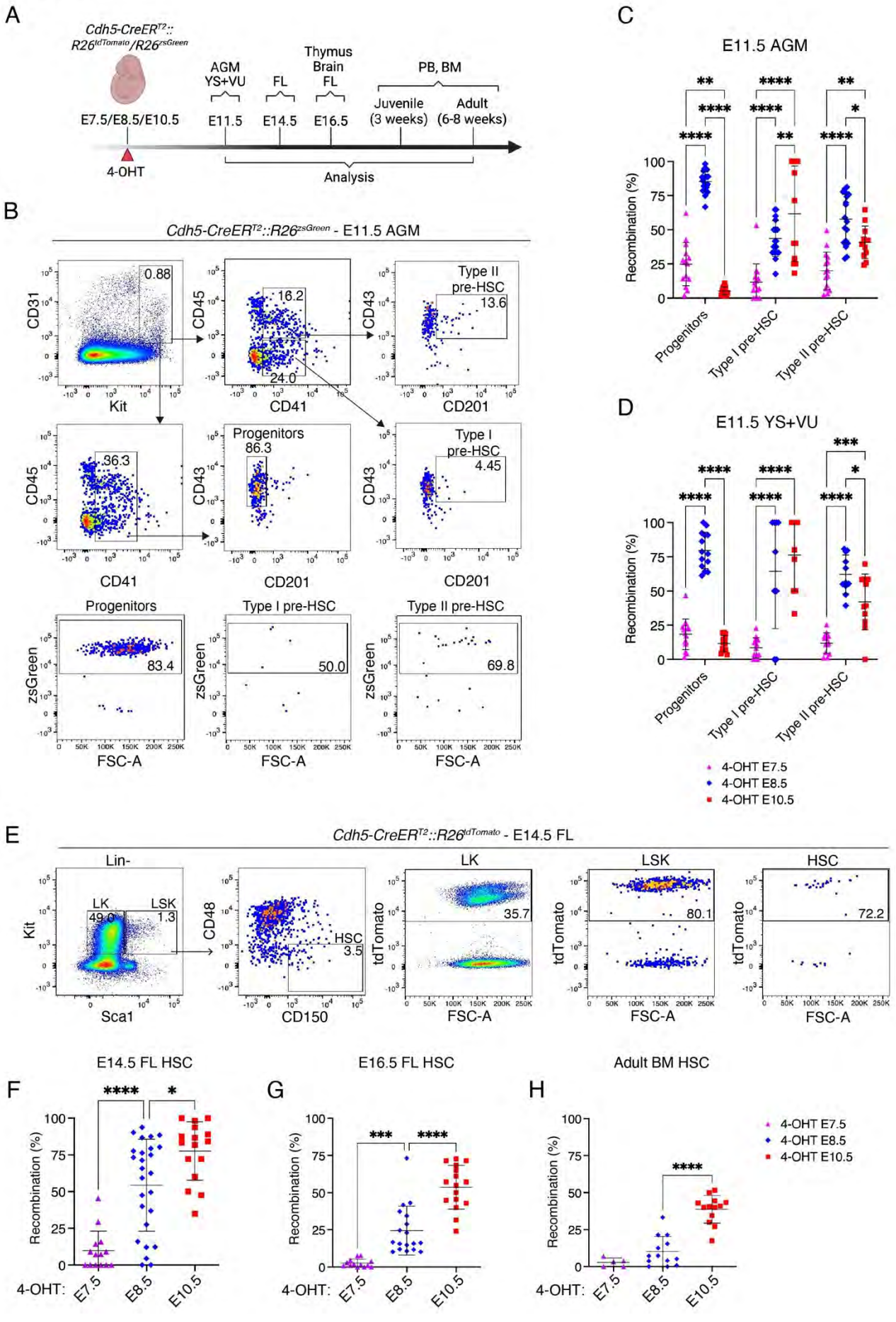
reagents and materials; ACu, SBr provided reagents and materials and interpreted data; ADi, AF performed experiments, interpreted data and provided reagents and materials; RP provided reagents and materials, analyzed and interpreted data and performed bioinformatic analysis; EA conceptualized, supervised and led the study, performed experiments, analyzed and interpreted data, made figures and wrote the manuscript. All authors read the manuscript and approved its final version.

Declaration of interests

The authors declare no competing interests.

MAIN FIGURE TITLES AND LEGENDS

Figure 1


A
Cdh5-CreER^T2::
R26^tdTomato/R26^zsGreen
E7.5/E8.5/E10.5
4-OHT
AGM
YS+VU
E11.5
FL
E14.5
Thymus
Brain
FL
E16.5
PB, BM
Juvenile
(3 weeks)
Adult
(6-8 weeks)
Analysis
B
Cdh5-CreER^T2::R26^zsGreen - E11.5 AGM
CD31
Kit
0.88
CD45
CD41
16.2
24.0
CD43
CD201
Type II
pre-HSC
13.6
CD45
36.3
CD41
CD43
CD201
Progenitors
86.3
CD43
CD201
Type I
pre-HSC
4.45
Progenitors
zsGreen
83.4
FSC-A
Type I pre-HSC
zsGreen
50.0
FSC-A
Type II pre-HSC
zsGreen
69.8
FSC-A
C
E11.5 AGM
Recombination (%)
Progenitors
Type I pre-HSC
Type II pre-HSC
D
E11.5 YS+VU
Recombination (%)
Progenitors
Type I pre-HSC
Type II pre-HSC
4-OHT E7.5
4-OHT E8.5
4-OHT E10.5
E
Cdh5-CreER^T2::R26^tdTomato - E14.5 FL
Lin-
Kit
LK
49.0
LSK
1.3
Sca1
CD48
HSC
3.5
CD150
LK
tdTomato
35.7
FSC-A
LSK
tdTomato
80.1
FSC-A
HSC
tdTomato
72.2
FSC-A
E14.5 FL HSC
F
Recombination (%)
4-OHT:
E7.5
E8.5
E10.5
E16.5 FL HSC
G
Recombination (%)
4-OHT:
E7.5
E8.5
E10.5
Adult BM HSC
H
Recombination (%)
4-OHT:
E7.5
E8.5
E10.5
4-OHT E7.5
4-OHT E8.5
4-OHT E10.5

577

Figure 1. Lineage tracing of HE between E8.5 and E9.5 labels a subset of phenotypic hematopoietic progenitors and HSCs.

(A) Visual schematic of lineage tracing experiments in *Cdh5-CreER^{T2}::R26^{tdTomato}/ R26^{zsGreen}* E11.5-E14.5-E16.5 embryos and postnatal mice, related to **Figure 1, 2, S1F-G, S2C-M and S3**.

(B) Representative flow cytometric analysis of aorta-gonad-mesonephros (AGM) region hematopoietic progenitors (Ter119- Kit+ CD31+ CD41+ CD43+ CD201-), pre-HSCs type I (Ter119- Kit+ CD31+ CD45- CD41+ CD43+ CD201+) and pre-HSCs type II (Ter119- Kit+ CD31+ CD45+ CD41+ CD43+ CD201-) in E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos. 4-OHT at E7.5 (n = 15), 4-OHT at E8.5 (n = 14), 4-OHT at E10.5 (n = 12) Yolk Sac + Vitelline and Umbilical artery (YS+VU) and 4-OHT at E7.5 (n = 15), 4-OHT at E8.5 (n = 23) , shown here, 4-OHT at E10.5 (n = 12) AGM were analyzed individually across 6 independent experiments.

(C-D) Quantification of flow cytometric analysis displayed in **(B)**. The percentage (%) of recombination is represented as the % of zsGreen+ cells within progenitors, type I and type II pre-HSCs in E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* AGM (C) and YS+VU (D). Error bars represent mean ± standard deviation (SD). *p<0.05, **p<0.01, ***:p<0.001, ****p<0.0001 (two-way ANOVA followed by Tukey's multiple comparisons test).

(E) Representative flow cytometric analysis of fetal liver (FL) HSCs (Lin- Kit+ Sca1+ CD48- CD150+), LK (Lin- Kit+ Sca1-) and LSK (Lin- Kit+ Sca1+) and relative labeling frequencies in E14.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* embryos. Lineage cocktail: B220, CD19, CD3e, F4/80, Gr1, Nk1.1, Ter119, 7-AAD. Embryos were activated with 4-OHT at E7.5 (n = 14), E8.5 (n = 26) or E10.5 (n = 16), shown here, and analyzed individually across six independent experiments.

(**F-H**) Recombination within phenotypic HSCs in E14.5 FL (**F**), E16.5 FL (**G**) and adult bone marrow (BM) (**H**). Data are shown as the percentage of tdTomato+ or zsGreen+ cells within HSCs. E16.5 FL: 4-OHT at E7.5 (n = 13), E8.5 (n = 18), E10.5 (n = 17), four independent experiments. Adult BM: 4-OHT at E7.5 (n = 5), E8.5 (n = 13), E10.5 (n = 13), seven independent experiments. Error bars represent mean ± standard deviation (SD). *:p<0.05, ***:p<0.001, ****:p<0.0001 (one-way ANOVA followed by Tukey's multiple comparisons test).

Figure 2

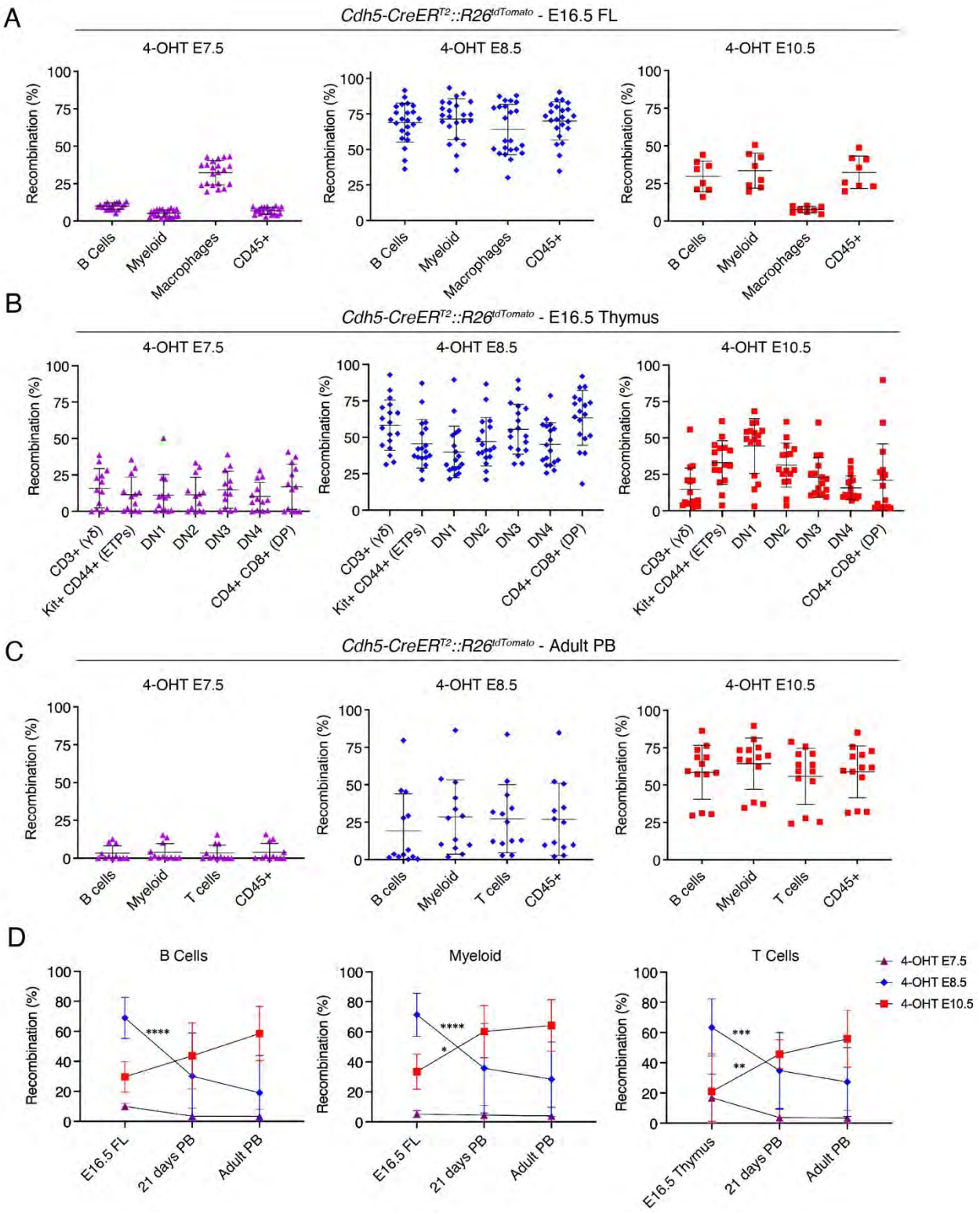


Figure 2. Extensive but transient lympho-myeloid contribution of fetal-restricted HSPCs at the end of gestation.

(A) Quantification of labeled (tdTomato+) B cells (CD45+ B220+), myeloid cells (CD45+ CD11b+), macrophages (CD45+ F4/80^{hi}) and total leukocytes (CD45+) in E16.5 FL from *Cdh5-CreER^{T2}::R26^{tdTomato}*

embryos activated at E7.5 (left, n = 21), E8.5 (middle, n = 23) or E10.5 (right, n=8), analyzed across 5 independent experiments (gating strategy in **S3A**). Error bars represent mean ± SD.

(B) Quantification of labeled thymocyte subsets in *Cdh5-CreER^{T2}::R26^{tdTomato}* E16.5 fetal thymus (gating strategy in **S3B**), activated with 4-OHT at E7.5 (left, n = 13), E8.5 (middle, n = 18) or E10.5 (right, n = 16). Thymuses were analyzed individually across 4 independent experiments. Error bars represent mean ± SD.

(C) Quantification of flow cytometric analysis of labeled B cells (CD45+ B220+), myeloid (CD45+ CD11b+), T cells (CD45+ CD3e+) and total leukocytes (CD45+) in adult (2 months old) PB from *Cdh5-CreER^{T2}::R26^{tdTomato}* mice (gating strategy in **S3D**), activated with 4-OHT at E7.5 (left, n = 12), E8.5 (middle, n = 13) or E10.5 (right, n = 13). Mice were analyzed individually across 7 independent experiments. Error bars represent mean ± SD.

(D) Longitudinal labeling of B and T lymphocytes and myeloid cells at the three analysis time points (E16.5, 21 days and 2 months/adult). Asterisks show significant differences between labeling at E16.5 and 21 days for the E8.5 and E10.5 activation time points. The T cells graph was made based on recombination frequencies within the CD4+ CD8+ subset. Error bars are colored for each activation time point and represent mean ± SD. *p<0.05, **p<0.01, ***:p<0.001, ****p<0.0001 (two-way ANOVA followed by Tukey's multiple comparisons test).

Figure 3

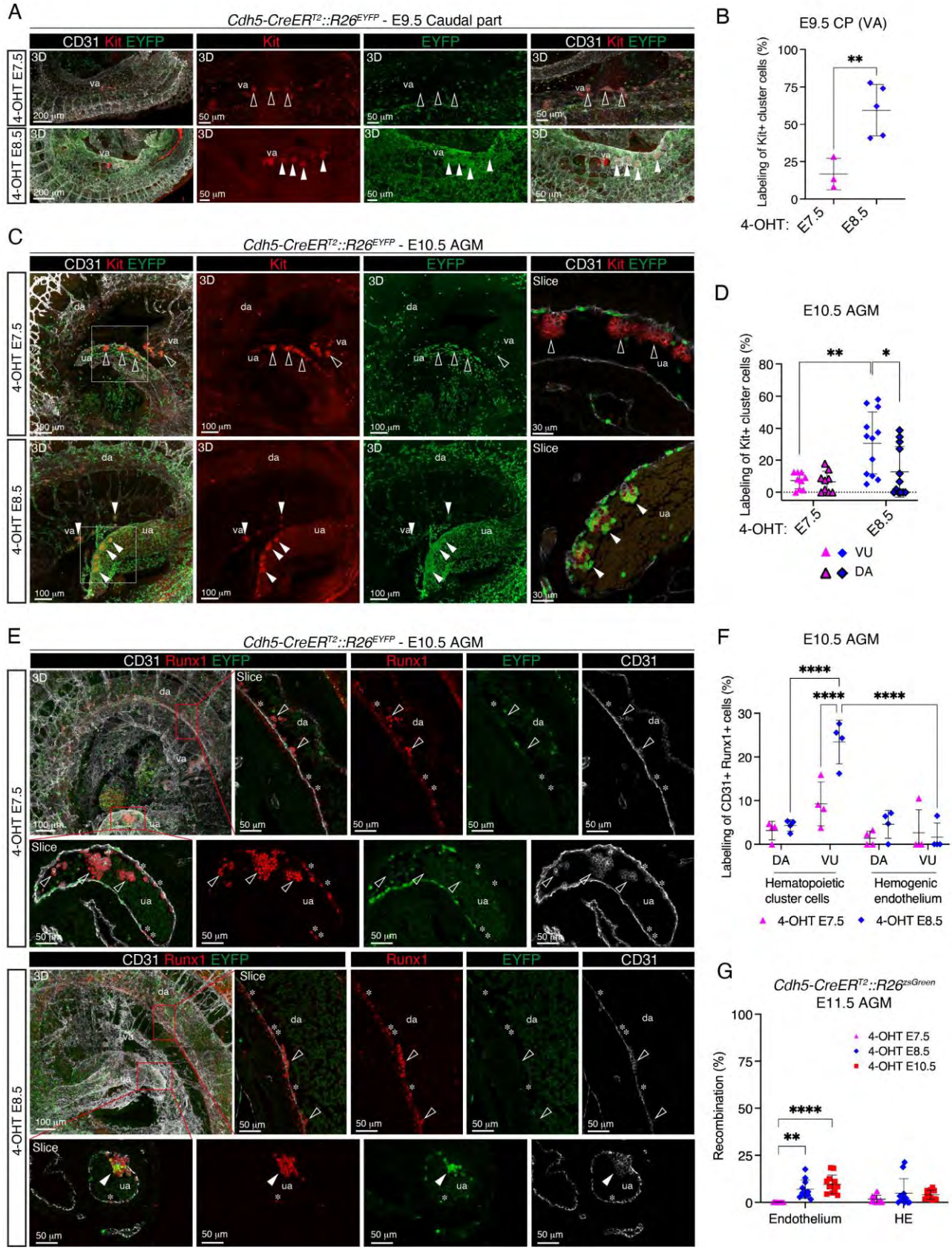

A
Cdh5-CreER^T2::R26^EYFP - E9.5 Caudal part
CD31 Kit EYFP
Kit
EYFP
CD31 Kit EYFP
4-OHT E7.5
4-OHT E8.5
3D
va
B
E9.5 CP (VA)
Labeling of Kit+ cluster cells (%)
4-OHT: E7.5 E8.5
C
Cdh5-CreER^T2::R26^EYFP - E10.5 AGM
CD31 Kit EYFP
Kit
EYFP
Slice
da
ua
va
D
E10.5 AGM
Labeling of Kit+ cluster cells (%)
4-OHT: E7.5 E8.5
VU
DA
E
Cdh5-CreER^T2::R26^EYFP - E10.5 AGM
CD31 Runx1 EYFP
Runx1
EYFP
CD31
F
E10.5 AGM
Labelling of CD31+ Runx1+ cells (%)
DA VU DA VU
Hematopoietic cluster cells
Hemogenic endothelium
4-OHT E7.5
4-OHT E8.5
G
Cdh5-CreER^T2::R26^zsGreen
E11.5 AGM
4-OHT E7.5
4-OHT E8.5
4-OHT E10.5
Recombination (%)
Endothelium
HE

Figure 3. Whole-mount confocal imaging localizes the initial emergence of fetal-restricted HSPCs to the vitelline and umbilical arteries.

(A) Whole-mount immunofluorescence (WM-IF) confocal analysis of E9.5 (21-27sp) *Cdh5-CreER^{T2}::R26^{EYFP}* embryos in the caudal part (CP). Maximum-intensity 3D projections are shown; middle and right panels are magnifications. Arrowheads indicate Kit$^+$ hematopoietic clusters in the vitelline artery (va), labeled following 4-OHT administration at E8.5 (filled arrowheads) but not at E7.5 (open arrowheads). Scale bars: 200 μm (left), 50 μm (magnifications).

(B) Quantification of EYFP labeling within Kit$^+$ cluster cells in the VA of E9.5 embryos shown in (**A**). Measurements were obtained from embryos activated at E7.5 (n = 3) or E8.5 (n = 5), using 1–4 images per embryo (11 images for E7.5; 6 images for E8.5). Data are mean ± SD. **:p<0.01 (two-tailed unpaired Student's *t*-test).

(C) WM-IF confocal analysis of the AGM region of E10.5 (32-36sp) *Cdh5-CreER^{T2}::R26^{EYFP}* embryos. Left and middle panels show maximum intensity 3D projections; the boxed region is shown as a single 2.5 μm optical slice (right). Arrowheads indicate Kit+ hematopoietic clusters in the umbilical artery (ua), unlabeled after E7.5 activation (open arrowheads; top), and labeled after E8.5 activation (filled arrowheads; bottom). Embryos activated at E7.5 (n = 9) or E8.5 (n = 12) were analyzed in seven independent experiments. Scale bars: 100 μm (3D), 30 μm (slice). va: vitelline artery; da: dorsal aorta.

(D) Quantification of labeling in Kit+ hematopoietic clusters within the AGM at E10.5, shown in (**C**). Clusters in the dorsal aorta (DA) and vitelline/umbilical arteries (VU) were quantified separately. Data derive from 9 embryos (E7.5 activation) and 12 embryos (E8.5 activation), with 3–6 images per embryo (39 images for E7.5; 56 images for E8.5). Data are mean ± SD. *:p<0.05, **:p<0.01 (two-way ANOVA with Tukey's multiple comparisons test).

(E) WM-IF confocal analysis of E10.5 AGM region of E10.5 (32-36sp) *Cdh5-CreER^{T2}::R26^{EYFP}* embryos. Boxed areas in maximum-intensity projections are magnified in single 2.5 μm optical slices. Runx1$^+$ CD31$^+$ round-shaped hematopoietic cluster cells (arrowheads) and flat-shaped hemogenic endothelium (asterisks) are indicated. Embryos activated at E7.5 (n = 4) or E8.5 (n = 4) were analyzed. Scale bars: 100 μm (3D), 60 μm (slice).

(F) Quantification of labeling in Runx1+ CD31+ cluster cells and hemogenic endothelium (HE) from (**E**), analyzed separately in DA and VU regions. Data derive from 4 (E7.5 activation) and 4 embryos (E8.5 activation), with 5–12 images per embryo (33 images for E7.5; 40 images for E8.5). Data are mean ± SD. ****:p<0.0001 (two-way ANOVA with Tukey's test).

(G) Flow cytometric quantification of labeled HE (Ter119- CD31+ Kit+ CD45- CD41- CD43-) and non-hemogenic endothelium (Ter119- CD31+ Kit- CD45- CD41- CD43-) in AGM of E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos (gates in **Figure S4C)**. Embryos activated at E7.5 (n = 15), E8.5 (n = 11) or

E10.5 (n = 12) were analyzed across five independent experiments. Data are mean ± SD. **:p<0.01, ****:p<0.0001 (two-way ANOVA with Tukey's test).

Figure 4

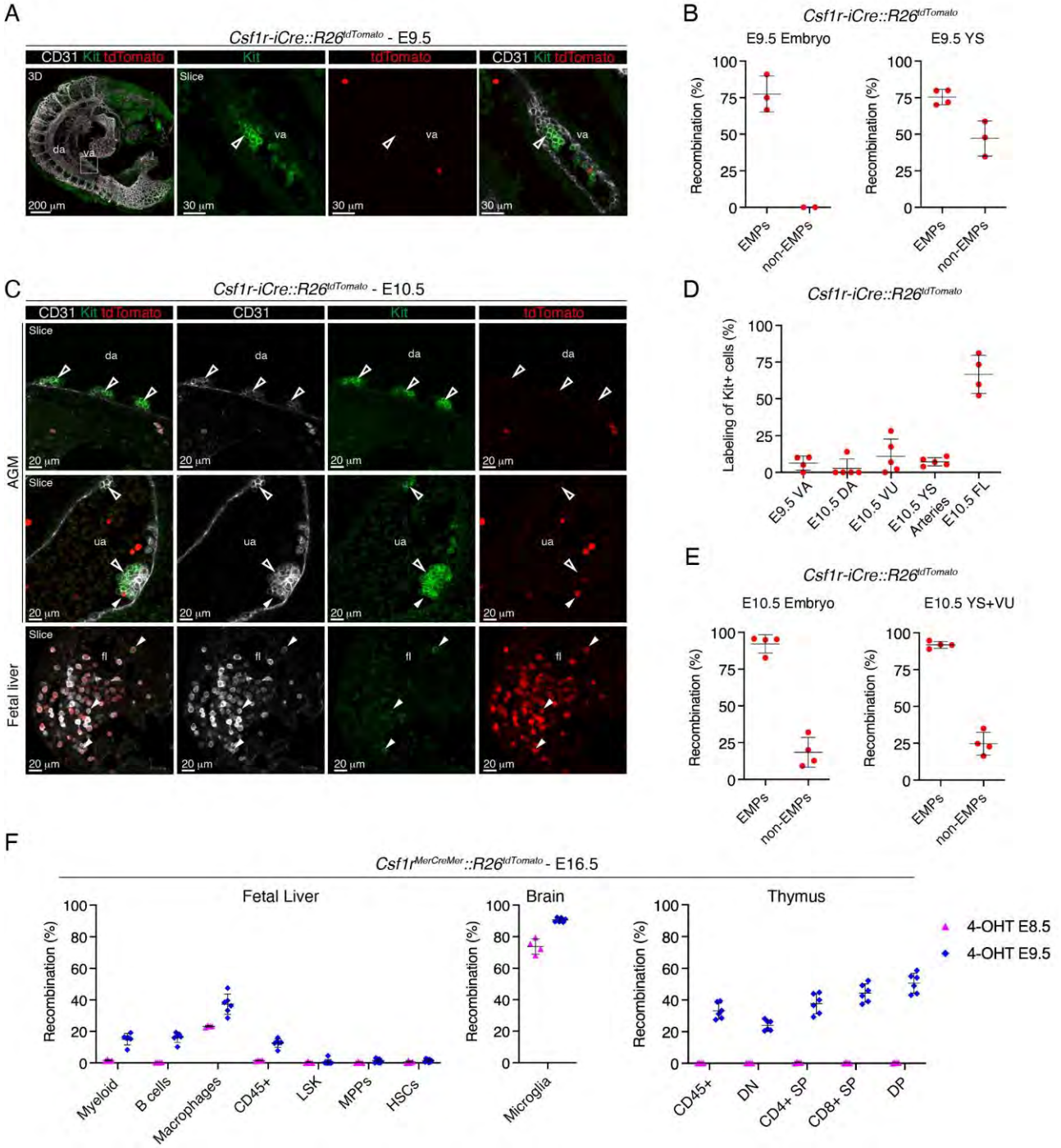


Figure 4. *Csf1r* lineage tracing reveals a non-EMP identity of arterial hematopoietic clusters and the contribution of Csf1r+ progenitors to fetal lympho-myelopoiesis.

(A) WM-IF confocal analysis of E9.5 *Csf1r-iCre::R26*^{tdTomato} embryos. 3D magnification (left) or single 2.5 µm-thick optical slices (middle, right) are shown. Open arrowheads indicate unlabeled (tdTomato-) Kit+ clusters in the va. Four embryos were analyzed in 2 independent experiments. Scale bar: 200 µm (3D), 30 µm (slices).

(B) Flow cytometric quantification of labeled (tdTomato+) EMPs (Ter119- Kit+ CD41^{lo} CD16/32+) and non-EMPs (Ter119- Kit+ CD41^{lo} CD16/32-) in E9.5 *Csf1r-iCre::R26*^{tdTomato} YS and embryos (gates in **Figure S1B**). Four YS and four embryos were analyzed. Data are mean ± SD.

(C) WM-IF confocal analysis of E10.5 *Csf1r-iCre::R26*^{tdTomato} AGM region (top rows) and FL (bottom). Single 2.5 μm optical sections are shown. Arrowheads indicate Kit+ hematopoietic clusters in the da, ua and hematopoietic cells in the FL. Open arrowheads mark unlabeled cells; filled arrowheads indicate tdTomato+ cells. Five embryos were analyzed across three independent experiments.

(D) Quantification of labeling from WM-IF images shown in (**A**) and (**C**). Measurements were obtained from E9.5 (n = 4) and E10.5 (n = 5) embryos, using 2–8 images per embryo (21 E9.5 images; 40 E10.5 images total). Data are mean ± SD.

(E) Flow cytometric quantification of labeled (tdTomato+) EMPs and non-EMPs in E10.5 *Csf1r-iCre::R26*^{tdTomato} YS+VU and embryos. Four YS and four embryos were analyzed. Data are mean ± SD.

(F) Flow cytometric quantification of E16.5 *Csf1r*^{MerCreMer}*::R26*^{tdTomato} FL (left), brain (middle) and thymus (right) following 4-OHT administration at E8.5 (n = 4) or E9.5 (n = 6). Shown is the percentage of tdTomato^{+} cells within FL myeloid cells (Ter119- CD45+ CD11b+ F4/80-), B cells (Ter119- CD45+ B220+), macrophages (Ter119- CD45+ CD11b- F4/80+), CD45+ cells (Ter119- CD45+), LSK (Lin- Kit+ Sca1+) MPPs (Lin- Kit+ Sca1+CD48+ CD150-), HSCs (Lin- Kit+ Sca1+CD48- CD150+); brain microglia (Kit+ CD45+ CD11b+ F4/80+); Thymus CD45+, DN (CD45+ CD4- CD8-) CD4+ SP (CD45+ CD4+ CD8-), CD8+ SP (CD45+ CD4- CD8+), DP (CD45+ CD4+ CD8+). Experimental scheme is shown in **Figure S5B**. Data are mean ± SD from two independent experiments.

Figure 5

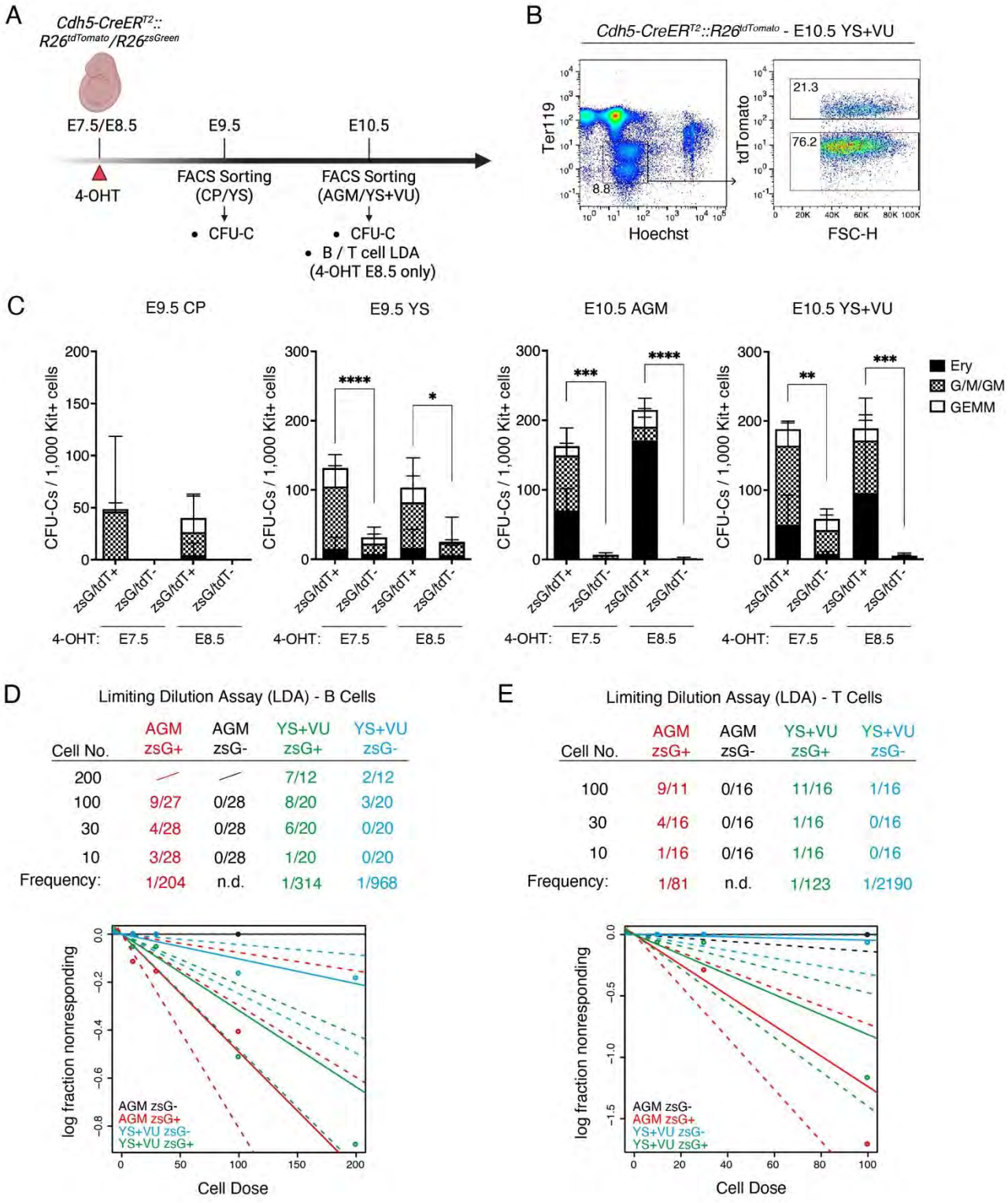


Figure 5. ***Ex vivo*** **B and T lymphoid potential emerges in intra- and extra-embryonic HE at E8.5, while erythro-myeloid potential is already present at E7.5**

(A) Schematic overview of colony-forming unit–culture (CFU-C) and limiting dilution assays used to assess *ex vivo* hematopoietic potential of traced and untraced hemato-endothelial cells.

(B) Representative flow cytometric gating strategy for isolation of Ter119- labeled (tdTomato+ or zsGreen+) and unlabeled (tdTomato- or zsGreen-) cell fractions from E10.5 *Cdh5-CreER^{T2}* YS+VU, following 4-OHT administration at E7.5 or E8.5 (shown).

(C) CFU-C output from labeled and unlabeled Ter119^{-} cell fractions isolated from *Cdh5-CreER^{T2}::R26$^{tdTomato/zsGreen}$* E9.5 caudal part (CP) and yolk sac (YS) (left), or E10.5 AGM and YS+VU (right), following 4-OHT activation at E7.5 or E8.5. Colony numbers are normalized to 1,000 Kit^{+} cells plated. E9.5: 4-OHT E7.5 (n = 12), 4-OHT E8.5 (n = 6); E10.5: 4-OHT E7.5 (n = 4), 4-OHT E8.5 (n = 4) samples analyzed across six independent experiments. Starting cell numbers ranged from 100–4,300 (YS) and 40–150,000 (CP/AGM). GEMM, granulocyte–erythroid–monocyte/macrophage–megakaryocyte; G/M/GM, granulocyte–monocyte/macrophage; Ery, erythroid. Data are mean ± SD.*:p<0.05, **:p<0.01, ***:p<0.001, ****:p<0.0001, E9.5 CP 4-OHT E7.5: p=0.1204, E9.5 CP 4-OHT E8.5: p= 0.0988 (two-tailed unpaired Student's *t*-test).

(D-E) B cell (**D**) or T cell **(E**) Limiting Dilution Assays (LDA) performed with *Cdh5-CreER^{T2}:: R26^{zsGreen}* E10.5 AGM and YS+VU cells isolated as shown in **Figure S6A,B**. Representative flow cytometric analyses are shown in **Figure S6C,D.** Top tables show, in the left columns, the number of cells seeded per well. Progenitor frequencies were calculated based on the presence of B or T lymphocytes in OP9 or OP9-DL1 co-cultures, respectively. ELDA software was used to perform this calculation and to generate the graphs. Samples were analyzed in 3 (B cell LDA) and 2 (T cell LDA) independent experiments.

Figure 6

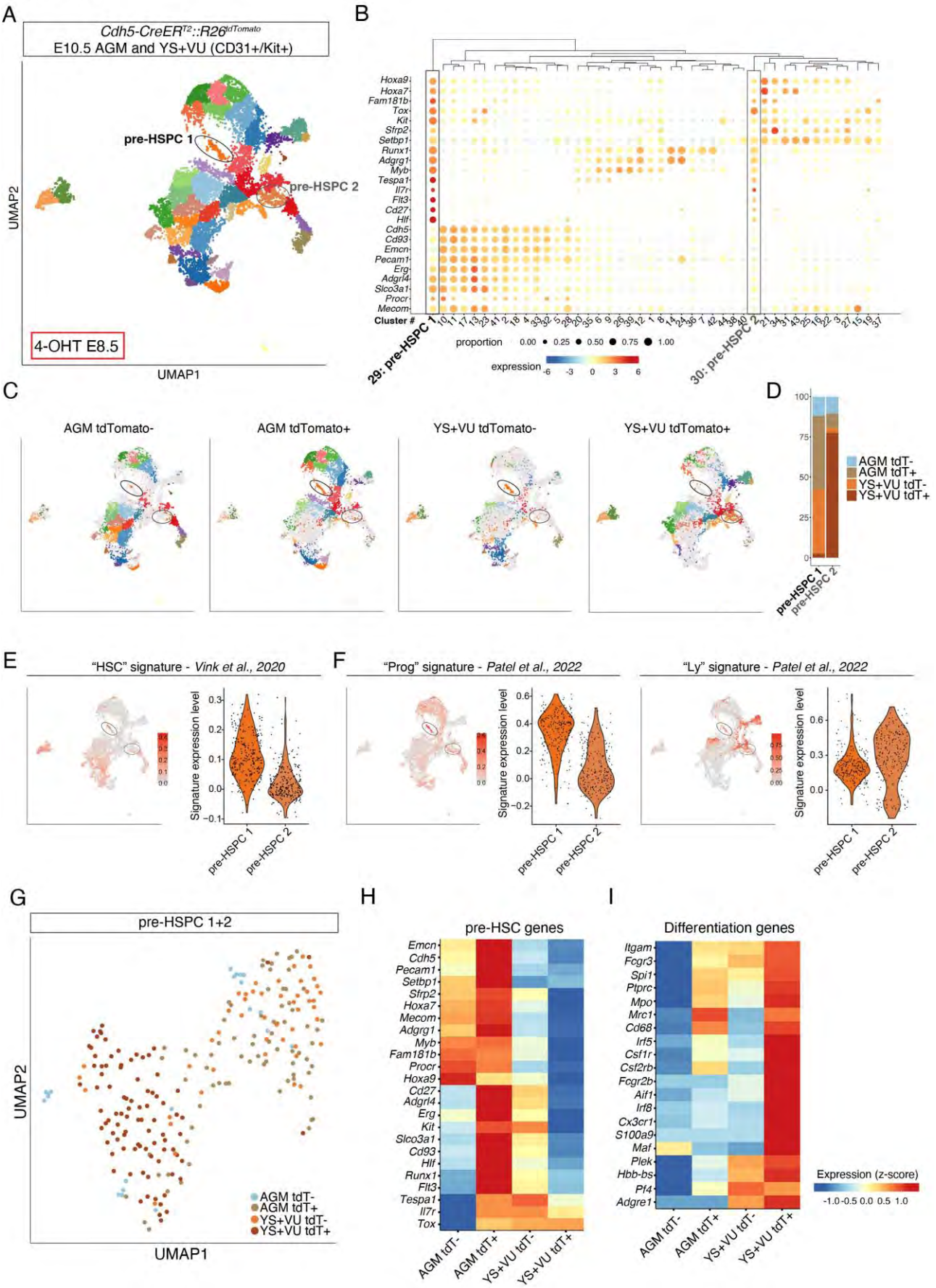

A
Cdh5-CreER^T2::R26^tdTomato
E10.5 AGM and YS+VU (CD31+/Kit+)
pre-HSPC 1
pre-HSPC 2
UMAP2
UMAP1
4-OHT E8.5
B
Hoxa9
Hoxa7
Fam181b
Tox
Kit
Sfrp2
Setbp1
Runx1
Adgrg1
Myb
Tespa1
Il7r
Flt3
Cd27
Hlf
Cdh5
Cd93
Emcn
Pecam1
Erg
Adgrl4
Slco3a1
Procr
Mecom
Cluster #
29: pre-HSPC 1
30: pre-HSPC 2
proportion 0.00 0.25 0.50 0.75 1.00
expression -6 -3 0 3 6
C
AGM tdTomato-
AGM tdTomato+
YS+VU tdTomato-
YS+VU tdTomato+
D
AGM tdT-
AGM tdT+
YS+VU tdT-
YS+VU tdT+
pre-HSPC 1
pre-HSPC 2
E
"HSC" signature - Vink et al., 2020
Signature expression level
pre-HSPC 1
pre-HSPC 2
F
"Prog" signature - Patel et al., 2022
"Ly" signature - Patel et al., 2022
G
pre-HSPC 1+2
UMAP2
UMAP1
AGM tdT-
AGM tdT+
YS+VU tdT-
YS+VU tdT+
H
pre-HSC genes
Emcn
Cdh5
Pecam1
Setbp1
Sfrp2
Hoxa7
Mecom
Adgrg1
Myb
Fam181b
Procr
Hoxa9
Cd27
Adgrl4
Erg
Kit
Slco3a1
Cd93
Hlf
Runx1
Flt3
Tespa1
Il7r
Tox
I
Differentiation genes
Itgam
Fcgr3
Spi1
Ptprc
Mpo
Mrc1
Cd68
Irf5
Csf1r
Csf2rb
Fcgr2b
Aif1
Irf8
Cx3cr1
S100a9
Maf
Plek
Hbb-bs
Pf4
Adgre1
Expression (z-score)
-1.0 -0.5 0.0 0.5 1.0

Figure 6. Combined HE lineage tracing and scRNA-Seq identify distinct subsets of E10.5 pre-HSPCs.

(A) Uniform Manifold Approximation and Projection (UMAP) of 26,180 cells (Ter119- Kit+/CD31+) isolated from AGM and YS+VU of E10.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* embryos activated with 4-OHT at E8.5. AGM tdTomato+ (5,465), AGM tdTomato- (8,135), YS+VU tdTomato+ (4,445) and YS+VU tdTomato- (8,135) cells were separately sequenced (strategy for cell isolation shown in **Figure S6A,B**). Cells were taken from 9 AGM and 14 YS+VU (1 litter of 9 embryos for AGM; 2 litters of 11 and 3 embryos for YS+VU). Cells are colored according to individual cluster identities, and pre-HSPCs clusters 1 and 2 are circled. A complete cell-type annotation is shown in **Figure S7A**.

(B) Bubble plot showing the expression level of HSC and hematopoietic progenitor genes for each cell cluster in the whole dataset. Dot size indicates the percentage of cells expressing each gene, and dot color represents gene expression level. Pre-HSPC 1 and 2 are highlighted with a black and grey box, respectively.

(C) UMAP plots showing cell cluster distribution of AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ samples. Pre-HSPC clusters 1 and 2 are circled.

(D) Bar plot showing the relative contribution of AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ to pre-HSPC clusters 1 and 2.

(E-F) Gene expression of published signatures for **(E)** HSC [55], **(F)** AGM-derived progenitors ("Prog") and lymphoid progenitors ("Ly") [29] in pre-HSPC 1 and pre-HSPC 2 clusters. The average expression of top50 genes for each signature was calculated on single cells using the AddModuleScore function in Seurat.

(G) UMAP plot showing the subclustering of pre-HSPC 1 and 2. Dots represent cells colored according to their origin.

(H) Heatmap showing the relative expression levels of pre-HSC signature genes among AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ cells within pre-HSPC clusters 1 and 2.

(I) Heatmaps showing the relative expression levels of hematopoietic differentiation genes among AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ cells within pre-HSPC clusters 1 and 2.

Figure 7

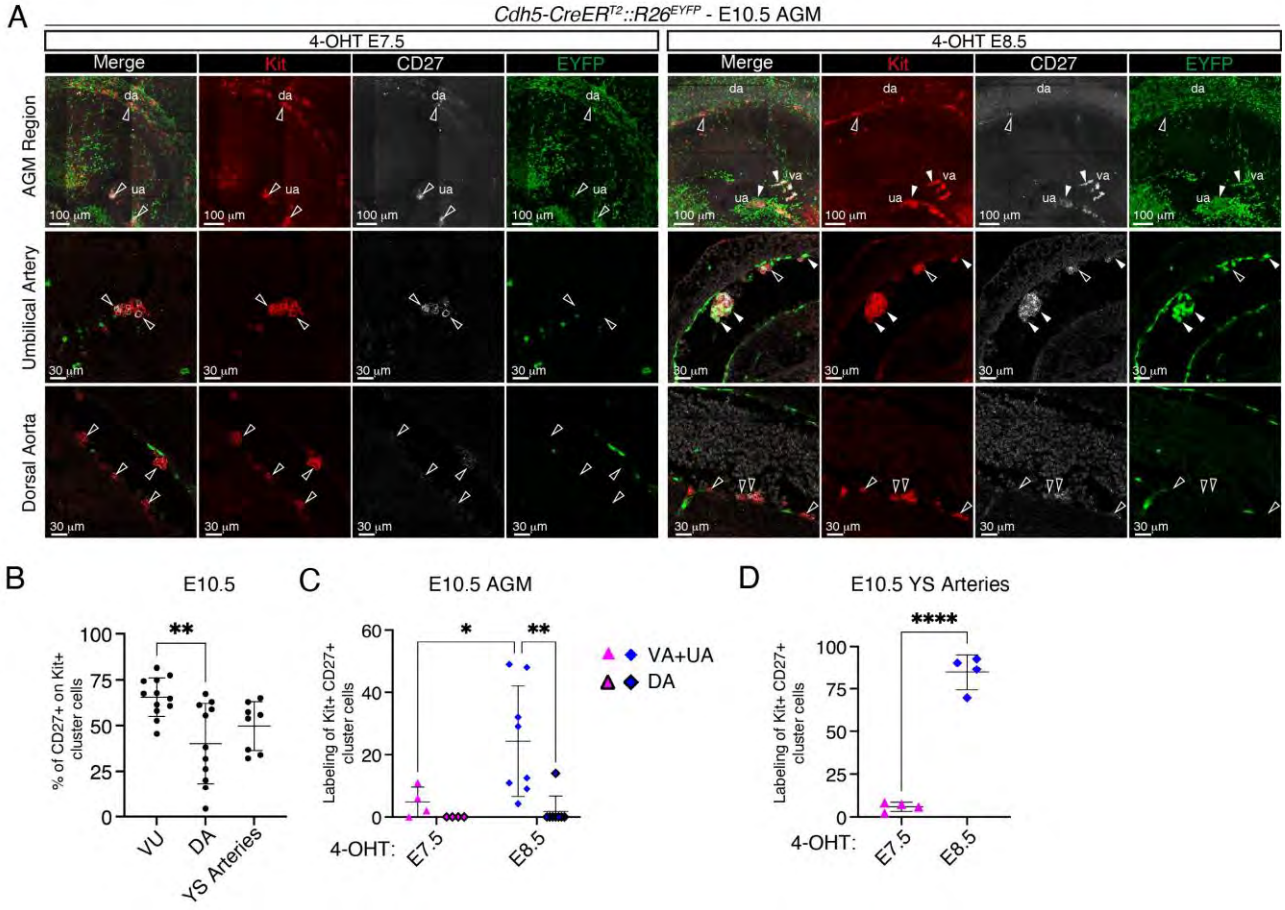


Figure 7. CD27+ hematopoietic clusters emerging from E8.5-E9.5 HE specifically localize to the VU arteries.

(A) WM-IF confocal analysis of E10.5 *Cdh5-CreER^{T2}::R26^{EYFP}* AGM (4-OHT at E7.5 or E8.5). Upper panels show a 3D maximum intensity projection. Middle (UA) and lower (DA) panels show single 2.5 µm-thick optical slices. Arrowheads indicate examples of labeled (EYFP+; filled arrowheads) or unlabeled (EYFP-; open arrowheads) Kit+ CD27+ cells. 4-OHT E7.5 (n = 4) and 4-OHT E8.5 (n = 8) different YS were analyzed in 3 independent experiments. Scale bars: 100 µm (3D), 30 µm (slices).

(B) Quantification of WM-IF confocal analysis shown in **(A)**, representing the percentage of CD27+ cells on the total of Kit+ cells. DA, VU (n=12) and YS arteries (n=8) were analyzed in 7 independent experiments. 3-9 images/sample were quantified; 105 (DA, VU) and 52 (YS arteries) total images. Error bars represent mean ± SD. **:p<0.01 (one-way ANOVA followed by Tukey's multiple comparisons test).

(C) Quantification of WM-IF confocal analysis in **(A)** representing the percentage of labeled EYFP+ cells on the total of Kit+ CD27+ AGM cluster cells. 4-OHT E7.5 (n= 4), and 4-OHT E8.5 (n=8) embryos were analyzed in 4 independent experiments. 5-15 images/sample were used; 34 (4-OHT E7.5), 71 (4-OHT E8.5) total images. Error bars represent mean ± SD. *:p<0.05, **:p<0.01 (two-way ANOVA followed by Tukey's multiple comparisons test).

(D) Quantification of WM-IF confocal analysis shown in **Figure S9B** representing the percentage of labeled EYFP+ cells on the total of Kit+ CD27+ cluster cells in YS arteries. 3-9 images/sample were quantified; 25 (4-OHT E7.5) and 27 (4-OHT E8.5) total images. Error bars represent mean ± SD. ****p<0.0001 (two-tailed unpaired Student's *t*-test).

Figure 8

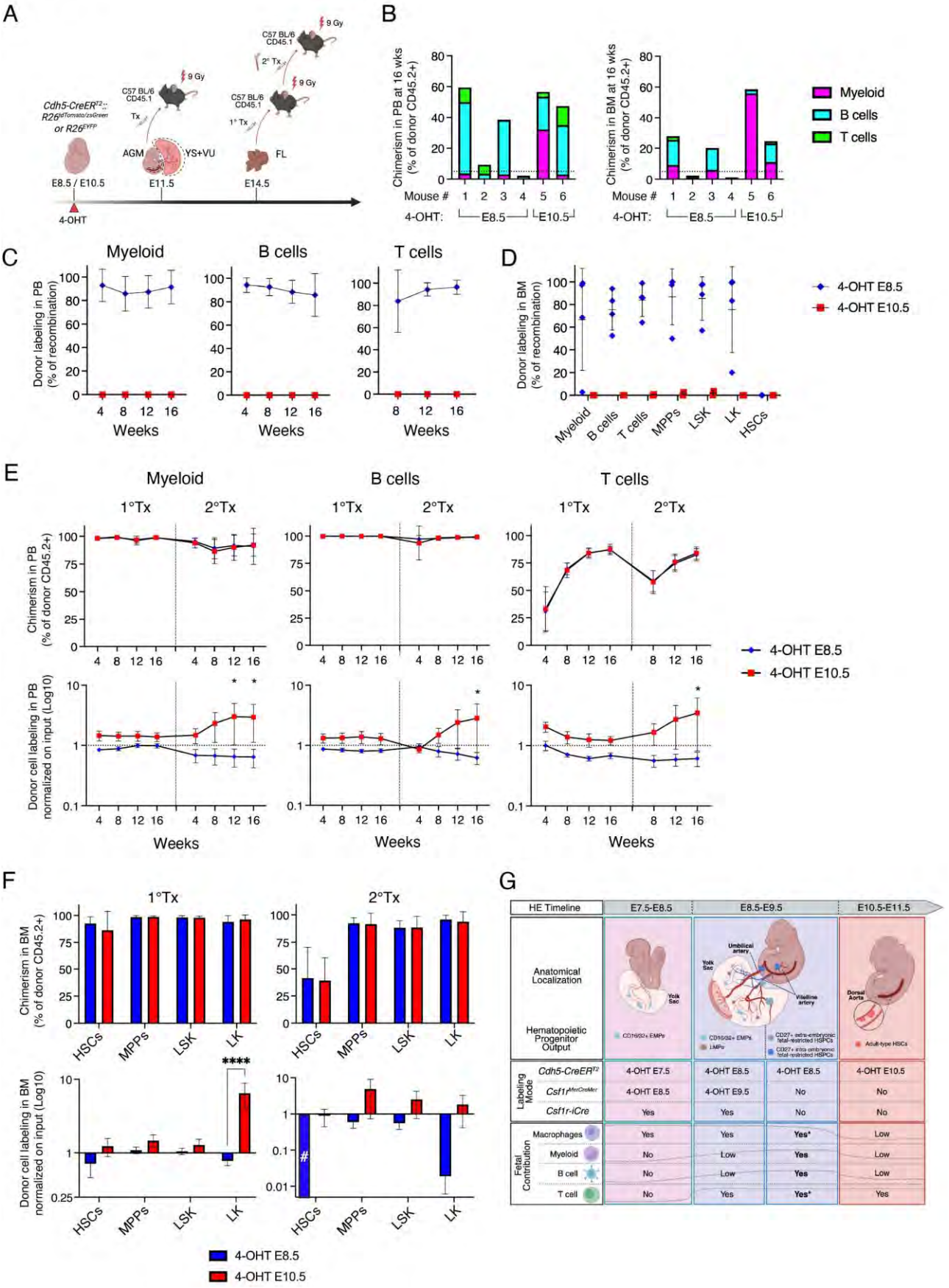

A
B
C
D
E
F
G
Myeloid
B cells
T cells
4-OHT E8.5
4-OHT E10.5
Weeks
1°Tx
2°Tx
HSCs
MPPs
LSK
LK
HE Timeline
E7.5-E8.5
E8.5-E9.5
E10.5-E11.5
Anatomical Localization
Hematopoietic Progenitor Output
Labeling Mode
Fetal Contribution

Figure 8. Fetal-restricted HSPCs yield multi-lineage engraftment, but they do not contain long-term HSCs and their contribution declines in serial transplantations.

(A) Experimental schematic of E11.5 AGM/YS+VU and E14.5 FL transplantation assays.

(B) Percentage (%) of chimerism (% of donor CD45.2+) in adult lethally irradiated mice transplanted with E11.5 *Cdh5-CreER^{T2}::R26$^{tdTomato/zsGreen}$* AGM (2 e.e./recipient), in PB (left) and BM (right) 16 weeks after transplantation. Each bar indicates a single recipient mouse. The dashed line indicates 5%, the threshold we considered for repopulation.

(C) Longitudinal analysis showing the frequency of labeled (zsGreen+ or tdTomato+) donor myeloid (CD45.2+ CD11b+), B cells (CD45.2+ B220+), T cells (CD45.2+ CD3e+) in PB of mice transplanted with E11.5 AGM. Gating strategy shown in **S10A**. Four (4-OHT E8.5) and two (4-OHT E10.5) mice were analyzed in four independent transplant experiments. Data are mean ± SD. A summary of E11.5 transplantation data in provided in **Table S4.**

(D) Labeling frequency of donor myeloid, B cells and T cells, MPPs (Lin- CD45.2+ Kit+ Sca1+CD48+ CD150-), LSK (Lin- CD45.2+ Kit+ Sca1+) LK (Lin- CD45.2+ Kit+ Sca1-) and HSCs (Lin- CD45.2+ Kit+ Sca1+CD48- CD150+) in BM of AGM transplanted mice. Gating strategy shown in **S10B**. Data are mean ± SD.

(E) Longitudinal analysis showing % of PB chimerism (top) and normalized labeling frequency (bottom) in adult lethally irradiated mice transplanted with E14.5 *Cdh5-CreER^{T2}::R26^{EYFP}* or *Cdh5-CreER^{T2}::R26^{tdTomato}* FL cells activated with 4-OHT at E8.5 or E10.5 (1° Tx) or adult BM cells from 1° TX (2° TX). The frequency of labeled (EYFP+/tdTomato+) donor myeloid, B and T cells was normalized on the % of labeled LSK before transplant and shown as Log10. Gating strategy shown in **S10E**. N=11 (4-OHT E8.5 1° Tx), n=13 (4-OHT E10.5 1° Tx), n=10 (4-OHT E8.5 2° Tx) and n=11 (4-OHT E10.5 2° Tx) recipient mice were analyzed in 2 independent experiments. Data are mean ± SD. A summary of PB data is provided in **Table S5** (1°Tx) and **Table S6** (2°Tx). *p<0.05 (two-way ANOVA followed by Tukey's multiple comparisons test).

(F) Percentage of BM chimerism (top) and normalized labeling frequency (bottom) of HSCs, MPPs, LSK, and LK subsets in primary and secondary transplanted mice from (**E**), 16 week after transplant. The frequency of labeled donor HSCs, MPPs, LSKs and LKs was normalized on the initial labeling of each cell subset in the donor tissue (FL for 1° Tx or BM for 2° TX) and shown as Log10. No labeled HSCs were found in 2°Tx recipients with 4-OHT at E8.5. Log10(0) is indicated with #. Data are mean ± SD. A summary of BM data in provided in **Table S5** (1°Tx) and **Table S6** (2°Tx). ****p<0.0001 (two-way ANOVA followed by Tukey's multiple comparisons test).

(G) Conceptual schematic summarizing timeline and anatomical sites of the subsets of HE and hematopoietic progenitors identified within this and other studies, along with labeling modes. Indicated is the contribution to fetal macrophages, myeloid cells, B and T lymphocytes, with E16.5 as reference timepoint. (*): Asterisk

indicates a contribution evaluated using multiple Cre lines; i.e., macrophages and T cells are contributed by Csf1r+ progenitors (*Csf1r*^{MerCreMer}, 4-OHT E9.5). However, because the combined contribution of Csf1r+ progenitors plus fetal-restricted HSPCs (*Cdh5-CreER*^{T2}, 4-OHT E8.5) is higher than the Csf1r+ contribution alone, an exclusive contribution of fetal-restricted HSPCs can be inferred. Figure created with BioRender.com.

METHODS

Mice and embryos

Cdh5-CreER^{T2} [35], *Csf1r-iCre* [72], *Csf1r^{MerCreMer}* [73], *R26^{zsGreen}*, *R26^{tdTomato}* [74] and *R26^{EYFP}* [75] transgenic mice were previously described and were genotyped according to reported protocols (further details and primers listed in **Table S1**). *R26^{zsGreen}*, *R26^{tdTomato}* or *R26^{EYFP}* females aged 6 to 16 weeks were subjected to overnight timed matings with *Cdh5-CreER^{T2}, Csf1r-iCre or Csf1r^{MerCreMer}* males. Successful mating was judged by the presence of vaginal plugs the morning after, which was considered 0.5 days post conception (E0.5). Embryos were collected and dissected as previously described [45, 51, 76]. E9.5-E11.5 embryos were carefully staged by counting of somite pairs; older embryos were staged by morphological criteria. For *Cdh5-CreER^{T2}* fate mapping, a single dose of 37.5 mg/kg of 4-hydroxytamoxifen (4-OHT) dissolved in corn oil was delivered by intra-peritoneal (i.p.) injections to pregnant females at E7.5, E8.5 or E10.5. To counteract adverse effects of 4-OHT on pregnancies, 4-OHT solutions were supplemented with progesterone (18.75 mg/kg). For *Csf1r^{MerCreMer}* fate mapping, a single dose of 75 mg/kg 4-OHT without progesterone was used.

All transgenic mouse lines were maintained on a CD45.2 C57BL/6 genetic background, with the exception of females used for *Csf1r^{MerCreMer}* timed matings and other matings for generation of adult mice with 4-OHT activation during embryogenesis, which were instead of C57BL/6/FVB mixed background (F1). Mice were housed with free access to food and water at the San Raffaele Scientific Institute, University of Oxford and University of Milan Institutional mouse facilities. All experiments were performed in accordance with experimental protocols approved by San Raffaele Scientific Institute and University of Milan Institutional Animal Care and Use Committees (IACUC) and the Italian Ministry of Health. All procedures carried out in Oxford were in compliance with United Kingdom Home Office regulations and the Oxford University Clinical Medicine Animal Welfare and Ethical Review Committee.

Flow cytometry analysis and cell sorting

Single cell suspensions were obtained from embryonic tissues (yolk sac, embryo/caudal part, fetal liver) by incubating for 15 min at 37 °C in calcium/magnesium free PBS supplemented with FBS 10%, Penicillin-Streptomycin 1%, EDTA 2mM and collagenase type I (Sigma) 0.12% (w/v), followed by mechanical dissociation by pipetting. Single cell suspension from fetal thymus was obtained by mechanical dissociation and passed through a 26-gauge needle. Peripheral blood (PB) samples were collected by tail vein bleeding using a scalpel; bone marrow (BM) was obtained by

flushing long bones using a syringe and filtered in 40μm strainers. PB, BM and fetal liver (FL) samples were treated with the appropriate amount of RBC Lysis Buffer. Single cell suspensions were incubated with conjugated antibodies and processed for flow cytometry as previously described [45, 51, 76]. A list of antibodies used for flow cytometry can be found in **Table S1**. Voltages, compensation and gates were set using unstained, single stained and fluorescence-minus-one (FMO) controls. Dead cells were excluded based on Hoechst 33258 (Hellobio) or 7-AAD (Sigma) incorporation. Flow cytometry data acquisition was carried out using a LSR Fortessa X-20 (BD) analyzer and BD FACSDiva software (version 8.0.2). Cell sorting was performed using a MoFLO Astrios cell sorter equipped with Summit software version 6.3 (both from Beckman Coulter), a BD FACSDiscover S8 with FACS Chorus software, or FACSAria II with BD FACSDiva software. An average sorting rate of 500-1000 events per second at a sorting pressure of 25 psi with a 100μm nozzle was maintained. Flow cytometric data was analyzed using FlowJo software version 10 (BD).

Colony-forming unit-culture (CFU-C) assays

YS+VU and caudal parts were dissected from E9.5 (21-27 sp) or E10.5 (32-36 sp) *Cdh5-CreER^{T2}::R26^{tdTomato}* , *Cdh5-CreER^{T2}::R26^{zsGreen}* or *Cdh5-CreER^{T2}::R26^{EYFP}* concepti. Three YS+VU and caudal parts were each pooled and processed into single cell suspensions. Labeled and unlabeled cells were isolated by flow cytometry. CFU-C assays were performed using Methocult M3434 (Stem Cell Technologies), as per manufacturer's instructions. Cells were plated in duplicate dishes and cultured at 37°C, 5% CO_2 in a humidified chamber. Colonies were scored after 7 days. Number of colonies are normalized to 1,000 Kit+ cells seeded, considering the average of percentage of Kit+ cells in labeled and unlabeled fraction of Ter119- cells of YS+VU and caudal parts from E9.5 or E10.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* , *Cdh5-CreER^{T2}::R26^{zsGreen}* or *Cdh5-CreER^{T2}::R26^{EYFP}* concepti, activated with 4-OHT at E7.5 or E8.5.

Limiting dilution OP9 and OP9-Dl1 co-cultures

For limiting dilution assays (LDA), YS+VU and AGM were dissected from E10.5 (32-36 sp) *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos activated with 4-OHT at E8.5. YS+VU and AGM from an entire litter (7 to 12 e.e.) were each pooled and processed into single cell suspensions. Sorted hemato-endothelial cells (sorting strategy shown in **Figure S6A,B**) were plated on confluent OP9 or Delta-like 1-expressing OP9 (OP9-Dl1)[77] stromal cells in 96-well plates at the doses of 10, 30, 100 or 200 cells per well. For B-cell differentiation, cultures were maintained in α MEM with 10% FBS supplemented

with 10 ng/ml IL-7 and 10 ng/ml Flt3 ligand. OP9 feeders and medium were refreshed every 4-5 days according to the status of the stromal layer, and non-adherent cells were collected and analyzed by flow cytometry on day 10. For T-cell differentiation, cultures were maintained in αMEM with 20% FBS supplemented with 10 ng/ml IL-7 and 10 ng/ml Flt3 ligand. OP9-Dl1 feeders and medium were refreshed every 4-5 days according to the status of the stromal layer, and non-adherent cells were collected and analyzed by flow cytometry on day 14. ELDA software was used to calculate the frequency of B and T cell progenitors [78].

Whole-mount immunofluorescence analysis and imaging

Whole-mount immunofluorescence (WM-IF) was performed as previously described [45, 79]. Briefly, embryos and yolk sacs were dissected and fixed in a 4% paraformaldehyde solution in PBS for 30 minutes to 2 hours at 4°C. For E10.5 embryos, limb buds and body wall were removed before fixation to expose the aorta. Next, samples were treated with a permeabilizing-blocking solution (0.2% Triton X-100, 2% donkey serum, 2% FBS) and incubated overnight with primary antibodies. A second step of incubation with appropriate secondary antibodies was then carried out. Antibodies used for WM-IF are listed in **Table S1**. After staining, embryos were cleared in a benzyl alcohol-benzyl benzoate solution (BABB) and mounted as previously described [79]. YS were cleared overnight in a 50% solution of glycerol in PBS at 4°C and then flat-mounted on Superfrost glass slides. Samples were imaged using a Zeiss 710 confocal microscope equipped with a LD LCI Plan-Apochromat 25x/0.8 Imm Corr DIC M27 objective or an EC Plan-Neofluar 40x/1.30 Oil DIC M27 objective. Confocal image acquisition was carried out using Zeiss Zen software version 2.3 SP1; image processing and analysis was carried out using ImarisViewer software version 10.2 and earlier versions (Bitplane), ImageJ/Fiji (versions 2.3.5-2.9.0) and Adobe Photoshop 2024 and earlier versions.

Single-Cell RNA Sequencing (scRNA-Seq) of E10.5 AGM and YS+VU

Aorta-Gonad-Mesonephros (AGM) and YS+VU dissected from E10.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* embryos (31-37 sp; 4-OHT at E8.5) were processed into single cell suspensions as described above. Live Ter119- CD31+/Kit+ cells were isolated by FACS as shown in **Figure S6D,E** and split into tdTomato+ and tdTomato-. Labeled and unlabeled cells were analyzed separately.

scRNA-Seq libraries were generated using a Chromium instrument (10x Genomics) with a Next GEM Single Cell 3' kit. Libraries were quantified using a Qubit fluorometer (Thermo Fisher) and their

profile was analyzed using a TapeStation instrument (Agilent). NGS sequences were generated using a Novaseq 6000 instrument (Illumina) with a target of 25000 reads/cell. Following multiplexing, raw fastq reads were processed using *cellranger* v6.1 and aligned against the mm10 mouse genome (GENCODE vM23/Ensembl 98) modified with the *mkref* tool to add an artificial '*tdTomato*' chromosome. The associated genome annotation GTF file was modified accordingly. Filtered count matrices generated with *cellranger* were processed with *Seurat* v4.0 [80] package implemented in R (versions 3.2.3 – 4.2.1). Cells with genes count > 300 and < 8000 and fraction of mitochondrial reads < 0.20 were kept for downstream processing.

After filtering, the dataset included 26,180 cells (5,465 AGM tdTomato+, 8,135 AGM tdTomato-, 4,445 YS+VU tdTomato+ and 8,135 YS+VU tdTomato-). After converting individual matrices in Seurat objects via the Read10X function, data were normalized and transformed using the *SCTransform* Variance Stabilizing Transformation using the *glmGamPoi* method, while also regressing-out for feature count, percent of mitochondrial counts and cell phase. Data generated from both samples were subsequently integrated with a Canonical Correlation Analysis using the *PrepSCTIntegration, FindIntegrationAnchors* and *IntegrateData* commands, by using SCT as the normalization method. Dimensionality reduction of the integrated data was initially carried-on using Principal Component Analysis (PCA) and subsequently with Uniform Manifold Approximation and Projection (UMAP) algorithms, by retaining the first 30 principal components of the PCA. Clusters were identified with the Louvain algorithm; their number was selected using the Clustree tool [81] (version 0.4.3) by maximizing cluster stability. Individual cell types were identified by using SingleR [82] (version 1.0.1) as well as with manual data curation.

Expression of publicly available gene signatures (top 50 DEGs) for nascent embryonic HSC and progenitors [29, 55] was assessed in our dataset using the AddModuleScore function in Seurat (ctrl = 100).

GO biological process gene set enrichment analysis (GSEA)

Gene Set Enrichment analysis (GSEA) was conducted using a ranked list of differentially expressed genes (DEG) ($p < 0.05$) between pre-HSPC 2 and pre-HSPC 1 (**Figure S7D**) and a ranked list of DEG between AGM tdT+ and YS+VU tdT+ pre-HSPCs (**Figure S8D**). Analyses were carried out using R package ClusterProfiler (version 4.8.3) [83] querying the Gene Ontology (GO) database. The analysis was performed by setting the number of permutations to 10000, the minimum gene set to 3 and maximum to 800. GO terms were considered significant with a selected cutoff p-value of 0.05.

In vivo transplantation

For AGM and YS+VU transplantation experiments, tissues were dissected from E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos (4-OHT at E8.5 or E10.5). Single cell suspensions were prepared as described above. Syngeneic C57BL/6 (CD45.1) recipient mice were lethally irradiated (9 Gy, split dose) before intra-venous transplantation of single cell suspensions obtained from 2 embryo equivalents of each tissue. For FL transplantation experiments, syngeneic C57BL/6 (CD45.1) recipient mice were lethally irradiated (9 Gy, split dose) before intra-venous transplantation of $1x10^6$ unfractionated FL cells (primary transplantation) or $2x10^6$ adult BM cells (secondary transplantation) from E14.5 *Cdh5-CreER^{T2}::R26^{TdTomato}* or *Cdh5-CreER^{T2}::R26^{EYFP}* mice (4-OHT at E8.5 or E10.5). For both AGM and YS+VU transplantation experiments, and FL transplantation experiments, donor-derived chimerism and percentage of labeling (tdTomato+, EYFP+ or zsGreen+) within donor cells was determined by flow cytometry in PB at 4, 8 and 12 weeks post transplantation, and in PB and BM at 16 weeks post transplantation. BM from primary and secondary transplanted mice was analyzed by flow cytometry to determine the percentage of labeling of donor hematopoietic stem/progenitor cells. A comprehensive summary of transplantation analysis data in provided in **Table S4** (E11.5 AGM transplants), **Table S5** (FL primary transplants) and **Table S6** (FL secondary transplants).

Quantification and statistical analysis

Statistical analyses were performed using GraphPad Prism v10.2.1 and later versions. No specific randomization method was used. Animals were allocated into experimental groups according to their genotype. No specific methods were used for blinding, but in general samples were collected from mice by one individual and then processed and analyzed by different individuals, at which time genotypes or experimental conditions of each sample were not known. To determine the level of significance, unpaired two-tailed Student *t*-test, one-way and two-way ANOVA followed by Tukey's multiple comparisons test were used as indicated in figure legends. $P<0.05$ was considered statistically significant, and the level of significance is indicated by asterisks: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

Data availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Emanuele Azzoni (Emanuele.azzoni@unimib.it)

Materials availability

This study did not generate any new unique reagents.

Data and code availability

- scRNA-seq data of YS+VU and AGM from E10.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* embryos (4-OHT at E8.5) have been deposited in the Sequence Read Archive (SRA) data repository (NCBI) with the accession number BioProject ID: PRJNA898269 (sample accession numbers: SRR28006358, SRR28006359, SRR28006360, SRR28006361). Data will be made publicly available upon publication.
- This study does not report original code.
- Any additional information required to reanalyze the data reported in this study is available from the lead contact upon reasonable request.

References

1. Barone, C., Orsenigo, R., Meneveri, R., Brunelli, S. & Azzoni, E. One Size Does Not Fit All: Heterogeneity in Developmental Hematopoiesis. *Cells* **11** (2022).
2. Dzierzak, E. & Bigas, A. Blood Development: Hematopoietic Stem Cell Dependence and Independence. *Cell Stem Cell* **22**, 639–651 (2018).
3. Karlsson, G., Sommarin, M.N.E. & Boiers, C. Defining the Emerging Blood System During Development at Single-Cell Resolution. *Front Cell Dev Biol* **9**, 660350 (2021).
4. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86**, 897–906 (1996).
5. Bertrand, J.Y. *et al.* Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108–111 (2010).
6. Boisset, J.C. *et al.* In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* **464**, 116–120 (2010).
7. de Bruijn, M.F., Speck, N.A., Peeters, M.C. & Dzierzak, E. Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J* **19**, 2465–2474 (2000).
8. Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–115 (2010).
9. Rybtsov, S. *et al.* Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43(-) embryonic precursor. *Stem Cell Reports* **3**, 489–501 (2014).
10. Rybtsov, S. *et al.* Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region. *J Exp Med* **208**, 1305–1315 (2011).
11. Rybtsov, S., Ivanovs, A., Zhao, S. & Medvinsky, A. Concealed expansion of immature precursors underpins acute burst of adult HSC activity in foetal liver. *Development* **143**, 1284–1289 (2016).
12. Hall, T.D. *et al.* Murine fetal bone marrow does not support functional hematopoietic stem and progenitor cells until birth. *Nat Commun* **13**, 5403 (2022).
13. Palis, J. Hematopoietic stem cell-independent hematopoiesis: emergence of erythroid, megakaryocyte, and myeloid potential in the mammalian embryo. *FEBS Lett* **590**, 3965–3974 (2016).
14. Frame, J.M., Fegan, K.H., Conway, S.J., McGrath, K.E. & Palis, J. Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity. *Stem Cells* **34**, 431–444 (2016).
15. Gomez Perdiguero, E. *et al.* Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547–551 (2015).
16. Hoeffel, G. *et al.* C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. *Immunity* **42**, 665–678 (2015).
17. Soares-da-Silva, F. *et al.* Yolk sac, but not hematopoietic stem cell-derived progenitors, sustain erythropoiesis throughout murine embryonic life. *J Exp Med* **218** (2021).
18. Dege, C. *et al.* Potently Cytotoxic Natural Killer Cells Initially Emerge from Erythro-Myeloid Progenitors during Mammalian Development. *Dev Cell* **53**, 229–239 e227 (2020).
19. McGrath, K.E. *et al.* Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Rep* **11**, 1892–1904 (2015).
20. Boiers, C. *et al.* Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell* **13**, 535–548 (2013).
21. Kobayashi, M. *et al.* Functional B-1 progenitor cells are present in the hematopoietic stem cell-deficient embryo and depend on Cbfbeta for their development. *Proc Natl Acad Sci U S A* **111**, 12151–12156 (2014).
22. Yoshimoto, M. *et al.* Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence. *Blood* **119**, 5706–5714 (2012).
23. Kobayashi, M. *et al.* HSC-independent definitive hematopoiesis persists into adult life. *Cell Rep* **42**, 112239 (2023).
24. Elsaid, R. *et al.* A wave of bipotent T/ILC-restricted progenitors shapes the embryonic thymus microenvironment in a time-dependent manner. *Blood* **137**, 1024–1036 (2021).

25. Luis, T.C. *et al.* Initial seeding of the embryonic thymus by immune-restricted lympho-myeloid progenitors. *Nat Immunol* **17**, 1424–1435 (2016).
26. Chen, M.J. *et al.* Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* **9**, 541–552 (2011).
27. Ulloa, B.A. *et al.* Definitive hematopoietic stem cells minimally contribute to embryonic hematopoiesis. *Cell Rep* **36**, 109703 (2021).
28. Yokomizo, T. *et al.* Independent origins of fetal liver haematopoietic stem and progenitor cells. *Nature* **609**, 779–784 (2022).
29. Patel, S.H. *et al.* Lifelong multilineage contribution by embryonic-born blood progenitors. *Nature* **606**, 747–753 (2022).
30. Shang, F. *et al.* Multipotent progenitors with distinct origins, clonal lineage fates, transcriptomes, and surface markers yield two hematopoietic trees. *Proc Natl Acad Sci U S A* **122**, e2505510122 (2025).
31. Dignum, T. *et al.* Multipotent progenitors and hematopoietic stem cells arise independently from hemogenic endothelium in the mouse embryo. *Cell Rep* **36**, 109675 (2021).
32. Beaudin, A.E. *et al.* A Transient Developmental Hematopoietic Stem Cell Gives Rise to Innate-like B and T Cells. *Cell Stem Cell* **19**, 768–783 (2016).
33. Ganuza, M. *et al.* Murine foetal liver supports limited detectable expansion of life-long haematopoietic progenitors. *Nat Cell Biol* **24**, 1475–1486 (2022).
34. Yokomizo, T. & Suda, T. Development of the hematopoietic system: expanding the concept of hematopoietic stem cell-independent hematopoiesis. *Trends Cell Biol* **34**, 161–172 (2024).
35. Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
36. Fowler, J.L. *et al.* Lineage-tracing hematopoietic stem cell origins in vivo to efficiently make human HLF+ HOXA+ hematopoietic progenitors from pluripotent stem cells. *Dev Cell* **59**, 1110–1131 e1122 (2024).
37. Martinez-Corral, I. & Makinen, T. Genetic Lineage Tracing of Lymphatic Endothelial Cells in Mice. *Methods Mol Biol* **1846**, 37–53 (2018).
38. Kierdorf, K. *et al.* Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways. *Nat Neurosci* **16**, 273–280 (2013).
39. Zhou, F. *et al.* Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487–492 (2016).
40. Gentek, R. *et al.* Hemogenic Endothelial Fate Mapping Reveals Dual Developmental Origin of Mast Cells. *Immunity* **48**, 1160–1171 e1165 (2018).
41. Ramond, C. *et al.* Two waves of distinct hematopoietic progenitor cells colonize the fetal thymus. *Nat Immunol* **15**, 27–35 (2014).
42. Azzoni, E. & Fantin, A. Fetal liver hematopoiesis revisited: a precast hierarchy. *Nature Cardiovascular Research* **1**, 872–873 (2022).
43. Boisset, J.C. *et al.* Progressive maturation toward hematopoietic stem cells in the mouse embryo aorta. *Blood* **125**, 465–469 (2015).
44. Yokomizo, T. & Dzierzak, E. Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos. *Development* **137**, 3651–3661 (2010).
45. Azzoni, E. *et al.* The onset of circulation triggers a metabolic switch required for endothelial to hematopoietic transition. *Cell Rep* **37**, 110103 (2021).
46. Yzaguirre, A.D. & Speck, N.A. Insights into blood cell formation from hemogenic endothelium in lesser-known anatomic sites. *Dev Dyn* **245**, 1011–1028 (2016).
47. Gordon-Keylock, S., Sobiesiak, M., Rybtsov, S., Moore, K. & Medvinsky, A. Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs. *Blood* **122**, 2338–2345 (2013).
48. Taoudi, S. & Medvinsky, A. Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta. *Proc Natl Acad Sci U S A* **104**, 9399–9403 (2007).
49. Ganuza, M. *et al.* Murine hematopoietic stem cell activity is derived from pre-circulation embryos but not yolk sacs. *Nat Commun* **9**, 5405 (2018).

50. Yokomizo, T. *et al.* Hlf marks the developmental pathway for hematopoietic stem cells but not for erythro-myeloid progenitors. *J Exp Med* **216**, 1599–1614 (2019).
51. Swiers, G. *et al.* Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. *Nat Commun* **4**, 2924 (2013).
52. Atkins, M.H. *et al.* Modeling human yolk sac hematopoiesis with pluripotent stem cells. *J Exp Med* **219** (2022).
53. Calvanese, V. *et al.* Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature* **604**, 534–540 (2022).
54. Li, Y., Gao, L., Hadland, B., Tan, K. & Speck, N.A. CD27 marks murine embryonic hematopoietic stem cells and type II prehematopoietic stem cells. *Blood* **130**, 372–376 (2017).
55. Vink, C.S. *et al.* Iterative Single-Cell Analyses Define the Transcriptome of the First Functional Hematopoietic Stem Cells. *Cell Rep* **31**, 107627 (2020).
56. Gao, S. *et al.* Identification of HSC/MPP expansion units in fetal liver by single-cell spatiotemporal transcriptomics. *Cell Res* **32**, 38–53 (2022).
57. Kim, P.G. *et al.* Interferon-alpha signaling promotes embryonic HSC maturation. *Blood* **128**, 204–216 (2016).
58. Sawamiphak, S., Kontarakis, Z. & Stainier, D.Y. Interferon gamma signaling positively regulates hematopoietic stem cell emergence. *Dev Cell* **31**, 640–653 (2014).
59. Espin-Palazon, R. *et al.* Proinflammatory signaling regulates hematopoietic stem cell emergence. *Cell* **159**, 1070–1085 (2014).
60. Li, Y. *et al.* Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. *Genes Dev* **28**, 2597–2612 (2014).
61. Wang, R. & Amoyel, M. mRNA Translation Is Dynamically Regulated to Instruct Stem Cell Fate. *Front Mol Biosci* **9**, 863885 (2022).
62. Morino-Koga, S. & Yokomizo, T. Deciphering hematopoietic stem cell development: key signaling pathways and mechanisms. *Front Cell Dev Biol* **12**, 1510198 (2024).
63. Cain, T.L., Derecka, M. & McKinney-Freeman, S. The role of the haematopoietic stem cell niche in development and ageing. *Nat Rev Mol Cell Biol* **26**, 32–50 (2025).
64. Xia, J. *et al.* Activation of lineage competence in hemogenic endothelium precedes the formation of hematopoietic stem cell heterogeneity. *Cell Res* **33**, 448–463 (2023).
65. Neo, W.H. *et al.* Single-cell profiling reveals three endothelial-to-hematopoietic transitions with divergent isoform expression landscapes. *Nat Cardiovasc Res* (2025).
66. Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat Cell Biol* **17**, 580–591 (2015).
67. Randolph, L.N., Castiglioni, C., Tavian, M., Sturgeon, C.M. & Ditadi, A. Bloodhounds chasing the origin of blood cells. *Trends Cell Biol* (2025).
68. Berthault, C. *et al.* Asynchronous lineage priming determines commitment to T cell and B cell lineages in fetal liver. *Nat Immunol* **18**, 1139–1149 (2017).
69. Kawamoto, H., Ikawa, T., Ohmura, K., Fujimoto, S. & Katsura, Y. T cell progenitors emerge earlier than B cell progenitors in the murine fetal liver. *Immunity* **12**, 441–450 (2000).
70. Nakamura-Ishizu, A., Ito, K. & Suda, T. Hematopoietic Stem Cell Metabolism during Development and Aging. *Dev Cell* **54**, 239–255 (2020).
71. Li, Y. *et al.* Single-Cell Analysis of Neonatal HSC Ontogeny Reveals Gradual and Uncoordinated Transcriptional Reprogramming that Begins before Birth. *Cell Stem Cell* **27**, 732–747 e737 (2020).
72. Deng, L. *et al.* A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *Am J Pathol* **176**, 952–967 (2010).
73. Qian, B.Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* **475**, 222–225 (2011).
74. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* **13**, 133–140 (2010).
75. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* **1**, 4 (2001).

76. Azzoni, E. *et al.* Kit ligand has a critical role in mouse yolk sac and aorta-gonad-mesonephros hematopoiesis. *EMBO Rep* **19** (2018).
77. Schmitt, T.M. *et al.* Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro. *Nat Immunol* **5**, 410–417 (2004).
78. Hu, Y. & Smyth, G.K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* **347**, 70–78 (2009).
79. Yokomizo, T. *et al.* Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos. *Nat Protoc* **7**, 421–431 (2012).
80. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 e3529 (2021).
81. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7** (2018).
82. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163–172 (2019).
83. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).

SUPPLEMENTAL FIGURE TITLES AND LEGENDS

Figure S1

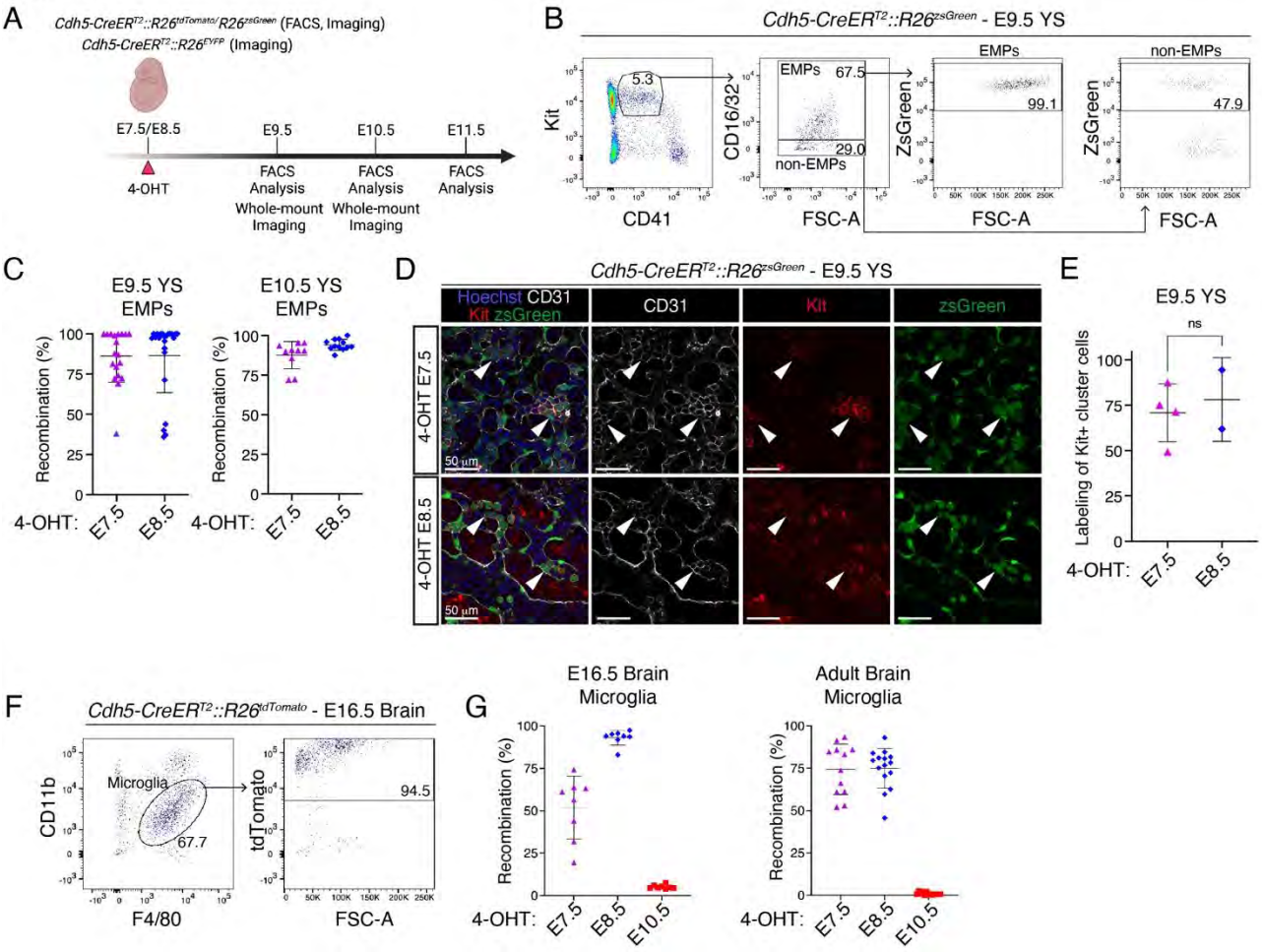


Figure S1 (Related to Figure 1). Lineage tracing analysis of EMPs and microglia in *Cdh5-CreER^{T2}* mice.

(A) Visual schematic of lineage tracing experiments in E9.5, E10.5 and E11.5 embryos.

(B) Representative flow cytometric analysis of labeled EMPs (Ter119- CD41^{lo} Kit+ CD16-32+) and non-EMPs (Ter119- CD41^{lo} Kit+ CD16-32-) in E9.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* yolk sac (YS), activated with 4-OHT E7.5.

(C) Quantification of flow cytometric analysis in **(B)**. E9.5 4-OHT E7.5 (n = 19), E9.5 4-OHT E8.5 (n = 23), E10.5 4-OHT E7.5 (n = 10), E10.5 4-OHT E8.5 (n = 12) YS were analyzed individually in 5 independent experiments. Error bars represent mean ± SD.

(D) Confocal whole mount immunofluorescence (WM-IF) analysis of E9.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* YS. Arrowheads indicate CD31+ Kit+ hematopoietic cell clusters, labeled with 4-OHT at E7.5 (top) and E8.5 (bottom). 4-OHT E7.5 (n = 4), and 4-OHT E8.5 (n = 2) YS analyzed. Scale bar: 50 µm.

(E) Quantification of WM-IF analysis in **(D)**. 3-9 images/sample were quantified; 12 (4-OHT E7.5) and 6 (4-OHT E8.5) total images Error bars represent mean ± SD. ns = non-significant (two-tailed unpaired Student's *t*-test).

(F) Representative flow cytometric analysis of brain microglia (CD45+ $CD11b^{low}$ F4/80+) in *Cdh5-CreER^{T2}::R26^{tdTomato}* E16.5 embryos, activated with 4-OHT at E7.5, E8.5 (shown here) or E10.5. 4-OHT E7.5 (n = 8), 4-OHT E8.5 (n = 8), 4-OHT E10.5 (n = 8) were analyzed individually in 3 independent experiments.

(G) Quantification of flow cytometric analysis of labeled brain microglia as shown in **(F)** in *Cdh5-CreER^{T2}::R26^{tdTomato}* E16.5 embryos (left) and 2-months-old adult mice (right), activated with 4-OHT at E7.5, E8.5 or E10.5. Replicates as shown. Error bars represent mean ± SD.

Figure S2

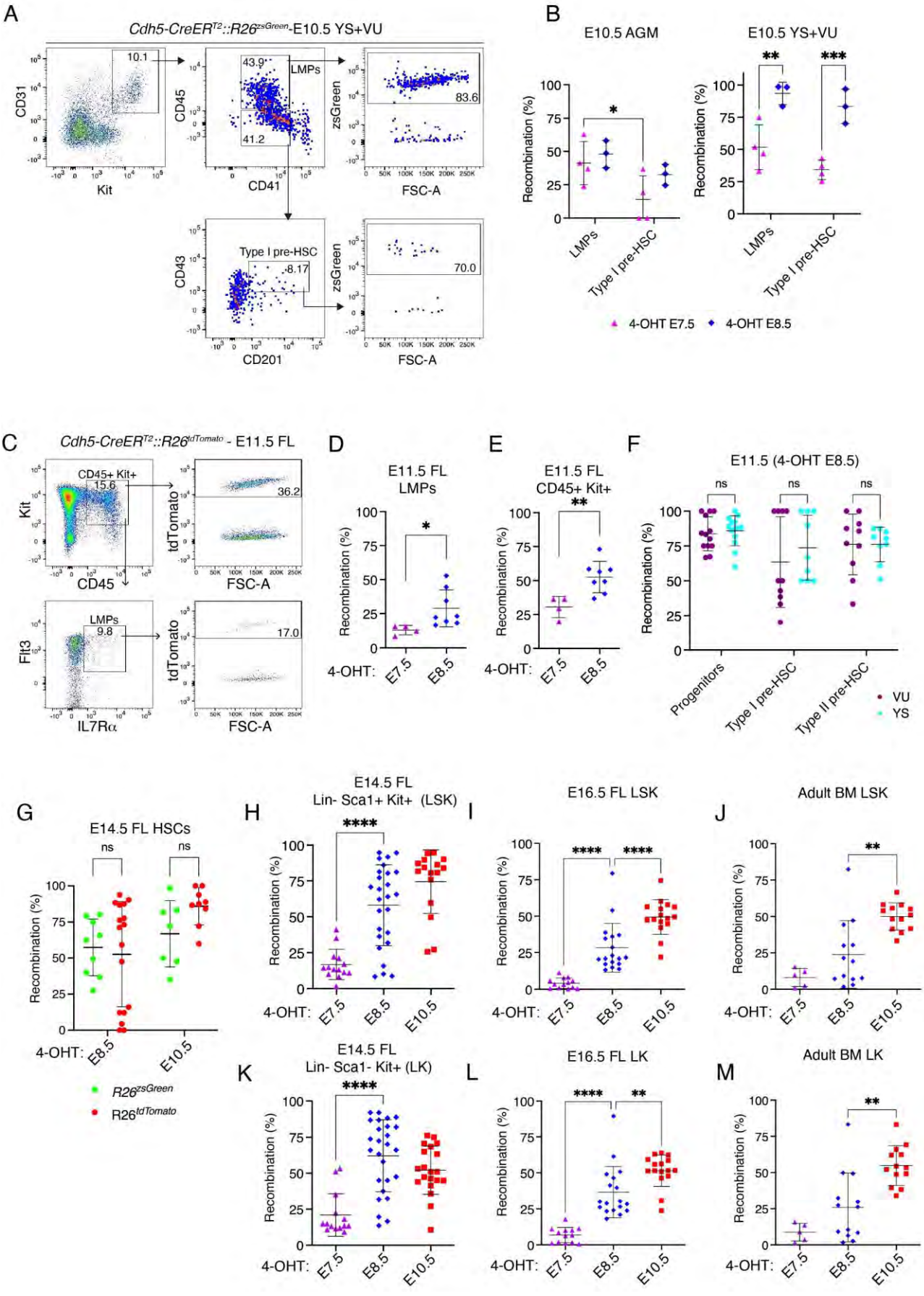

A
Cdh5-CreER^T2::R26^zsGreen-E10.5 YS+VU
CD31
Kit
10.1
CD45
CD41
43.9
LMPs
41.2
zsGreen
FSC-A
83.6
CD43
CD201
Type I pre-HSC
8.17
70.0
B
E10.5 AGM
E10.5 YS+VU
Recombination (%)
LMPs
Type I pre-HSC
4-OHT E7.5
4-OHT E8.5
C
Cdh5-CreER^T2::R26^tdTomato - E11.5 FL
CD45+ Kit+
15.6
tdTomato
36.2
Flt3
IL7Rα
9.8
17.0
D
E11.5 FL LMPs
E
E11.5 FL CD45+ Kit+
F
E11.5 (4-OHT E8.5)
ns
Progenitors
Type I pre-HSC
Type II pre-HSC
VU
YS
4-OHT:
E7.5
E8.5
G
E14.5 FL HSCs
E10.5
R26^zsGreen
R26^tdTomato
H
E14.5 FL Lin- Sca1+ Kit+ (LSK)
I
E16.5 FL LSK
J
Adult BM LSK
K
E14.5 FL Lin- Sca1- Kit+ (LK)
L
E16.5 FL LK
M
Adult BM LK

Figure S2 (Related to Figure 1). Flow cytometric analysis of LMPs, LSK and LK labeling in *Cdh5-CreER^{T2}* embryos and adult mice.

(A) Representative flow cytometric analysis of LMPs (Ter119- CD31+ Kit+ CD45+ CD41+) and type I pre-HSC (Ter119- CD31+ Kit+ CD45- CD41+ CD43+ CD201+) in E10.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* AGM and YS+VU, labeled with 4-OHT at E7.5, E8.5 (shown here). 4-OHT E7.5 (n = 4), 4-OHT E8.5 (n = 3) were analyzed individually in 2 independent experiments.

(B) Quantification of flow cytometric analysis of labeled LMPs and type I pre-HSCs, as shown in (A), in E10.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* AGM (left) and YS+VU (right), labeled with 4-OHT at E7.5 or 4-OHT E8.5. Error bars represent mean ± SD. * $p < 0.05$,** $p < 0.01$, *** $p < 0.001$ (two-way ANOVA followed by Tukey's multiple comparisons test).

(C) Flow cytometric analysis of labeled LMPs (Lin- CD45+ Kit+ Flt3+ IL7Rα+) in E11.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* fetal liver (FL). 4-OHT E7.5 (n = 4), and 4-OHT E8.5 (n = 8) FL were analyzed individually in 2 independent experiments.

(D) Quantification of flow cytometric analysis of LMPs shown in **(C)**. Error bars represent mean ± SD. *$p<0.05$ (two-tailed unpaired Student's *t*-test).

(E) Quantification of flow cytometric analysis of labeled hematopoietic progenitor cells (CD45+ Kit+) shown in **(C)**. Error bars represent mean ± SD. ** $p < 0.01$, (two-tailed unpaired Student's *t*-test).

(F) Quantification of flow cytometric analysis of labeled (zsGreen+) progenitors, type I and type II pre-HSCs in E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* VU and YS when dissected separately (gating strategy shown in **Figure 1B**. Error bars represent mean ± standard deviation (SD). ns = not significant (two-way ANOVA followed by Tukey's multiple comparisons test).

(G) Comparison of flow cytometric analysis of E14.5 FL HSCs as shown in **Figure 1E**, shown separately for *Cdh5-CreER^{T2}::R26^{zsGreen} and Cdh5-CreER^{T2}::R26^{tdTomato}* embryos. Error bars represent mean ± SD.

(H-M) Quantification of flow cytometric analysis of labeled LSK (Lineage- Kit+ Sca1+) and LK hematopoietic progenitor cells (Lineage- Kit+ Sca1-) related to **Figure 1E** in *Cdh5-CreER^{T2}::R26^{tdTomato}* or *Cdh5-CreER^{T2}::R26^{zsGreen}* E14.5 FL **(H** for LSK**, K** for LK**)**, E16.5 FL **(I** for LSK**, L** for LK**)** and adult BM (2-months-old) **(J** for LSK**, M** for LK**)** activated with 4-OHT at E7.5, E8.5 or E10.5. (E14.5 4-OHT at E7.5 (n = 14), E14.5 4-OHT at E8.5 (n = 26), E14.5 4-OHT at E10.5 (n = 16), E16.5 4-OHT E7.5 (n = 13), E16.5 4-OHT E8.5 (n = 18), E16.5 4-OHT E10.5 (n = 17), 2 months 4-OHT E7.5 (n = 5), 2 months 4-OHT E8.5 (n = 13), 2 months 4-OHT E10.5 (n = 13) were analyzed individually in 17 independent experiments. Error bars represent mean ± SD. ** $p < 0.01$, **** $p < 0.0001$, (one-way ANOVA followed by Tukey's multiple comparisons test).

Figure S3

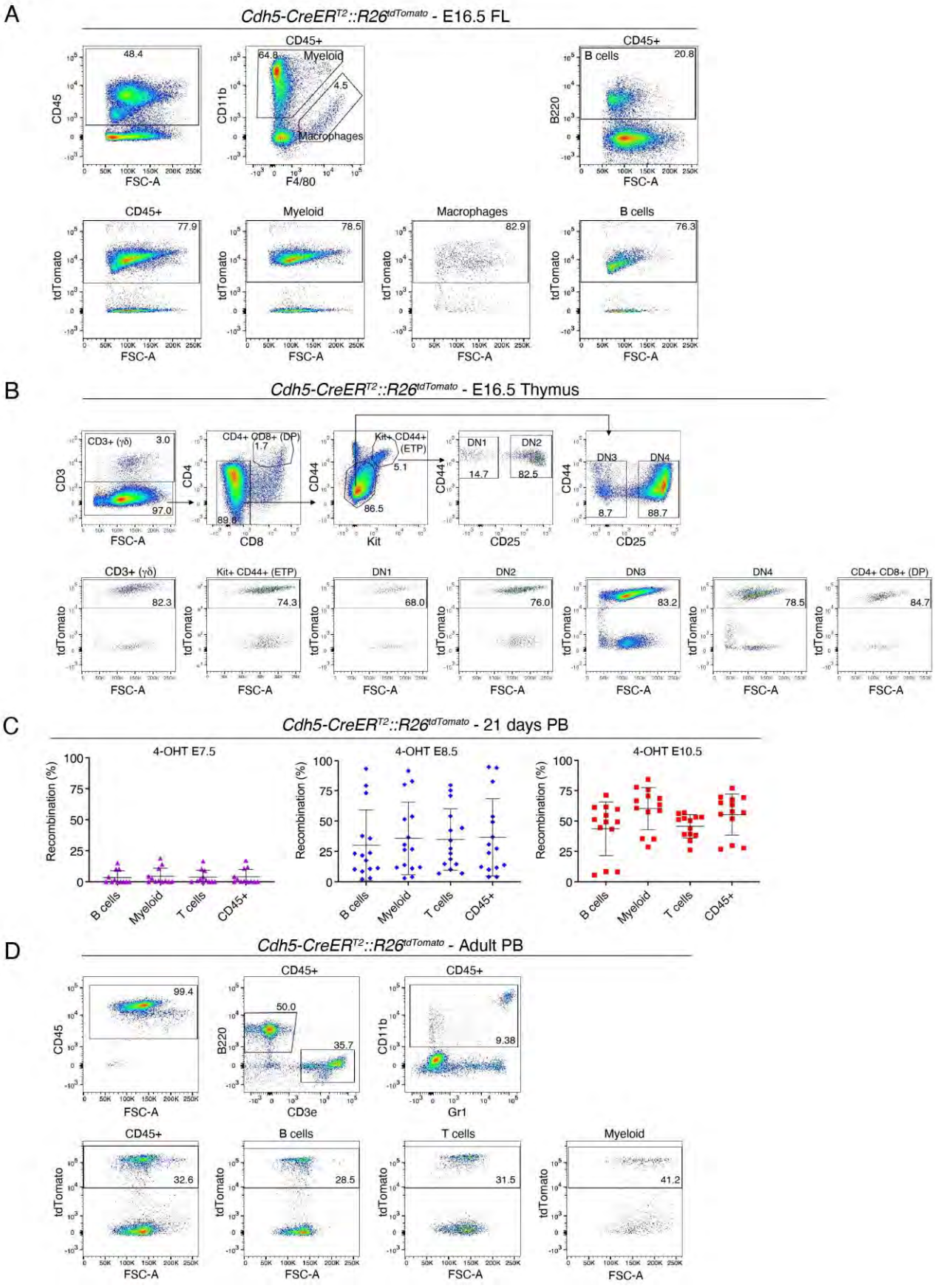

A
Cdh5-CreER^T2::R26^tdTomato - E16.5 FL
CD45+
Myeloid
Macrophages
B cells
CD45
CD11b
B220
F4/80
FSC-A
tdTomato
B
Cdh5-CreER^T2::R26^tdTomato - E16.5 Thymus
CD3+ (γδ)
CD4+ CD8+ (DP)
Kit+ CD44+ (ETP)
DN1
DN2
DN3
DN4
CD3
CD4
CD8
CD44
Kit
CD25
C
Cdh5-CreER^T2::R26^tdTomato - 21 days PB
4-OHT E7.5
4-OHT E8.5
4-OHT E10.5
Recombination (%)
B cells
Myeloid
T cells
CD45+
D
Cdh5-CreER^T2::R26^tdTomato - Adult PB
CD3e
Gr1

Figure S3 (Related to Figure 2). Flow cytometric analysis of fetal and postnatal lympho-myeloid contribution in *Cdh5-CreER^{T2}* mice.

(A) Representative flow cytometric analysis of CD45+ (leukocytes), myeloid cells (CD45+ CD11b+), macrophages (CD45+ F4/80^{hi}), B cells (CD45+ B220+) in E16.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* FL, labeled with 4-OHT at E7.5, E8.5 (shown here) or E10.5. 4-OHT E7.5 (n = 21), 4-OHT E8.5 (n = 23), 4-OHT E10.5 (n= 8) FL were analyzed individually in 5 independent experiments. The corresponding quantification is shown in **Figure 2A**.

(B) Representative flow cytometric analysis of thymocytes in E16.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* or *Cdh5-CreER^{T2}::R26^{zsGreen}* fetal thymus, labeled with 4-OHT at E7.5, E8.5 (shown here) or E10.5. 4-OHT E7.5 (n = 13), 4-OHT E8.5 (n = 18), 4-OHT E10.5 (n = 16) thymuses were individually analyzed in 4 independent experiments. Quantification shown in **Figure 2B**.

(C) Quantification of flow cytometric analysis of B cells (CD45+ B220+), myeloid cells (CD45+ CD11b+), T cells (CD45+ CD3e+), CD45+ (leukocytes) in juvenile 21 days old *Cdh5-CreER^{T2}::R26^{tdTomato}* PB. 4-OHT E7.5 (n = 12), 4-OHT E8.5 (n = 15), and 4-OHT E10.5 (n = 13) mice were analyzed individually in 7 independent experiments. Error bars represent mean ± SD.

(D) Representative flow cytometric analysis of postnatal CD45+ (leukocytes), B cells (CD45+ B220+), T cells (CD45+ CD3e+), myeloid cells (CD45+ CD11b+) in *Cdh5-CreER^{T2}::R26^{tdTomato}* PB labeled with 4-OHT at E7.5, E8.5 (2-months-old adult mice; shown here) or E10.5. 4-OHT E7.5 (n = 12), 4-OHT E8.5 (n = 13), and 4-OHT E10.5 (n = 13) mice were analyzed individually in 7 independent experiments. The corresponding quantification is shown in **Figure 2C**.

Figure S4

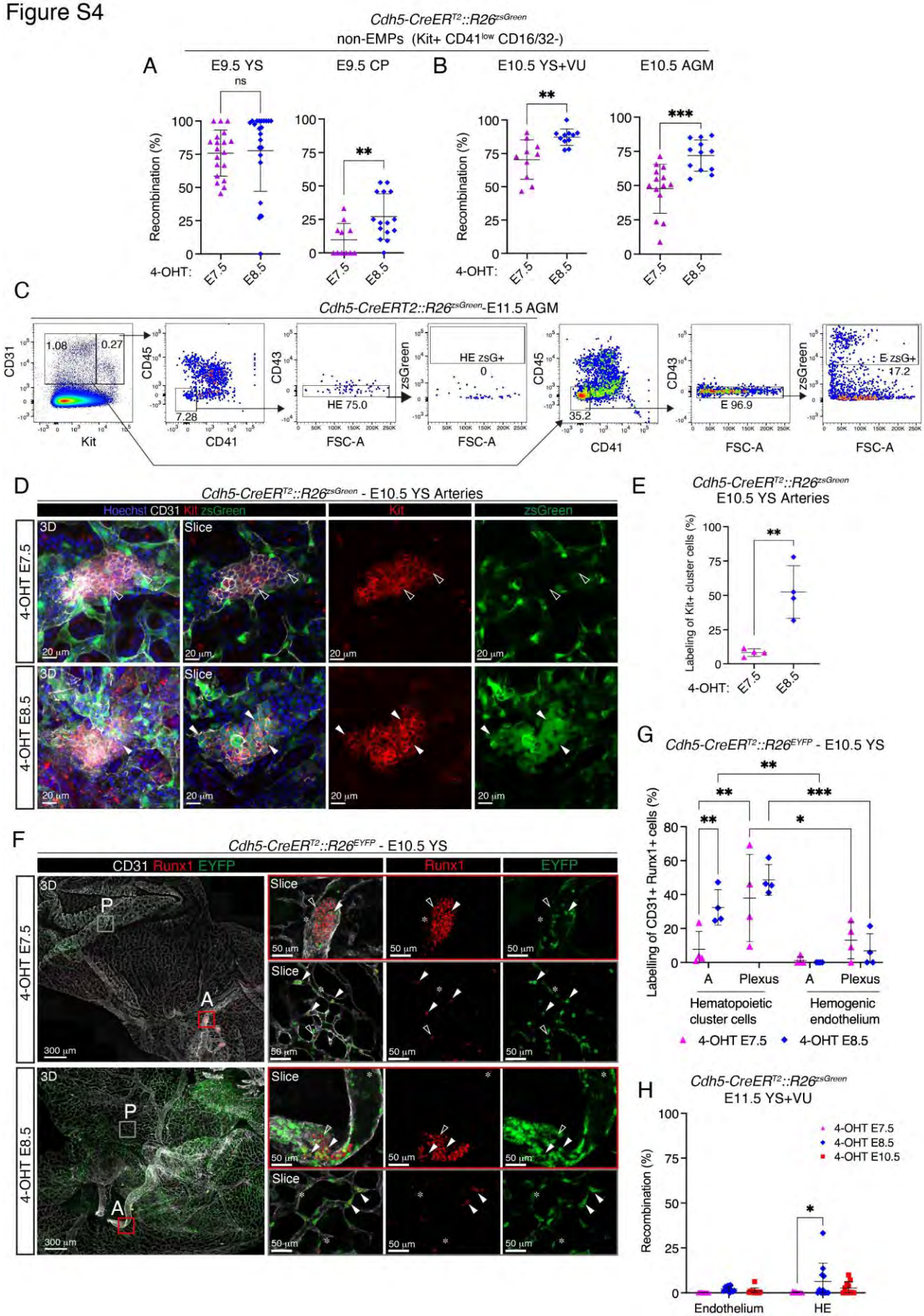

Cdh5-CreER^T2::R26^zsGreen
non-EMPs (Kit+ CD41^low CD16/32-)
A
E9.5 YS
E9.5 CP
B
E10.5 YS+VU
E10.5 AGM
Recombination (%)
4-OHT:
E7.5
E8.5
ns
C
Cdh5-CreERT2::R26^zsGreen-E11.5 AGM
CD31
Kit
CD45
CD41
CD43
FSC-A
zsGreen
HE zsG+
HE 75.0
D
Cdh5-CreER^T2::R26^zsGreen - E10.5 YS Arteries
Hoechst CD31 Kit zsGreen
4-OHT E7.5
4-OHT E8.5
3D
Slice
E
Cdh5-CreER^T2::R26^zsGreen
E10.5 YS Arteries
Labeling of Kit+ cluster cells (%)
F
Cdh5-CreER^T2::R26^EYFP - E10.5 YS
CD31 Runx1 EYFP
Runx1
EYFP
P
A
G
Cdh5-CreER^T2::R26^EYFP - E10.5 YS
Labelling of CD31+ Runx1+ cells (%)
Hematopoietic cluster cells
Hemogenic endothelium
4-OHT E7.5
4-OHT E8.5
H
Cdh5-CreER^T2::R26^zsGreen
E11.5 YS+VU
4-OHT E10.5
Endothelium
HE

Figure S4 (Related to Figure 3). Flow cytometric analysis and whole-mount confocal imaging of non-EMPs, endothelial cells and hematopoietic clusters in *Cdh5-CreER^{T2}* embryos and YS.

(A) Quantification of flow cytometric analysis of non-EMP (Ter119- Kit+ CD41^{low} CD16/32-) cells in *Cdh5-CreER^{T2}::R26^{zsGreen}* E9.5 YS (left) and caudal part (CP, right), labeled with 4-OHT at E7.5 or E8.5. Gates as shown in **Figure S1B**. 4-OHT E7.5 (n = 19), and 4-OHT E8.5 (n = 23) YS analyzed in 4 independent experiments. 4-OHT E7.5 (n = 11), and 4-OHT E8.5 (n = 15) CP analyzed in 3 independent experiments. Error bars represent mean ± SD. ns = non-significant; ** $p < 0.01$, (two-tailed unpaired Student's *t*-test).

(B) Quantification of flow cytometric analysis of non-EMP (Ter119- Kit+ CD41^{low} CD16/32-) cells in *Cdh5-CreER^{T2}::R26^{zsGreen}* E10.5 YS including vitelline and umbilical arteries (VU), left and AGM, right, labeled with 4-OHT E7.5 or E8.5. Gates as shown in **Figure S1B**. 4-OHT E7.5 (n = 10), and 4-OHT E8.5 (n = 12) YS analyzed in 4 independent experiments. 4-OHT E7.5 (n = 14), and 4-OHT E8.5 (n = 12) AGM analyzed in 4 independent experiments. Error bars represent mean ± SD. ** $p < 0.01$; *** $p < 0.001$ (two-tailed unpaired Student's *t*-test).

(C) Representative flow cytometric analysis, related to **Figure 3G**, of AGM region hemogenic endothelium (HE, Ter119- Kit+ CD31+ CD41- CD43-) and endothelium (Ter119- Kit- CD31+ CD45- CD41- CD43) in E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos, activated with 4-OHT at E10.5.

(D) Confocal WM-IF analysis of E10.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* YS large arteries. Left panels show 3D maximum intensity projections. Middle and right panels show single 2.5 μm-thick optical slices. Arrowheads indicate large Kit+ hematopoietic clusters, unlabeled with 4-OHT at E7.5 (empty arrowheads; top), or labeled at E8.5 (white arrowheads; bottom). 4-OHT E7.5 (n=4), and 4-OHT E8.5 (n=4) different YS were analyzed in 2 independent experiments. Scale bars: 20 μm.

(E) Labeling quantification of Kit+ cluster cells in the YS large arteries as displayed in **(D)**. Measurements were performed on images from 4-OHT E7.5 (n=4), and 4-OHT E8.5 (n=4) different YSs (2-4 images / YS); 12 (4-OHT E7.5), 12 (4-OHT E8.5) different images used. Error bars represent mean ± SD. **:$p<0.01$ (two-tailed unpaired Student's *t*-test).

(F) Confocal WM-IF analysis of E10.5 (32-36sp) *Cdh5-CreER^{T2}::R26^{EYFP} YS*. The left panels show maximum intensity 3D projections. Boxed area in the merged image is magnified in the right panels and shows a single 2.5 μm-thick optical slice. Color-code indicates magnified arterial (red box) and vascular plexus (grey box) regions. Arrowheads indicate Runx1+ CD31+ round-shaped hematopoietic cluster cells, labeled (white arrowheads) and unlabeled (empty arrowheads). Asterisks indicate Runx1+ CD31+ flat-shaped hemogenic endothelium cells. 4-OHT E7.5 (n = 4), and 4-OHT E8.5 (n = 4) different embryos were analyzed Scale bars: 300 μm (3D), 50 μm (slice). A: artery; P: plexus.

(G) Labeling quantification of Runx1+ CD31+ round-shaped hematopoietic cluster cells and Runx1+ CD31+ flat-shaped hemogenic endothelium cells in E10.5 *Cdh5-CreER^{T2}::R26^{EYFP}* YS as shown in **(F)**. Cells located in the arteries (A) cells were quantified separately from cells in the YS vascular plexus. Measurements were performed on images from 4-OHT E7.5 (n = 4), and 4-OHT E8.5 (n =4) different embryos (3-6 images / YS); Error bars represent mean ± SD. *:p<0.05, **:p<0.01, ***:p<0.001 (two-way ANOVA followed by Tukey's multiple comparisons test).

(H) Quantification of flow cytometric analysis of labeled hemogenic endothelium (HE) (Ter119- CD31+ Kit+ CD45- CD41- CD43-) and endothelium (Ter119- CD31+ Kit- CD45- CD41- CD43-) in YS+VU of E11.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* embryos (gates as in **(C)**). 4-OHT at E7.5 (n = 15), 4-OHT at E8.5 (n = 11), 4-OHT at E10.5 (n = 12) AGM were analyzed individually in 5 independent experiments. Error bars represent mean ± SD. *:p<0.05 (two-way ANOVA followed by Tukey's multiple comparisons test)

Figure S5

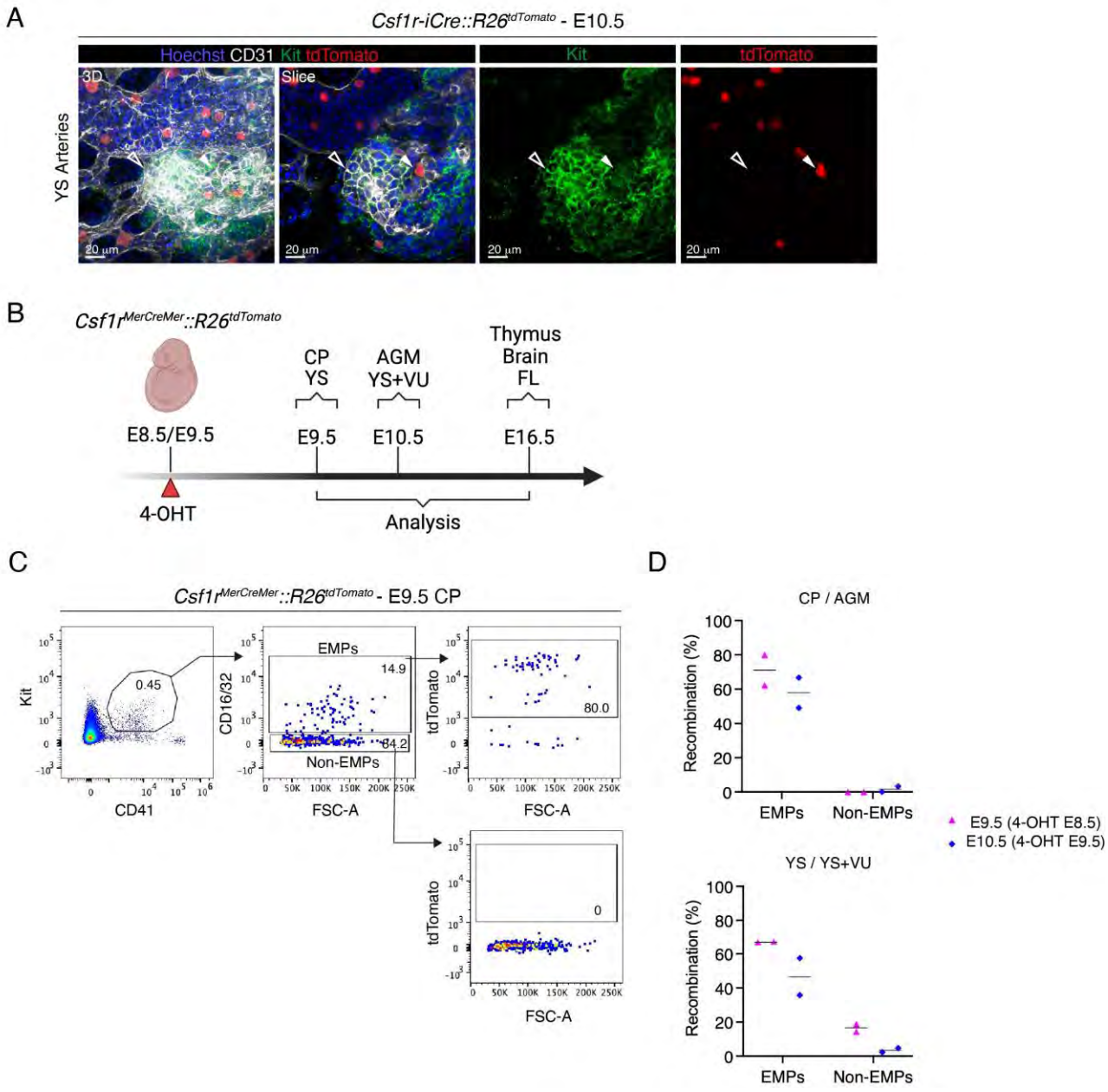


Figure S5 (Related to Figure 4). Csf1r lineage tracing yields highly specific labeling of immunophenotypically defined EMPs.

(A) Confocal WM-IF analysis of E10.5 *Csf1r-iCre::R26*^{tdTomato} YS large arteries. Left panel shows a 3D maximum intensity projection; other images are single 2.5 µm-thick optical slices. Arrowheads indicate Kit+ hematopoietic clusters. Empty arrowheads indicate unlabeled (tdTomato-) cells, white arrowheads indicate labeled (tdTomato+) ones. Quantification is shown in **Figure 4D**. A total number of 5 different embryos were analyzed in 3 independent experiments.

(B) Visual schematic of lineage tracing experiments in *Csf1r*^{MerCreMer}*::R26*^{tdTomato} E9.5-E10.5-E16.5 embryos, related to **Figure 4F** and panels **C**-**D**.

(C) Representative flow cytometric analysis of labeled EMPs (Ter119- Kit+ CD41^{lo} CD16/32+) and non-EMPs (Ter119-, Kit+ CD41^{lo} CD16/32-) in *Csf1r*^{MerCreMer}*::R26*^{tdTomato} E9.5 embryos, activated with 4-OHT at E8.5. E9.5 embryos 4-OHT E8.5 (n=2), and E10.5 embryos 4-OHT E8.5 (n=2) were analyzed in 2 independent experiments.

(D) Quantification of flow cytometric analysis of labeled (tdTomato+) EMPs and non-EMPs in E9.5 *Csf1r*^{MerCreMer}*::R26*^{tdTomato} CP (top) and YS (bottom) and E10.5 *Csf1r*^{MerCreMer}*::R26*^{tdTomato} AGM (top) and YS+VU (bottom).

Figure S6

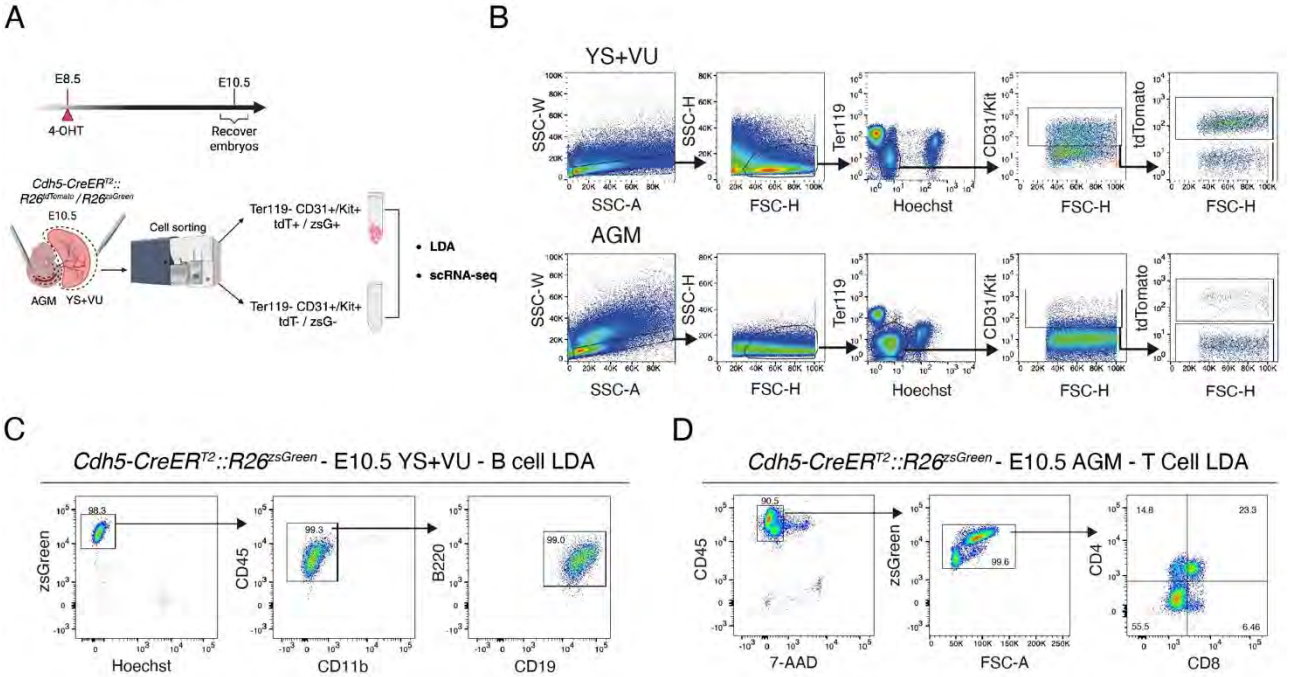


Figure S6 (Related to Figure 5). Hemato-endothelial cell sorting strategy and *ex vivo* assessment of the lympho-myeloid potential of fetal-restricted HSPCs through limiting dilution co-culture assays.

(A) Experimental schematic showing the cell sorting strategy used for scRNA-Seq and for Limiting Dilution Assays (LDA) on AGM and YS+VU from E10.5 *Cdh5-CreER^{T2}::R26$^{tdTomato/zsGreen}$* embryos, with 4-OHT activation at E8.5. Hemato-endothelial (Ter119- CD31+ and/or Kit+) cells were selected and tdTomato/zsGreen+ or tdTomato/zsGreen- fractions were separately sorted from either AGM or YS+VU.

(B) Representative flow cytometric gating strategy for the isolation of live Ter119- CD31+and/or Kit+ tdTomato+ and tdTomato- for scRNA-Seq or LDA as represented in (A).

(C) Representative flow cytometric analysis of B cell LDA OP9 co-cultures from E10.5 Cdh5-*CreER^{T2}::R26^{zsGreen}* YS+VU with 4-OHT at E8.5. 3 independent experiments were performed. The corresponding quantification is shown in **Figure 5D**.

(D) Representative flow cytometric analysis of T cell LDA OP9-Dl1 co-cultures from E10.5 Cdh5-*CreER^{T2}::R26^{zsGreen}* AGM with 4-OHT at E8.5. 2 independent experiments were performed. The corresponding quantification is shown in **Figure 5E**.

Figure S7

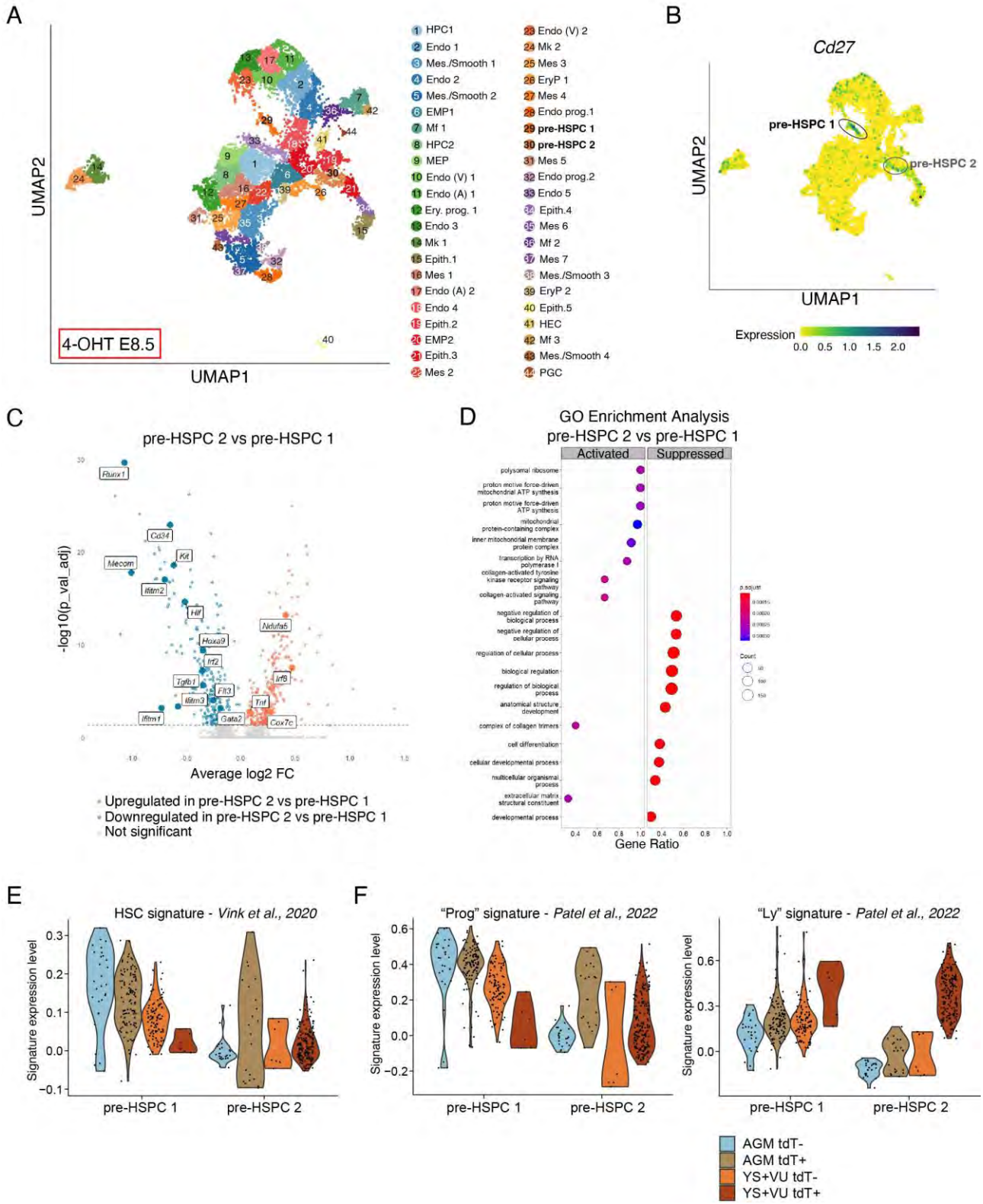

A
UMAP2
UMAP1
4-OHT E8.5
1 HPC1
2 Endo 1
3 Mes./Smooth 1
4 Endo 2
5 Mes./Smooth 2
6 EMP1
7 Mf 1
8 HPC2
9 MEP
10 Endo (V) 1
11 Endo (A) 1
12 Ery. prog. 1
13 Endo 3
14 Mk 1
15 Epith.1
16 Mes 1
17 Endo (A) 2
18 Endo 4
19 Epith.2
20 EMP2
21 Epith.3
22 Mes 2
23 Endo (V) 2
24 Mk 2
25 Mes 3
26 EryP 1
27 Mes 4
28 Endo prog.1
29 pre-HSPC 1
30 pre-HSPC 2
31 Mes 5
32 Endo prog.2
33 Endo 5
34 Epith.4
35 Mes 6
36 Mf 2
37 Mes 7
38 Mes./Smooth 3
39 EryP 2
40 Epith.5
41 HEC
42 Mf 3
43 Mes./Smooth 4
44 PGC
B
Cd27
pre-HSPC 1
pre-HSPC 2
UMAP2
UMAP1
Expression
0.0 0.5 1.0 1.5 2.0
C
pre-HSPC 2 vs pre-HSPC 1
-log10(p_val_adj)
Average log2 FC
Runx1
Cd34
Kit
Mecom
Ifitm2
Hlf
Hoxa9
Irf2
Tgfb1
Flt3
Ifitm3
Ifitm1
Gata2
Tnf
Cox7c
Ndufa5
Irf8
Upregulated in pre-HSPC 2 vs pre-HSPC 1
Downregulated in pre-HSPC 2 vs pre-HSPC 1
Not significant
D
GO Enrichment Analysis
pre-HSPC 2 vs pre-HSPC 1
Activated
Suppressed
polysomal ribosome
proton motive force-driven mitochondrial ATP synthesis
proton motive force-driven ATP synthesis
mitochondrial protein-containing complex
inner mitochondrial membrane protein complex
transcription by RNA polymerase I
collagen-activated tyrosine kinase receptor signaling pathway
collagen-activated signaling pathway
negative regulation of biological process
negative regulation of cellular process
regulation of cellular process
biological regulation
regulation of biological process
anatomical structure development
complex of collagen trimers
cell differentiation
cellular developmental process
multicellular organismal process
extracellular matrix structural constituent
developmental process
Gene Ratio
p.adjust
Count
E
HSC signature - Vink et al., 2020
Signature expression level
pre-HSPC 1
pre-HSPC 2
F
"Prog" signature - Patel et al., 2022
"Ly" signature - Patel et al., 2022
AGM tdT-
AGM tdT+
YS+VU tdT-
YS+VU tdT+

Figure S7 (Related to Figure 6). scRNA-Seq of E10.5 AGM and YS+VU hemato-endothelial cells isolated from *Cdh5-CreER^{T2}::R26^{tdTomato}* embryos (4-OHT at E8.5) allows the identification and comparison of subsets of pre-HSPCs with distinct origins.

(A) UMAP plot showing complete clustering annotation of the scRNA-Seq dataset (cell isolation strategy as shown in **Figure S6D-E**).

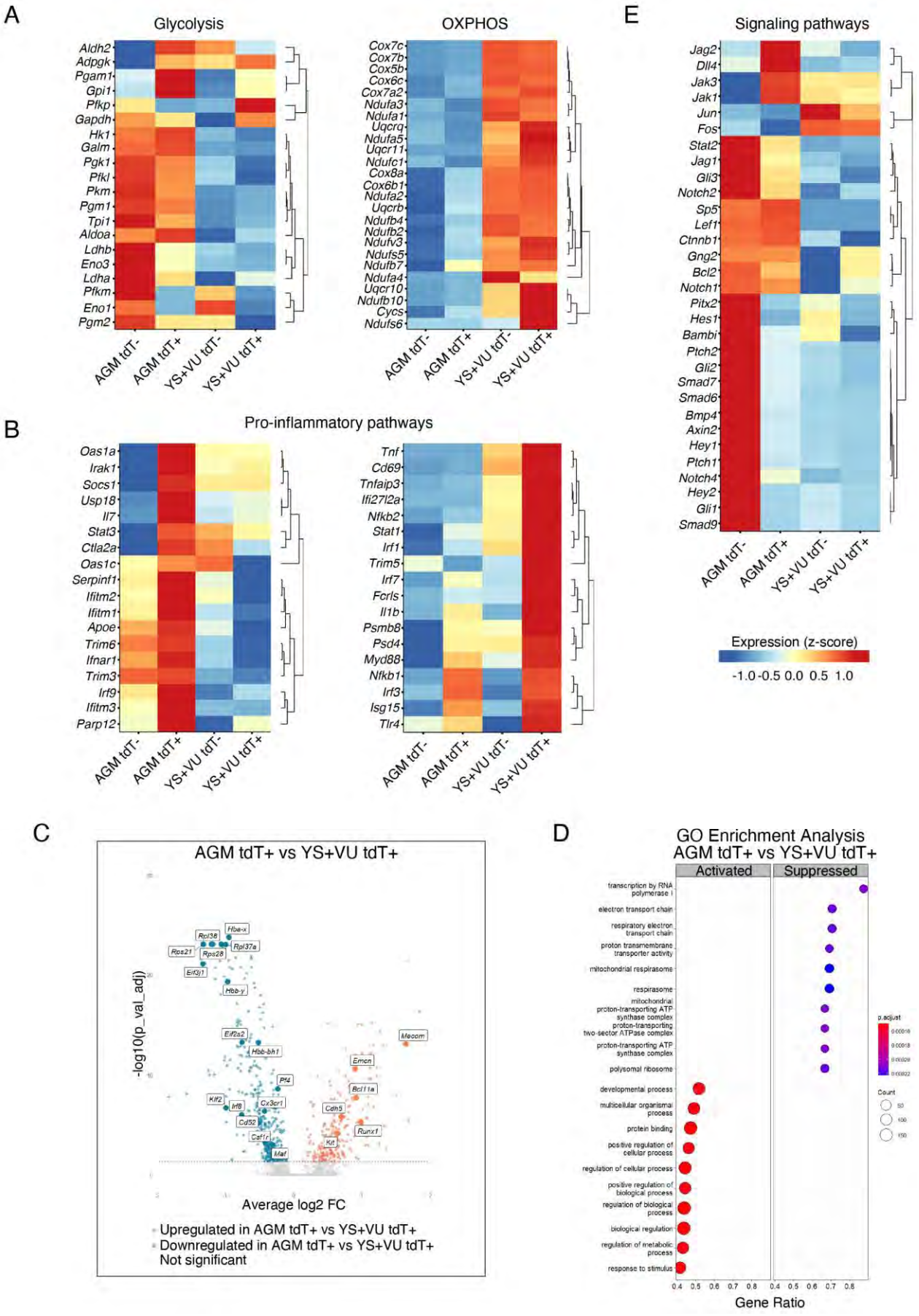
(B) UMAP plot showing *Cd27* expression. Pre-HSPC clusters 1 and 2 are circled.

(C) Volcano plot (Average log2 FC versus negative log of adjusted P value) used to visualize statistically significant gene expression changes (adjusted P value <0.05) between pre-HSPC 2 and pre-HSPC 1. Upregulated genes are labeled in red and downregulated genes labels in blue. Highlighted in labels are known genes related with hematopoiesis, inflammation and metabolic processes. A total number of 587 genes were differentially expressed (complete list in **Table S2**).

(D) GO biological process gene set enrichment analysis (GSEA) performed on a ranked list of differentially expressed genes between pre-HSPC 2 and pre-HSPC 1. Dot size indicates the gene count and color intensity represents enrichment score (adjusted P value).

(E-F) Gene expression of published signatures for (E) HSC (Vink et al., Cell Reports 2020), (F) AGM-derived progenitors ("Prog") and lymphoid progenitors ("Ly") (Patel et al., Nature 2022) in pre-HSPC 1 and pre-HSPC 2 clusters, considering the tissue origin (AGM tdT-, AGM tdT+, YS+VU tdT-, YS+VU tdT+). The average expression of top50 genes for each signature was calculated on single cells using the AddModuleScore function in Seurat.

Figure S8



211

Figure S8 (Related to Figure 6). Pre-HSPCs of distinct origin show different expression levels of genes involved in metabolism, inflammation and hematopoiesis.

(A) Heatmap showing the relative expression levels of selected genes involved in the metabolic pathways of glycolysis (left) and oxidative phosphorylation (OXPHOS, right) among AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ pre-HSPCs.

(B) Heatmap showing the relative expression levels of selected genes involved in pro-inflammatory pathways among AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ pre-HSPCs.

(C) Volcano plot (Average log2 FC versus negative log of adjusted P value) used to visualize statistically significant gene expression changes (adjusted P value <0.05) between AGM tdTomato+ and YS+VU tdTomato+ pre-HSPCs. Upregulated genes are labeled in red and downregulated genes labels in blue. Highlighted in labels are known hematopoietic and ribosomal genes. A total number of 698 genes were differentially expressed (complete list in **Table S3**).

(D) GO biological process gene set enrichment analysis (GSEA) performed on a ranked list of differentially expressed genes between AGM tdTomato+ and YS+VU tdTomato+ pre-HSPCs. Dot size indicates the gene count and color intensity represents enrichment score (adjusted P value).

(E) Heatmap showing the relative expression levels of selected genes involved in developmental and hematopoietic processes (e.g. Notch, BMP, Shh signaling pathways) among AGM tdTomato-, AGM tdTomato+, YS+VU tdTomato- and YS+VU tdTomato+ pre-HSPCs.

Figure S9

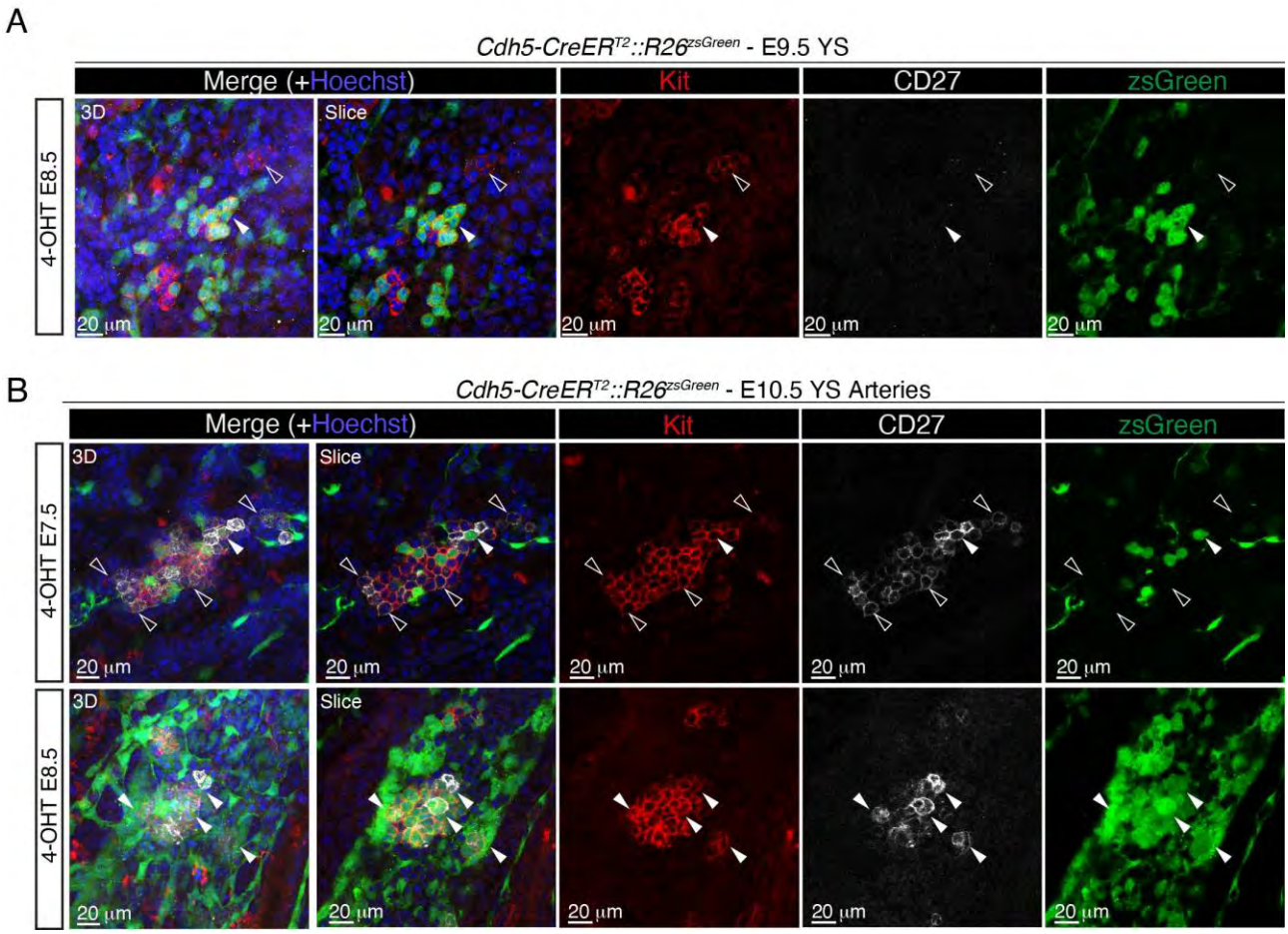


Figure S9 (Related to Figure 7). Whole-mount confocal imaging analysis of Kit+ CD27- and Kit+ CD27+ clusters in the YS.

(A) Confocal WM-IF analysis of E9.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* YS (4-OHT E8.5) showing lack of CD27 expression within Kit+ clusters. Left panel shows maximum intensity 3D projection. Middle and right panels show single 2.5 µm-thick slices. Arrowheads indicate Kit+ CD27- clusters, labeled (zsGreen+; white arrowhead) or unlabeled (zsGreen-; empty arrowhead). A total number of 3 different YS were analyzed in 2 independent experiments. Scale bar: 20 µm.

(B) Confocal WM-IF analysis of E10.5 *Cdh5-CreER^{T2}::R26^{zsGreen}* YS large arteries. Images show single 2.5 µm-thick optical slices. Arrowheads indicate examples of labeled (zsGreen+; white arrowheads) or unlabeled (zsGreen-; empty arrowheads) Kit+ CD27+ cells. 4-OHT E7.5 (n = 4) and 4-OHT E8.5 (n = 4) different YS were analyzed in 3 independent experiments. Scale bars: 20 µm

Figure S10

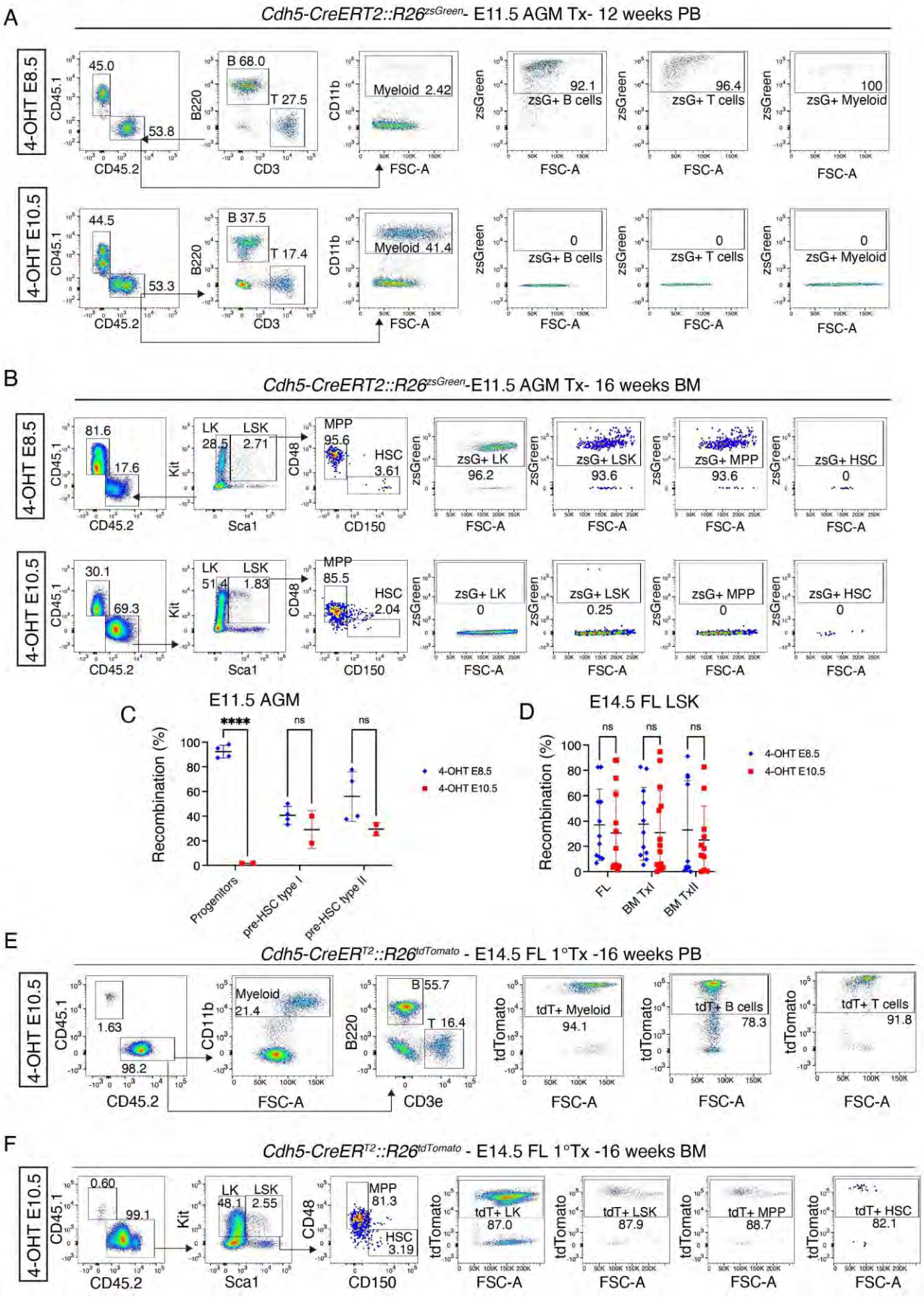

A
Cdh5-CreERT2::R26zsGreen- E11.5 AGM Tx- 12 weeks PB
4-OHT E8.5
4-OHT E10.5
CD45.1
CD45.2
B220
CD3
CD11b
FSC-A
zsGreen
45.0
53.8
B 68.0
T 27.5
Myeloid 2.42
92.1
zsG+ B cells
96.4
zsG+ T cells
100
zsG+ Myeloid
44.5
53.3
B 37.5
T 17.4
Myeloid 41.4
0
zsG+ B cells
0
zsG+ T cells
0
zsG+ Myeloid
B
Cdh5-CreERT2::R26zsGreen-E11.5 AGM Tx- 16 weeks BM
Kit
Sca1
CD48
CD150
LK
LSK
MPP
HSC
81.6
17.6
28.5
2.71
95.6
3.61
zsG+ LK
96.2
zsG+ LSK
93.6
zsG+ MPP
93.6
zsG+ HSC
0
30.1
69.3
51.4
1.83
85.5
2.04
zsG+ LK
0
zsG+ LSK
0.25
zsG+ MPP
0
zsG+ HSC
0
C
E11.5 AGM
Recombination (%)
****
ns
ns
Progenitors
pre-HSC type I
pre-HSC type II
4-OHT E8.5
4-OHT E10.5
D
E14.5 FL LSK
Recombination (%)
ns
ns
ns
FL
BM TxI
BM TxII
4-OHT E8.5
4-OHT E10.5
E
Cdh5-CreERT2::R26tdTomato - E14.5 FL 1°Tx -16 weeks PB
4-OHT E10.5
CD45.1
CD45.2
CD11b
B220
CD3e
FSC-A
tdTomato
1.63
98.2
Myeloid
21.4
B 55.7
T 16.4
tdT+ Myeloid
94.1
tdT+ B cells
78.3
tdT+ T cells
91.8
F
Cdh5-CreERT2::R26tdTomato - E14.5 FL 1°Tx -16 weeks BM
4-OHT E10.5
CD45.1
CD45.2
Kit
Sca1
CD48
CD150
FSC-A
tdTomato
0.60
99.1
LK
LSK
48.1
2.55
MPP
81.3
HSC
3.19
tdT+ LK
87.0
tdT+ LSK
87.9
tdT+ MPP
88.7
tdT+ HSC
82.1

Figure S10 (Related to Figure 8). Flow cytometric analysis of E11.5 AGM and E14.5 FL transplants.

(A) Representative flow cytometric analysis of PB from adult C57 BL/6 CD45.1 mice transplanted with *Cdh5-CreER^{T2}::R26^{zsGreen}* E11.5 AGM, activated with 4-OHT at E8.5 (top) or 4-OHT at E10.5 (bottom). FACS plots show hematopoietic populations and labeling percentages of donor myeloid cells, B cells and T cells. The corresponding quantification is shown in **Figure 8B-C**.

(B) Representative flow cytometric analysis of BM from adult C57 BL/6 CD45.1 mice transplanted with *Cdh5-CreER^{T2}::R26^{zsGreen}* E11.5 AGM, activated with 4-OHT at E8.5 (top) or 4-OHT at E10.5 (bottom). FACS plots show hematopoietic progenitor populations and labeling percentages of donor LKs, LSK, MPPs and HSCs. The corresponding quantification is shown in **Figure 8D**.

(C) Quantification of flow cytometric analysis of immunophenotypic pre-HSC labeling in *Cdh5-CreER^{T2}::R26^{zsGreen} and Cdh5-CreER^{T2}::R26^{tdTomato}* E11.5 transplanted AGM (related to **Figure 8A-D**), labeled with 4-OHT at E8.5 or E10.5. Number of replicates as in **Figure 8C**. Error bars represent mean ± SD. **** $p < 0.0001$, ns = non-significant (two-way ANOVA followed by Tukey's multiple comparisons test).

(D) Quantification of flow cytometric analysis of LSK cells labeling in *Cdh5-CreER^{T2}::R26^{EYFP} and Cdh5-CreER^{T2}::R26^{tdTomato}* E14.5 transplanted FL and bone marrow (BM) of primary and secondary transplanted mice labeled with 4-OHT at E8.5 or E10.5. Number of replicates as in **Figure 8E**. Error bars represent mean ± SD. ns = non-significant (two-way ANOVA followed by Tukey's multiple comparisons test).

(E) Representative flow cytometric analysis of PB from adult C57 BL/6 CD45.1 mice transplanted with *Cdh5-CreER^{T2}::R26^{tdTomato}* E14.5 FL, activated with 4-OHT at E8.5 or 4-OHT at E10.5 (shown here). FACS plots show hematopoietic populations and labeling percentages of donor myeloid cells, B cells and T cells. The corresponding quantification is shown in **Figure 8E**.

(F) Representative flow cytometric analysis of BM from adult C57 BL/6 CD45.1 mice transplanted with *Cdh5-CreER^{T2}::R26^{tdTomato}* E14.5 FL, activated with 4-OHT at E8.5 or 4-OHT at E10.5 (shown here). FACS plots show hematopoietic progenitor populations and labeling percentages of donor LKs, LSK, MPPs and HSCs. The corresponding quantification is shown in **Figure 8F**.

278

279 **SUPPLEMENTAL TABLES**

REAGENT or RESOURCE	SOURCE	IDENTIFIER	Experiment/Use
FACS Antibodies			
Rat monoclonal anti-Ter119 APC-fire750 (clone TER-119)	BioLegend	Cat#116250; RRID: AB_2819833	AGM, YS-VU analysis (for erythroid cells exclusion)
Rat monoclonal anti-CD117 (c-Kit) FITC (clone 2B8)	eBioscience	Ref: 11-1171-85; RRID: AB_465187	FL and BM LSK analysis
Rat monoclonal anti-CD41 PE-Cy7 (clone eBioMWReg30)	eBioscience	Ref: 25-0411-82; RRID: AB_1234970	AGM/YS EMPs, pre-HSCs and LMPs analysis
Rat monoclonal anti-CD16/32 APC (clone 93)	BioLegend	Cat#101326; RRID: AB_1953273	AGM/YS EMPs analysis
Rat monoclonal anti-CD150 PE-Cy7 (SLAM) (clone TC15-12F12.2)	BioLegend	Cat#115914; RRID: AB_439797	FL and BM LSK analysis
Armenian hamster monoclonal anti-CD48 APC (clone HM48-1)	BioLegend	Cat#103412; RRID: AB_571997	FL and BM LSK analysis
Rat monoclonal anti-Ly6A/E (Sca-1) Pacific Blue (clone E13-161.7)	BioLegend	Cat#122520; RRID: AB_2143237	FL and BM LSK analysis
Rat monoclonal anti-CD117 (c-kit) APC-Cy7 (clone 2B8)	BioLegend	Cat#105826; RRID: AB_1626278	FL and BM LSK analysis, fetal thymus analysis and bulk OP9-DL1 co-culture readout
Rat monoclonal anti-CD45 APC-eFluor 780 (clone 30-F11)	eBioscience	Cat#47-0451-82; RRID: AB_1548781	FL, BM an PB hematopoietic cells detection
Rat monoclonal anti-CD117 (c-kit) PE-Cy7 (clone 2B8)	BioLegend	Cat#105814; RRID: AB_313223	FL and BM LSK analysis
Rat monoclonal anti-CD135 Brilliant Violet 421 (clone A2F10)	BioLegend	Cat#135313; RRID: AB_2562338	E11.5 FL LMPPs detection
Rat monoclonal anti-CD127 (IL-7Ra) PE (clone A7R34)	BioLegend	Cat#135009; RRID: AB_1937252	E11.5 FL LMPPs detection
Rat monoclonal anti-Ter119 PE-Cy5 (clone TER-119)	BioLegend	Cat#116210; RRID: AB_313711	FL and BM LSK analysis (lineage cocktail), FACS sorting for scRNAseq, LDA and CFU-C
Armenian hamster monoclonal anti-CD3e PE-Cy5 (clone 145-2C11)	BioLegend	Cat#100310; RRID: AB_312675	FL and BM LSK analysis (lineage cocktail), adult BM an PB T cells detection
Rat monoclonal anti-F4/80 PE-Cy5 (clone BM8)	BioLegend	Cat#123112; RRID: AB_893482	FL and BM LSK analysis (lineage cocktail)
Mouse monoclonal anti-NK1.1 PE-Cy5 (clone PK136)	BioLegend	Cat#108716; RRID: AB_493590	FL and BM LSK analysis (lineage cocktail)
Rat monoclonal anti-Ly6G/Ly6C (Gr1) PE-Cy5 (clone RB6-8C5)	BioLegend	Cat#108410; RRID: AB_313375	FL and BM LSK analysis (lineage cocktail)
Rat monoclonal anti-CD19 PE-Cy5 (clone 6D5)	BioLegend	Cat#115509; RRID: AB_313644	FL and BM LSK analysis (lineage cocktail), bulk and

			LDA OP9 co-culture readout
Rat monoclonal anti-CD45R/B220 PE-Cy5 (clone RA3-6B2)	BioLegend	Cat#103210; RRID: AB_312995	FL B cells detection and FL and BM LSK analysis (lineage cocktail)
Rat monoclonal anti-CD45 PE (clone 30-F11)	BioLegend	Cat#103106; RRID: AB_312971	FL, BM an PB hematopoietic cells detection
Rat monoclonal anti-CD45R/B220 APC (clone RA3-6B2)	BioLegend	Cat#103212; RRID: AB_312997	FL, BM an PB B cells detection, LDA co-culture readout
Rat monoclonal anti-CD45R/B220 APC-Cy7 (clone RA3-6B2)	BioLegend	Cat#103224; RRID: AB_313007	FL, BM an PB B cells detection, bulk OP9 co-culture readout
Rat monoclonal anti-CD11b PE-Cy7 (clone M1/70)	BioLegend	Cat#101216; RRID: AB_312799	FL, BM an PB myeloid cells detection and adult brain microglia detection
Rat monoclonal anti-Ly6G/Ly6C (Gr1) APC (clone RB6-8C5)	BioLegend	Cat#108412; RRID: AB_313377	FL, BM an PB myeloid cells detection
Mouse monoclonal anti-CD45.2 FITC (clone 104)	BioLegend	Cat#109806; RRID: AB_313443	BM and PB of transplanted mice, for chimerism evaluation.
Mouse monoclonal anti-CD45.1 PE (clone A20)	BioLegend	Cat#110708; RRID: AB_313497	BM and PB of transplanted mice, for chimerism evaluation.
Mouse monoclonal anti-CD45.1 BV786 (clone A20)	BioLegend	Cat#110743; RRID: AB_2563379	BM and PB of transplanted mice, for chimerism evaluation
Rat monoclonal anti-CD93 (AA4.1) APC (clone AA4.1)	BioLegend	Cat#136510; RRID: AB_2275868	Bulk OP9 co-culture readout
Rat monoclonal anti-F4/80 APC (clone BM8)	BioLegend	Cat#123116; RRID: AB_893481	FL macrophage detection and adult brain microglia detection
Rat monoclonal anti-CD4 PE-Cy5 (clone RM4-5)	BioLegend	Cat#100514; RRID: AB_312717	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Rat monoclonal anti-CD8a PE-Cy7 (clone 53-6.7)	BioLegend	Cat#100714; RRID: AB_312753	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Rat monoclonal anti-CD25 PE-Cy7 (clone PC61)	BioLegend	Cat#102016; RRID: AB_312865	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Rat monoclonal anti-CD44 APC (clone IM7)	BioLegend	Cat#103012; RRID: AB_312963	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Mouse monoclonal anti-CD45.2 APC-Cy7 (clone 104)	BioLegend	Cat#109824; RRID: AB_830789	AGM/YS pre-HSC and LMP analysis. BM and PB of transplanted mice,

			for chimerism evaluation.
Rat monoclonal anti-CD31 APC (clone 390)	BioLegend	Cat#102410; RRID: AB_312905	AGM/YS pre-HSC and LMP analysis, FACS sorting for scRNAseq and LDA
Rat monoclonal anti-CD117 (c-kit) BV 786 (clone 2B8)	BD Horizon	Cat#564012; RRID: AB_2732005	AGM/YS pre-HSC and LMP analysis
Rat monoclonal anti-CD117 (c-kit) APC (clone 2B8)	BioLegend	Cat#105812; RRID: AB_313221	FACS sorting for scRNAseq and LDA
Rat monoclonal anti-CD117 (c-kit) PE (clone 2B8)	BioLegend	Cat#105807; RRID: AB_313216	FL and BM LSK analysis
Rat monoclonal anti-CD201 PE (clone 1560)	BD Pharmingen	Cat#566337; AB_2739694	AGM/YS pre-HSC and LMP analysis
Rat monoclonal anti-mouse CD16/CD32 antibody (Fc Block), clone 2.4G2	BD Biosciences	Cat# 553142; RRID:AB_394657	Fc-block
Rat monoclonal Anti-Mouse CD44 BV510 (clone IM7)	BD Biosciences	Cat#563114; RRID: AB_2738011	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Armenian Hamster monoclonal Anti-Mouse CD3e BUV395 (clone 145-2C11)	BD Biosciences	Cat#563565; RRID: AB_2738278	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Rat monoclonal Anti-Mouse CD4 BV786 (clone RM4-5)	BD Biosciences	Cat#563727; RRID: AB_2728707	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Rat monoclonal Anti-Mouse CD8a Pacific Blue (clone 53-6.7)	BioLegend	Cat#100725; RRID: AB_493425	Fetal thymus analyisis and bulk OP9-DL1 co-culture readout
Rat monoclonal Anti-Mouse CD43 Alexa Fluor 700 (clone S11)	BioLegend	Cat#143213; RRID: AB_2800660	AGM/YS pre-HSC and LMP analysis
Human recombinant monoclonal Anti-Mouse CD45 APC Vio770 (clone REA737)	Miltenyi	Cat#130-110-662; RRID: AB_2658231	LDA co-colture readout
Rat monoclonal Anti-Mouse CD25 APC (clone PC61)	BD Biosciences	Cat# 557192; RRID:AB_398623	LDA co-colture readout
Rat monoclonal Anti-Mouse CD4 Pacific Blue (clone RM4-5)	BD Biosciences	Cat# 558107; RRID:AB_397030	LDA co-colture readout
Rat monoclonal Anti-Mouse CD8a PE (clone 53-6.7)	BD Biosciences	Cat# 553033; RRID:AB_394571	LDA co-colture readout
Rat monoclonal Anti-Mouse CD45 BV605 (clone 30-F11)	BioLegend	Cat# 103139; RRID: AB_2562341	Brain-Thymus-FL *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse CD11b FITC (clone M1/70)	BioLegend	Cat#101206; RRID: AB_312789	Brain *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse CD117 (cKit) APC (clone 2B8)	BioLegend	Cat#105812; RRID: AB_313221	Brain *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse F4/80 AF700 (clone BM8)	BioLegend	Cat#123130; RRID: AB_2293450	Brain *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse CD4 BV786 (clone RM4-5)	eBioscience	Cat# 417-0042-82; RRID: AB_2921053	Thymus *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse CD8a PE-Cy7 (clone 53-6.7)	eBioscience	Cat# 25-0081-82; RRID: AB_469584	Thymus *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse CD45R/B220 APC (clone RA3-6B2)	BD Biosciences	Cat#553092; RRID: AB_398531	FL *Csf1r*^{MerCreMer} analysis
Rat monoclonal Anti-Mouse F4/80 FITC (clone BM8)	BioLegend	Cat#123108; RRID: AB_893502	FL

			Csf1r^{MerCreMer} analysis
Rat monoclonal Anti-Mouse Ter119 APC-ef780 (clone TER119)	eBioscience	Cat# 47-5921-80; RRID: AB_1548786	FL *Csf1r*^{MerCreMer} analysis
Streptavidin BV785	BioLegend	Cat#405233	FL *Csf1r*^{MerCreMer} LSK analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse Sca1 FITC (clone E13-161.7)	BioLegend	Cat# 122506; RRID: AB_756191	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse CD150 PE-Cy7 (clone mShad150)	eBioscience	Cat# 25-1502-82 RRID: AB_10805742	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Armenian Hamster monoclonal Anti-Mouse CD48 APC-Cy7 (clone HM48-1)	BD Biosciences	Cat# 561242; RRID: AB_10644381	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse CD45R/B220 Biotin (clone RA3-6B2)	BD Biosciences	Cat# 553086; RRID: AB_394615	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Armenian Hamster monoclonal Anti-Mouse CD3e Biotin (clone 145-2C11)	BD Biosciences	Cat#553060; RRID: AB_394593	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse F4/80 Biotin (clone BM8)	eBioscience	Cat# 13-4801-85; RRID: AB_466657	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse Gr-1 Biotin (clone RB6-8C5)	BD Biosciences	Cat# 553125; RRID: AB_394641	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Mouse monoclonal Anti-Mouse Nk1.1 Biotin (clone PK136)	Biolegend	Cat# 108704; RRID: AB_313391	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse Csf1r Biotin (clone AFS98)	BioLegend	Cat# 135508; RRID: AB_2085223	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse CD4 Biotin (clone GK1.5)	BD Biosciences	Cat# 553728; RRID: AB_395012	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse CD8a Biotin (clone 53-6.7)	eBioscience	Cat# 13-0081-85; RRID: AB_466346	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Armenian Hamster monoclonal CD11c Biotin (clone N418)	BioLegend	Cat# 117304; RRID: AB_313773	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Rat monoclonal Anti-Mouse Ter119 Biotin (clone TER119)	BD Biosciences	Cat# 553672; RRID: AB_394985	FL *Csf1r*^{MerCreMer} analysis (Lineage cocktail)
Whole-mount Immunofluorescence Antibodies			
Goat polyclonal anti-m/rCD31/PECAM1	R&D systems	Cat#AF3628	AGM, YS endothelium
Rabbit polyclonal anti-GFP	Invitrogen	Cat#A11122; RRID: AB_221569	AGM, YS EYFP detection
Chicken polyclonal anti-GFP	Invitrogen	Cat#A10262; RRID: AB_2534023	AGM, YS EYFP detection
Rat monoclonal anti-mouse CD117 (c-Kit) (clone 2B8)	eBioscience	Cat# 14-1171-82; RRID: AB_467433	AGM, YS hematopoietic cluster cells

Rabbit polyclonal anti-RFP	Rockland	Cat# 600-401-379;	AGM, YS tdTomato detection
Armenian hamster monoclonal anti-mouse CD27 (clone LG-7F9)	eBioscience	Cat#14-0271-82; RRID: AB_467183	AGM, YS hematopoietic cluster cells with lymphoid potential
Rabbit anti-mouse/rat/human Runx1, clone EPR3099	Abcam	Cat# ab92336; RRID: AB_2049267	AGM, YS hematopoietic cluster cells and HE
Donkey polyclonal anti-goat Alexa Fluor Plus 647	Invitrogen	Cat#A32849; RRID: AB_2762840	Secondary antibodies
Donkey polyclonal anti-rabbit Alexa Fluor Plus 488	Invitrogen	Cat#A32790; RRID: AB_2762833	Secondary antibodies
Donkey polyclonal anti-rabbit Alexa Fluor Plus 555	Invitrogen	Cat#A32794; RRID: AB_2762834	Secondary antibodies
Donkey polyclonal anti-rat Alexa Fluor 488	Invitrogen	Cat#21208; RRID: AB_2535794	Secondary antibodies
Donkey polyclonal anti-rat CF568	Biotium	Cat#20092; RRID: AB_	Secondary antibodies
Donkey anti-chicken IGY AF488	Invitrogen	Cat#A78948; RRID: AB_2921070	Secondary antibodies
Goat polyclonal anti-armenian hamster DyLight 649 (clone Poly4055)	BioLegend	Cat#405505; RRID: AB_1575122	Secondary antibodies
Reagents and Media			
7-aminoactinomycin D (7-AAD)	BioLegend	Cat#420404	Live/dead
Hoechst 33258	Hellobio	Cat#HB0786	Live/dead
Benzyl alcohol	Sigma-Aldrich	Cat#305197	Whole-mount immunofluorescence
Benzyl benzoate	Sigma-Aldrich	Cat#68183	Whole-mount immunofluorescence
MEM Alpha Medium	Gibco	Ref: 12000-063	Cell cultures
Sodium bicarbonate	Euroclone	Cat#ECM0980D	Cell cultures
FBS HyClone	Cytiva	Cat#SH30071.03	Cell cultures
Murine IL-7	PeproTech	Cat#217-17	Cell cultures
Murine Flt3-Ligand	PeproTech	Cat#250-31L	Cell cultures
Collagenase I	Merck		Single cell suspensions
4-hydroxytamoxifen (≥98% Z isomer)	Sigma Aldrich	Cat#H6278	In vivo Cre induction
Hot StarTaq Master Mix	QIAGEN	Cat#203445	PCR
Progesterone	Sigma Aldrich	Cat#P0130	In vivo Cre induction
Methocult GF M3434	STEMCELL Technologies	Cat# 03434	CFU-C assay
Deposited Data			
Single cell RNA sequencing data of YS+VU and AGM from E10.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* embryos (4-OHT at E8.5)	This paper	BioProject ID: PRJNA898269 accession numbers: SRR28006358, SRR28006359, SRR28006360, SRR28006361	
Cell Lines			
OP9	Gift from Marella de Bruijn	N/A	
OP9-Dl1	Gift from Juan Carlos	(Schmitt *et al*, 2004)	

	Zúniga-Pflücker		
Mouse Strains			
Cdh5-CreER^{T2}	Gift from R.Adams	(Wang *et al*, 2010)	
R26^{zsGreen}	Gift from M.Iannacone	RRID:IMSR_JAX:007906	
R26^{tdTomato}	The Jackson Laboratory	RRID:IMSR_JAX:007909	
R26^{EYFP}	The Jackson Laboratory	RRID:IMSR_JAX:006148	
Csf1r-iCre	(Deng *et al*, 2010)	RRID:IMSR_JAX:021024	
Csf1r^{MerCreMer}	(Qian *et al*, 2011)	RRID:IMSR_JAX:019098	
B6.SJL-Ptprca Pepcb/BoyJ (B6 CD45.1)	Gift from L. Naldini	RRID:IMSR_JAX:002014	
PCR Primers			
CreFW: TGATGGACATGTTCAGGGATC	Metabion	(Wang *et al.*, 2010)	*Cdh5-CreER^{T2},Csf1r-iCre* genotyping
CreRV: CAGCCACCAGCTTGCATGA	Metabion		
zsGreen WT FW: AAGGGAGCTGCAGTGGAGTA	Metabion	(Madisen *et al*, 2010)	*R26^{zsGreen}* genotyping
zsGreen WT RV: CCGAAAATCTGTGGGAAGTC	Metabion		
zsGreen TG FW: AACCAGAAGTGGCACCTGAC	Metabion		
zsGreen TG RV: GGCATTAAAGCAGCGTATCC	Metabion		
EYFP WT FW: CTGGCTTCTGAGGACCG	Metabion	(Srinivas *et al*, 2001)	*R26^{EYFP}* genotyping
EYFP WT RV: CAGGACAACGCCCACACA	Metabion		
EYFP TG FW: AGGGCGAGGAGCTGTTCA	Metabion		
EYFP TG RV: TGAAGTCGATGCCCTTCAG	Metabion		
tdTomato WT FW: AAGGGAGCTGCAGTGGAGTA	Metabion	(Madisen *et al.*, 2010)	*R26^{tdTomato}* genotyping
tdTomato WT RV: CCGAAAATCTGTGGGAAGTC	Metabion		
tdTomato TG FW: GGCATTAAAGCAGCGTATCC	Metabion		
tdTomato TG RV: CTGTTCCTGTACGGCATGG	Metabion		
Csf1rMer-RV: TGAACCAGCTCCCTATCTGC	Eurofins	(Qian *et al*, 2011)	*Csf1r^{MerCreMer} genotyping*
Csf1rMer-FW: CTTCCAAAGCATGGTCCAGT	Eurofins		
Software and Algorithms			
Imaris (v. 9.7.2)	Bitplane	RRID: SCR_007370	Image analysis
Zeiss Zen (v. 2.3 SP1)	Zeiss	https://www.zeiss.com/microscopy/int/products/microscope-software/zen.html	Image analysis
FlowJo (v. 10)	BD	RRID:SCR_008520	Flow cytometry analysis
GraphPad Prism (v. 9.4.1)	GraphPad Software	RRID:SCR_002798	Statistics
Photoshop CC 2019	Adobe	RRID:SCR_014199	Image analysis

Illustrator 2019	Adobe	RRID:SCR_010279	Figure preparation
Microsoft Excel (v. 16.41)	Microsoft	RRID:SCR_016137	Statistics
Seurat (v. 4.0)	(Hao *et al*, 2021)	N/A	scRNA-Seq analysis
SCTransform	(Hafemeister & Satija, 2019)	N/A	scRNA-Seq analysis
R (R-3.2.3 – R-4.2.1)	The R Foundation	https://www.r-project.org	scRNA-Seq analysis
Louvain	(Blondel *et al*, 2008)	N/A	scRNA-Seq analysis
CellRanger (v 6.1)	10x Genomics	https://support.10xgenomics.com	scRNA-Seq analysis
Clustree (v. 0.4.3)	(Zappia & Oshlack, 2018)	N/A	scRNA-Seq analysis
SingleR (v1.0.1)	(Aran *et al*, 2019)	N/A	scRNA-Seq analysis
BioRender	BioRender.com	RRID:SCR_018361	Schematics
ImageJ (v. 1.54f)	ImageJ	https://imagej.net/ij/	Image analysis
ShinyCell	(Liu et al., 2020)	http://shinycell1.ddnetbio.com	scRNA-Seq analysis
ClusterProfiler (v. 4.8.3)	(Wu *et al*, 2021)	N/A	scRNA-Seq analysis

Table S1 - Reagents and Resources.

Table S2 (enclosed as a separate Excel file) – Differential gene expression analysis between pre-HSPC 2 and pre-HSPC 1

Differentially expressed gene (DEG) list between pre-HSPC 2 and pre-HSPC 1 in the E10.5 AGM and YS+VU *Cdh5-CreER^{T2}::R26^{tdTomato}* scRNA-Seq dataset (4-OHT at E8.5). Statistically significant gene expression changes were considered as adjusted p value: <0.05.

Table S3 (enclosed as a separate Excel file) – Differential gene expression analysis between AGM tdTomato+ pre-HSPCs and YS+VU tdTomato+ pre-HSPCs

Differentially expressed gene (DEG) list between AGM tdTomato+ pre-HSPCs vs YS+VU tdTomato+ pre-HSPCs in the E10.5 *Cdh5-CreER^{T2}::R26^{tdTomato}* scRNA-Seq dataset (4-OHT at E8.5). Statistically significant gene expression changes were considered as adjusted p value: <0.05.

Table S4 (enclosed as a separate Excel file) –E11.5 AGM transplantation data

Summary of peripheral blood and bone marrow analysis of adult C57 BL/6 CD45.1 mice transplanted with *Cdh5-CreER^{T2}::R26^{zsGreen}* and *Cdh5-CreER^{T2}::R26^{tdTomato}* E11.5 FL, activated with 4-OHT at E8.5 or E10.5.

Table S5 (enclosed as a separate Excel file) –E14.5 FL primary transplantation data

Summary of peripheral blood and bone marrow analysis of adult C57 BL/6 CD45.1 mice transplanted with *Cdh5-CreER^{T2}::R26^{tdTomato}* E14.5 FL, activated with 4-OHT at E8.5 or E10.5.

Table S6 (enclosed as a separate Excel file) –E14.5 FL secondary transplantation data

Summary of peripheral blood and bone marrow analysis of adult C57 BL/6 CD45.1 mice transplanted with BM from E14.5 primary transplant recipients.

Supplemental References

Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR *et al* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20: 163–172

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008: P10008

Deng L, Zhou JF, Sellers RS, Li JF, Nguyen AV, Wang Y, Orlofsky A, Liu Q, Hume DA, Pollard JW *et al* (2010) A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *Am J Pathol* 176: 952–967

Hafemeister C, Satija R (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 20: 296

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M *et al* (2021) Integrated analysis of multimodal single-cell data. *Cell* 184: 3573–3587 e3529

Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, Ng LL, Palmiter RD, Hawrylycz MJ, Jones AR *et al* (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat Neurosci* 13: 133–140

Qian BZ, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, Kaiser EA, Snyder LA, Pollard JW (2011) CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* 475: 222–225

Schmitt TM, de Pooter RF, Gronski MA, Cho SK, Ohashi PS, Zuniga-Pflucker JC (2004) Induction of T cell development and establishment of T cell competence from embryonic stem cells differentiated in vitro. *Nat Immunol* 5: 410–417

Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F (2001) Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* 1: 4

Wang Y, Nakayama M, Pitulescu ME, Schmidt TS, Bochenek ML, Sakakibara A, Adams S, Davy A, Deutsch U, Luthi U *et al* (2010) Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* 465: 483–486

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L *et al* (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2: 100141

339 Zappia L, Oshlack A (2018) Clustering trees: a visualization for evaluating clusterings at multiple
340 resolutions. *Gigascience* 7
341

A new mouse model of JMML highlights differential susceptibility of embryonic hematopoietic stem/progenitor cells to the Kras^{G12D} mutation

Giulia Quattrini[1], Cristiana Barone[1], Alessandro Muratore[1], Gloria Zambelli[1], Filipa Timóteo-Ferreira[1], Thea Milanesi[1], Tiphanie Durfort[1], Mahdieh Naghavi Alhosseini[1], Silvia Bombelli[3], Deborah D'Aliberti[1], Valentina Sangiorgio[4], Cristina D'Orlando[1], Raffaella Meneveri[1], Barbara Vergani[1], Rocco Piazza[1,2], Silvia Brunelli[1,2], Emanuele Azzoni[1,2]

[1]*School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy*

[2]*Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy*

[3]*Human Technopole, Milan, Italy*

[4]*Hematology Division and Bone Marrow Unit, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico San Gerardo dei Tintori, Monza, Italy.*

Abstract

Juvenile Myelomonocytic Leukemia (JMML) is a rare pediatric myeloproliferative neoplasm, driven by mutations in the RAS pathway. Developing faithful models of JMML remains a major challenge, likely due to the critical role of the cellular context in which the mutation occurs and the prenatal origin of half of JMML cases. To address this, we developed a transgenic mouse model of JMML with prenatal onset that allows selective targeting of $Kras^{G12D}$ to specific subsets of hematopoietic stem/progenitor cells (HSPCs) during embryogenesis (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$). Here, we show that embryonic HSPCs exhibit differential susceptibility to $Kras^{G12D}$. Interestingly, while EMPs may not be involved in JMML pathogenesis, both HSC-independent progenitors and HSCs could act as cells of origin of JMML. Remarkably, by E14.5, fetal liver (FL) cells already showed defining features of JMML, such as GM-CSF hypersensitivity, along with a transcriptional and epigenetic reprogramming of fetal HSPCs, including metabolic rewiring with the expression of hypoxia-related pathways and Nfkb1-driven inflammation. These data identify an early pre-leukemic state, supporting the *in utero* origin of the disease. When $Kras^{G12D}$ was targeted to fetal HSCs, the JMML-resembling leukemic phenotype in adult mice was more aggressive and transplantable. Interestingly, at the transcriptional level BM cells showed a myeloid-biased differentiation program and enhanced inflammatory signature, further supported by the epigenetic enrichment of inflammatory-related transcriptional factors (FOS, FOSL2). These data suggest that clinical heterogeneity of JMML could be, at least in part, explained by variability in the cell of origin. By providing a faithful model of prenatal-onset JMML, our approach offers a valuable platform for studying the early pathogenic events underlying disease initiation and progression, thus providing an opportunity to identify targetable vulnerabilities.

Introduction

Juvenile myelomonocytic leukemia (JMML) is a rare and aggressive myeloproliferative disease of early childhood, with a median age of diagnosis of 2 years and an incidence rate of 1.2 cases per million[1,2]. The disease is clinically heterogeneous, but most of the cases are characterized by uncontrolled proliferation of myeloid and monocytic lineages, thrombocytopenia, anemia, splenomegaly, pulmonary infiltrations, and hypersensitivity to Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF)[3–5]. The prognosis is very poor, with allogenic stem cell transplantation being currently the most effective cure. However, the relapse rate is high, and the 5-years free survival rate only reaches about 50% of the cases[1,6,7] . This therapeutic gap highlights the urgent need for new and specific therapeutic approaches for this disease. JMML is characterized by hyperactivation of the RAS/MAPK pathways. In fact, the majority of cases of JMML (95%) are caused by a mutation in the RAS canonical pathway, including NF1, NRAS, KRAS, PTPN11, CBL, RRAS, with most of them being mutually exclusive[8–11]. In up to 30% of patients, secondary subclonal mutations have been identified, most commonly in SETBP1 and JAK3, associated with worse prognosis[12,13].

JMML-driving mutation can occur as germline or somatic, defining different disease subtypes with different clinical course. Germline mutations often associated with RASopathies, such as neurofibromatosis type 1 (NF1 mutations), CBL syndrome (CBL mutations), or Noonan syndrome (PTPN11 mutations) which predispose to JMML. Interestingly, about 5% of Noonan syndrome patients develop a mild and transitory myeloproliferative disorder, and only a minority of the cases progress to aggressive JMML[14]. On the contrary, somatic mutations usually show a more aggressive course, and they are most commonly found in PTPN11, KRAS and NRAS[9].

In nearly half of the cases of JMML, canonical somatic mutations were already present at birth, therefore underscoring the potential *in utero* origin of JMML[8,15]. However, the precise cellular origin of the disease has not yet been identified. Recent evidence has in fact pointed out the heterogeneity of disease-propagating cells in JMML, which includes – but are not restricted to – HSCs compartment[16,17].

Embryonic hematopoiesis is a complex and dynamic process, which relies not only on HSCs, but also on multiple subpopulations of HSCs-independent progenitors, emerging in sequential and partially overlapping waves during fetal development[18,19]. These early progenitors play essential roles in fetal hematopoiesis, and they can persist into adulthood[19], suggesting that they may also represent the targets of mutations responsible for leukemic transformation. Identifying the cell of origin of the

disease is crucial for a better understanding of its pathogenesis and for uncovering early molecular and epigenetic events that drive leukemic transformation.

However, investigating the early cellular events driving the development of JMML remains one of the major challenges. This disease is particularly difficult to model in vitro[4], as primary JMML cell in culture tends to rapidly differentiate and become apoptotic. Currently, immortalized lines from JMML cells are still lacking. Several strategies have been explored, each providing unique insights but also presenting important limitations. Recently, iPSC-based and 3D culture systems have been established, enabling the study of human JMML cells in controlled environments, capturing specific disease features such as GM-CSF hypersensitivity and abnormal myeloid expansion[20,21]. These in vitro approaches are a useful tool to study JMML features, but they lack the full complexity of the hematopoietic niche and do not reproduce in vivo developmental cues. Xenografts are currently the most promising model[17], preserving human disease characteristics and allowing therapeutic testing, but they have limitations mainly due to the rarity of the patients' samples. Moreover, this model does not allow the investigation of *in utero* pre-leukemic stage of the disease. Genetically engineered mouse models have been widely used to study disease mechanisms by introducing JMML-associated mutations into the hematopoietic system during development or postnatally[22–24]. However, due to the lack of knowledge of the cellular origin of JMML, these models are not fully representative of the early pathological events in humans and of the developmental context in which the driving mutations arise.

To fill this knowledge gap, we developed a new preclinical model that allows the introduction of JMML-associated driving mutation ($Kras^{G12D}$) in selected hematopoietic cell populations at specific timepoints during embryogenesis. Considering their essential contribution of HSCs-independent progenitors to fetal hematopoiesis and their persistence in adulthood[19], we hypothesize that JMML may arise not only from HSCs but also from one or more of these early progenitor populations. The different cellular origins of the disease could also correlate with the clinical heterogeneity of JMML.

By combining phenotypic, transcriptional and epigenetic analysis, we show that JMML can originate prenatally form both HSC-independent progenitors and HSCs, and the presence of a preleukemic state of fetal cells induced by fetal induction of $Kras^{G12D}$ mutation. Moreover, we show a myeloid-biased differentiation and a marked inflammatory signature in adult leukemic mice driven by embryonic HSCs-induced $Kras^{G12D}$. Therefore, our model provides a faithful platform to recapitulating the prenatal origin of JMML, enabling the identification of critical therapeutic vulnerabilities and the development of novel targeted treatments.

Results

Targeting Kras^{G12D} to erythro-myeloid progenitors (EMPs) leads to low penetrance slow progressing myeloproliferative disease

To investigate the prenatal origin of JMML, we developed a murine model that enables the selective targeting of Kras^{G12D} to different subsets of embryonic hematopoietic progenitors. We took advantage of a transgenic mouse model in which a tamoxifen-inducible Cre recombinase is under the control of the Cdh5 promoter (Cdh5-CreER^{T2})[18,19,25], expressed in endothelial cells, including hemogenic endothelium (HE), precursor to all definitive blood cells[18,26] . We crossed this mouse line with a LSL-Kras^{G12D}::R26^{tdTomato} [27,28] mouse, in which the expression of Kras^{G12D} is induced only upon Cre recombination, together with TdTomato as a Cre reporter gene. By administering 4-hydroxytamoxifen at specific embryonic timepoints (**Figure 1A**), we were able to target Kras^{G12D} to specific subsets of embryonic hematopoietic stem/progenitor cells (HSPCs) in the restricted time window in which they emerge from hemogenic endothelium.

In order to specifically target Kras^{G12D} to the first HE-derived hematopoietic wave, Cre was activated at embryonic day (E) 7.5. At this timepoint, the mutation is targeted to erythro-myeloid progenitors (EMPs), in addition to a subset of primitive erythrocytes, macrophages and megakaryocytes[18]. The expected Mendelian ratio between wild-type and transgenic littermates was observed at both fetal and postnatal stage, without visible developmental defects. At E12.5, transgenic embryos did not show any significant alteration in morphology or in the proportions of different subsets of erythroid and differentiated cells in the fetal liver (FL). Adult offspring were followed up by peripheral blood (PB) analysis without showing any significant difference between Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} and Cdh5-CreER^{T2}::R26^{tdTomato} littermate controls. At 38 weeks after birth, just one transgenic mouse out of four showed a significant increase in the absolute count of white blood cells (WBC) and monocytes, and in the frequency of myeloid cells in both PB and bone marrow (BM), together with a decreasing trend in the frequency of B and T cells (**Figure 1B-D**). No alterations were observed in the frequency of HSPCs in the BM (**Figure 1D**). This mouse also showed splenomegaly and pathological infiltration in spleen, liver and lungs (**Figure 1E, F**), and the occasional presence of undifferentiated blasts in PB, suggesting the presence of a slow progressing myeloproliferative disease with late onset (**Figure 1G**). All the other transgenic mice did not show any significant alterations compared to wild-type littermate controls. The frequency of recombination of tdTomato reporter gene revealed a progressive decrease in all the different cellular populations over the weeks in all the mice (**Figure 1H**). However, the percentage of recombination in the mice

showing features of myeloproliferative disease resulted increase at all the time points compared to both transgenic and wild type littermates (**Figure 1H**).

Taken together, these data suggest that EMPs are not susceptible to $Kras^{G12D}$ mutation, and RAS pathway does not appear to have a role in EMPs differentiation. Therefore, these hematopoietic progenitors are unlikely to represent the main cell of origin of JMML, consistently with what previously reported in a Ptpn11 model[29]. However, in a minority of cases, the late onset of a slow-progressing myeloproliferative disorder was observed, suggesting that leukemic transformation in these progenitors can sometimes occur.

Targeting $Kras^{G12D}$ to fetal-restricted hematopoietic progenitors leads to a severe lethal fetal liver anemia

In order to target $Kras^{G12D}$ to fetal-restricted hematopoietic stem/progenitor cells, 4-OHT was administered at E8.5. This activation mode allows to target the mutation to HSC-independent progenitors, including late EMPs and LMPs[18]. We have previously demonstrated that this wave of hematopoietic progenitors, originating from a specific subset of HE in the vitelline and umbilical arteries, represents the major contributor to fetal lympho-myelopoiesis[30]. Unexpectedly, targeting $Kras^{G12D}$ to these HSCs-independent progenitors caused embryonic lethality at early fetal stages (**Figure 2A**). The Mendelian ratio between transgenic and wild-type was altered due to the high rate of embryonic death occurring shortly after the activation of $Kras^{G12D}$ mutation. The latest stage that we were able to analyze was E12.5, at which transgenic embryos already showed a delay in their development (**Figure 2B**). The pale appearance of transgenic embryos suggested the presence of anemia, subsequently confirmed by FL cells analysis. Indeed, a significant reduction of the Ter119+ subpopulation was detected, as well as a prominent decrease of EMPs ($Ter119^{-}$ Kit^{+} $CD41^{+}$ $CD16/32^{+}$) (**Figure 2D**). A more detailed analysis of the erythroid subpopulations showed a reduction of the more differentiated subpopulations (S2, $CD71^{+}$ $Ter119^{low}$, and S3, $CD71^{+}$ $Ter119^{+}$,) in the transgenic embryo, and a concomitant increase in the frequency of early erythroid progenitors ($CD71^{-/low}$ $Ter119^{-}$) (**Figure 2E, F**), suggesting a block in erythroid lineage maturation, which is likely the cause of the observed developmental delay. These data suggest that the Ras pathway plays a crucial role along the erythroid differentiation route of fetal-restricted HSPCs, which appear to be particularly susceptible to $Kras^{G12D}$ mutation. Furthermore, our data is consistent with these progenitors significantly contributing to fetal erythropoiesis.

Targeting Kras^{G12D} to embryonic HSCs causes fetal anemia with perinatal lethality

To specifically target the mutation to the hematopoietic stem cell wave[18], Kras^{G12D} was induced at E10.5. While this activation strategy delayed the onset of lethality compared to earlier activation, it still resulted in embryonic death due to fetal anemia, though at later stage of development as compared to the E8.5 activation (**Figure 3A**). At E13.5/E14.5 embryos did not show any relevant morphological abnormalities (**Figure 3B**). However, a decrease in the differentiated erythrocytes (S3, CD71^{+} Ter119^{+}) can be detected, with a concomitant increase of the early erythroid progenitors (S0, CD71^{-} Ter119^{-}, and S2, CD71^{low} Ter119^{-}) (**Figure 3C-E**). Later on in development, the anemia phenotype became more severe. By E18.5, all the transgenic embryos show developmental delay, reduced growth and pale appearance (**Figure 3B**), and the erythroid differentiation defect involved all the stages of erythroid maturation. In fact, at this developmental stage erythroid cells were almost undetectable, suggesting a complete block in erythroid differentiation (**Figure 3C, F, G**).

Overall, these data indicate that the acquisition of Kras^{G12D} in the definitive waves of hematopoietic progenitors severely impairs erythroid lineage differentiation, ultimately leading to embryonic lethality. This suggests that Ras pathway is involved in erythroid differentiation of embryonic HSCs, which contribute to erythropoiesis at a later developmental stage than fetal-restricted HSPCs.

Mosaic-induced Kras^{G12D} fetal liver cells show an increase in the frequency of fetal HSPCs and the appearance of defining features of JMML

The severe anemia and consequent embryonic lethality caused by the targeting of Kras^{G12D} to fetal-restricted HSPCs (activation at E8.5) or to the HSC wave (activation at E10.5) did not allow the analysis of late fetal or postnatal stages. To circumvent this, we decided to reduce the number of targeted cells by lowering the dose of 4-OHT, thus effectively inducing Kras^{G12D} as a genetic mosaic in the target cell populations. First, half of the normal dose of 4-OHT (0,5 mg) was used (**Figure 4A**). This strategy allowed the analysis of later stages of embryonic development, with both activation time points. The recombination efficiency was assessed by Deep sequencing of the Kras locus in TdTomato^{+} and TdTomato^{-} fractions of E14.5 FL cells following the mosaic activation at E8.5 or E10.5. E 14.5 FL cells from Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos were sorted based on the expression of the tdTomato reporter and the sequencing was performed on cDNA to specifically quantify the relative expression of Kras^{G12D} and wild-type transcripts. The Variant Allele Frequency (VAF) relative to the Kras^{G12D} mutation within the tdTomato^{+} fraction was 0.43 in FL cells

activated at E8.5 and 0.39 in FL cells activated at E10.5, confirming the expected recombination of the LSL-Kras^{G12D} allele. In contrast, the tdTomato- fraction showed unexpected elevated VAFs (0.52 and 0.42 respectively), probably indicating that the recombination of Kras^{G12D} and TdTomato alleles did not always occur together. In fact, they are carried in two different transgenes, which undergo independent Cre recombination events. Despite the lack of concordance, these data indicate that the mosaic activation still resulted in a high level of recombination of the Kras^{G12D} allele.

Mosaic activation with a half-dose of 4-OHT resulted in a milder anemia phenotype as compared to the full-dose activation; and although it delayed embryonic lethality, it still did not allow postnatal analysis. Following activation at E8.5 or E10.5, embryonic development proceeded relatively normally until E14.5. At this stage, FLs were collected and analyzed. Organ size and the total cell number were significantly reduced in the transgenic embryos compared to littermate controls (**Figure 4B**). The erythroid compartment showed a delay in maturation, similar to what previously observed with the full dose activation, though less pronounced and again more severe when the mutation was induced at E8.5 (**Figure 4C**). The analysis of HSPCs compartment revealed a significant expansion in the LSK (Lineage$^-$ Kit$^+$ Sca1$^+$) population following both the activation at E8.5 and E10.5, more evident when the mutation was induced in embryonic HSCs (**Figure 4E, F**). Notably, activation at E10.5 led to a strong increase in MPPs (Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^+$ CD150$^-$) frequency, along with an increase in both frequency and absolute count of immunophenotypic HSCs (Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^-$ CD150$^+$). A similar but milder effect resulted from the activation at E8.5 (**Figure 4E, G, H**).

Mosaic-activated E14.5 FL cells were also tested for sensitivity to Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF), as it is one of the defining features of JMML[31]. Both HSC-independent progenitors-induced (4-OHT E8.5) and HSCs-induced (4-OHT E10.5) FL cells exhibited a dose-dependent increase in colony formation upon exposure to GM-CSF stimulation; however, the effect was significantly greater when Kras^{G12D} was induced in embryonic HSCs (4-OHT E10.5). (**Figure 4D**). This data indicates that the Kras^{G12D} mutation, if acquired during fetal development, drives the emergence of JMML-associated features in embryonic HSPCs, suggesting the onset of a pre-leukemic stage of the disease.

To characterize the transcriptomic changes following the acquisition of the Kras^{G12D} mutation in embryonic HSCs (4-OHT E10.5), wild type (Cdh5-CreER^{T2}::R26^{tdTomato}) and transgenic (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}) Lin$^-$ Kit$^+$ E14.5 FL cells were analyzed by single-cell RNA sequencing (scRNA-seq) (**Figure 5A**). This time point was chosen because HSPCs compartment expansion and GM-CSF hypersensitivity were notably more pronounced with this experimental condition. Cells distributed in three major trajectories in UMAP space, representing erythroid

(clusters 3, 4, 6, 9, 11, 14, 16, 17, 20, 21, 22, 23), myeloid (clusters 0,1, 7, 8, 15, 19, 25, 26), and HSPC lineages (clusters 2, 5, 10, 12, 13) (**Figure 5B**). While the proportions of cells in G1, S, and G2/M phases were comparable between conditions in myeloid and erythroid subsets, HSPCs compartment showed a proportional increase of cells in G2M phase in the transgenic sample compared to control, suggesting enhanced cell cycle progression (**Figure 5C**). Moreover, the relative abundance of the three main subpopulations showed marked differences between wild-type and transgenic samples, with a significant increase in myeloid and HSPCs subpopulations and a reduction in the erythroid subset, consistent with the previously observed anemia (**Figure 5D**). Transgenic cells, in particular the myeloid subset, showed an overall transcriptional reprogramming with a significant activation of metabolic pathways, cell cycle, and apoptotic regulation (**Figure 5E, G**). Enriched pathways included amino sugar metabolism, seleno compound metabolism, and HIF-1 signaling, suggesting a possible metabolic rewiring as a stress response to hypoxia, triggered by the erythroid differentiation defect. Moreover, the fetal HSPCs subsets revealed a significant upregulation of metabolic and biosynthetic pathways and DNA damage repair mechanisms, compatible with early features of leukemic transformation (**Figure 5F**). Notably, $Kras^{G12D}$-induced fetal HSPCs exhibited a significant upregulation of key myeloid markers, pointing to an early transcriptional reprogramming toward the myeloid fate (**Figure 5H**). Taken together, these data identify a pre-leukemic state of E14.5 FL cells upon acquisition of the $Kras^{G12D}$ mutation in embryonic HSCs, revealing an early transcriptional dysregulation in the HSPCs and myeloid cell compartment. Interestingly, Nfkb1, a key inflammatory transcription factor, was upregulated, suggesting the presence of a pro-inflammatory environment, similar to what recently reported in other JMML models[32–34]. Moreover, its expression was enhanced across all cell types (**Figure 5I**), suggesting a potential contribution of inflammation to the functional reprogramming of fetal HSPCs and the progression toward a leukemic phenotype.

We assessed genome-wide chromatin accessibility in wild type and $Kras^{G12D}$-induced $Lineage^{-}$ Kit^{+} FL cells through ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing). A total of 30 enhancers were identified as WT-specific, while 81 were TG-specific. Among the transcription factors enriched in transgenic samples, RUNX1, a master hematopoietic transcription factor often deregulated in other leukemias, emerged as the most significantly associated motif (**Figure 5J**) , underscoring the presence of early epigenetic reprogramming driven by $Kras^{G12D}$. Consistently, RUNX1 expression was found to be upregulated in transgenic samples analyzed by scRNA-seq (**Figure 5K**), further supporting the activation of RUNX1-regulated transcriptional programs.

Mosaic activation of Kras^{G12D} targeted to fetal HSPCs leads to aggressive JMML-resembling myeloproliferative disease in adult mice

In order to circumvent embryonic lethality and obtain viable offspring, the mosaic strategy was employed by further reducing the dose of 4-OHT to 10% of the standard dose (0,1 mg), thus activating the Kras^{G12D} mutation in a small subset of the target cell population. 4-OHT was administered at E8.5 or E10.5 (**Figure 6A**) With both activation time points, live pups were born in the expected Mendelian ratio. The recombination efficiency was assessed by Deep sequencing of the Kras locus in BM cells of adult mice following the mosaic activation at E8.5 or E10.5. BM cells from Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} mice were sorted based on the expression of the tdTomato reporter gene and sequencing was performed on cDNA. The Variant Allele Frequency (VAF) of the tdTomato+ fraction was 0.78 in BM cells activated at E8.5 and 0.72 in BM cells activated at E10.5, while the VAF of the tdTomato- fraction was 0,03 and 0,2, respectively, confirming the high recombination of the Kras^{G12D} despite the low dose of 4-OHT, likely due to the clonal expansion of cells in which the recombination of both transgenes occurred. All transgenic mice developed features of a myeloproliferative disease closely resembling JMML, starting around 8 weeks after birth. However, the disease progression differed based on the activation time point: in mice activated at E10.5 the disease was lethal between 10 and 20 weeks after birth, while in mice activated at E8.5 between 14 and 28 weeks (**Figure 6B**). All transgenic mice showed a progressive increase in monocytes and myeloid cells (CD45$^+$ CD11b$^+$) in the peripheral blood (PB), and a concomitant reduction of B (CD45$^+$ B220$^+$ CD3$^-$) and T lymphocytes (CD45$^+$ B220$^-$ CD3$^+$) (**Figure 6C, E**). Cytological analysis confirmed the enrichment of monocytic cells and the occasional presence of blasts in transgenic samples (**Figure 6D**). Mice in which the mutation was targeted to embryonic HSCs (4-OHT at E10.5) exhibited a more aggressive and fully penetrant leukemic phenotype, also showing anemia and a more pronounced splenomegaly (**Figure 6 C, F**), ultimately leading to death of all mice by 20 weeks of age. In contrast, activation in fetal-restricted hematopoietic progenitors (4-OHT E8.5) caused a milder phenotype with incomplete penetrance (**Figure 6B**). The increase in myeloid cells in both PB and BM was less marked and appeared at later timepoints compared to HSC-induced mice (**Figure 6E, H**). The increase in WBC and monocytes was detected sporadically (**Figure 6C**). Moreover, 2 out of 13 mice showed minimal clinical manifestations of the disease, indicating a more indolent phenotype. Interestingly, in one out of thirteen cases analyzed, the activation of Kras^{G12D} at E8.5 led to a leukemic phenotype resembling T-cell lymphoblastic leukemia in adult mice, with a marked expansion of T cell compartment in both PB and BM and the presence of splenomegaly (data not shown). These findings suggest a link between the clinical heterogeneity of JMML and the specific cell in which the driving mutation occurs. With both activations, pathological extra-medullary infiltrates could be observed in liver,

predominantly localized in the perivascular and intrasinusoidal regions. In the lungs, myeloid microaggregates could be detected mainly in peribronchial areas. Myeloid infiltrations were also detectable in spleen, along with the occasional presence of megakaryocytes, leading to the alteration of normal splenic architecture (**Figure 6G**). Bone marrow (BM) analysis revealed the reduction of hematopoietic progenitor populations frequency in both experimental conditions, with the most significant decrease observed in the HSC compartment (**Figure 6H**). These data suggest that both HSCs-independent progenitors and HSCs can be involved in the pathogenesis of the disease and that variance in the cell of origin might potentially correlate with the heterogeneity of the leukemic phenotype observed in JMML patients.

Transplantation of Kras^{G12D} HSC-induced, but not fetal-restricted HSPCs, causes a fully penetrant JMML-resembling disease originating prenatally

To confirm the prenatal origin of the disease, we performed transplantation of E14.5 FL cells in which the Kras^{G12D} mutation was targeted to HSCs-independent progenitors (E8.5) or fetal HSCs (E10.5) (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}) (**Figure 7A**), along with wild type (Cdh5-CreER^{T2}::R26^{tdTomato}) cells using a reduced dose of 4-OHT (0,5 mg). Mice transplanted with embryonic HSCs-induced FL cells (4-OHT E10.5) started showing typical myeloproliferative features around 5 weeks post-transplant, which led to death of all transplanted mice 16 weeks after transplant (**Figure7B**). Surprisingly, despite high levels of chimerism (**Figure 7C**), mice transplanted with FL cells activated at E8.5 did not exhibit any overt signs of leukemia and their survival rate was comparable to mice transplanted with wild-type cells (**Figure 7B, C, D, F, G, H, I, L**), suggesting that activation of the Kras^{G12D} mutation in fetal-restricted hematopoietic progenitors could not result in an engraftable leukemia. Notably, only one out of five transplanted mice showed a marked increase in reporter gene recombination (**Figure 7E**), leading to death within 12 weeks post-transplant. However, this mouse did not show any features of the JMML-resembling phenotype observed in mice transplanted with fetal HSCs-induced FL cells. The phenotype observed after transplantation of embryonic HSCs-induced FL cells was characterized a progressive increase in WBC, monocytes, and a marked expansion of the myeloid compartment in both PB and BM (**Figure 7C, D, G, H, I**). All these mice also showed a slight reduction of RBC and a decrease in B and T lymphocytes in PB (**Figure 7C, D, E, F**). Notably, all mice engrafted with embryonic HSC-induced cells also exhibited hepatosplenomegaly (**Figure 7G, H**), accompanied by a loss of normal splenic architecture and the presence of immature myeloid infiltrates in these organs, suggesting the presence of extramedullary hematopoiesis (**Figure 7I**). Significant infiltrates were also observed in both pulmonary and

extrapulmonary regions (**Figure 7I**). At BM level, mice transplanted with transgenic cells show a marked reduction of HSCs frequency, which was nearly undetectable in most of the mice (**Figure 7J**). This phenotype closely recapitulated the one observed in the adult mosaic model, but the leukemia was even more aggressive. To further confirm these data, transplantation was also performed with E14.5 FL cells following the activation of the $Kras^{G12D}$ mutation with the full dose of 4-OHT in fetal HSCs (4-OHT E10.5). The resulting phenotype closely mirrored that observed upon mosaic activation in fetal HSCs (4-OHT E10.5), supporting the central role of cellular context in which the mutation arises (**Figure S1**).

Single cell transcriptomics of embryonic HSCs-derived $Kras^{G12D}$ leukemias highlights aberrant activation of inflammatory pathways in the adult BM

To investigate the transcriptional landscape of $Kras^{G12D}$ leukemias arising from induction in embryonic HSCs (4-OHT at E10.5), BM cells (Lineage$^-$ Kit$^+$ CD45.1$^-$ CD45.2$^+$) isolated from FL-transplanted mice were then analyzed by scRNAseq (**Figure 8A, B**). Cell proportions confirmed the flow cytometry data (**Figure 7C, D, G**), showing a marked reduction of HSCs and myeloid-committed progenitors (**Figure 8C**). Notably, a highly enriched subpopulation of hematopoietic progenitors could be detected in the transgenic sample (clusters 15, 21, 22, 30, 32) (**Figure 8B, C**). These cells express an intermediate signature between erythroid and HSCs lineages (**Figure 8D**) Cluster-specific genes of this population indicate a transcriptional profile related to increased metabolic activity, mitochondrial involvement, and alterations in ribosomal activity, commonly observed in rapidly proliferating cells. Cell cycle analysis showed significant difference between wild-type and transgenic samples. In particular, in the JMML-enriched progenitors subset, a marked imbalance in the proportions of cells in G2/M and S phases could be observed in the transgenic sample. (**Figure8E**). Nfkb1 expression, previously found to be upregulated in HSPCs subset of FL cells (**Figure 5H**), was found to be significantly overexpressed in the HSCs subset of transgenic BM (**Figure 8F**), suggesting a progressive enhancement of the pro-inflammatory environment during disease progression. To further investigate the relationship between inflammation and disease progression, we analyzed the expression of main pro-inflammatory mediators. Interestingly, the myeloid subset of transgenic BM showed higher expression of several key inflammatory regulators (**Figure 8H**), suggesting a progressive pro-inflammatory transcriptional shift that involves different inflammatory pathways. To gain insight on the functional state of hematopoietic progenitors upon acquisition of $Kras^{G12D}$ mutation, a panel of key regulatory HSPCs genes (Myc, Ccnd1, Cdk4, Cdk6, Kit, Flt3, Gata2, Hoxa9, Runx1, Notch1, Jak2, Bcl2, Mpl, Socs3, Cxcr4) was analyzed, suggesting a

potential functional impairment of the transgenic HSCs subset, involving both proliferation and lineage differentiation (**Figure 8H**). These data suggest that activation of $Kras^{G12D}$ mutation in embryonic HSCs (4-OHT at E10.5) disrupts essential transcriptional programs required for HSC maintenance and hematopoietic homeostasis, ultimately leading to the enhancement of a myeloid-biased differentiation program.

To investigate the chromatin accessibility landscape of $Kras^{G12D}$ leukemias arising from induction in embryonic HSCs, $Lineage^{-}$ $c\text{-}Kit^{+}$ bone marrow (BM) cells isolated from these mice were analyzed by ATAC-seq. A total of 84 enhancers were identified as WT-specific, while 89 enhancers were TG-specific. Among the transcription factors enriched in TG-specific enhancers, FOSL2 emerged as the most significantly associated motif (**Figure 8I**), followed by FOS and STAT6. These transcriptional factors are RAS downstream targets, and they were demonstrated to be involved in the regulation of monocytic and neutrophilic lineages differentiation[35,36]. This suggests that the observed myeloid/monocytic bias in adult BM may be driven, at least in part, by enhanced activity of these transcription factors. However, the role of these transcriptional factors has not yet been investigated in JMML. Notably, Fos and Fosl2 are reported to correlate with inflammation[37,38], consistent with the inflammatory signature observed at the transcriptional level.

These results show that $Kras^{G12D}$ acquisition in embryonic HSCs leads to a transcriptional and epigenetic alteration of progenitor cells in the adult BM, consistent with activation of RAS signaling pathways. This reprogramming involves a transcriptional shift toward myeloid lineage differentiation and the enhanced expression of key inflammatory markers, particularly in HSCs and myeloid compartment. The enhanced activity of inflammation-related transcriptional factors further supports the potential role of inflammation in disease progression.

To assess the translational relevance of these results, a comparative reanalysis of scRNAseq data from bone marrow HSPCs (Lin- CD34+) of two JMML patients, one with an NF1 mutation and the other with a PTPN11 mutation, was performed[16]. Differential expression analysis revealed shared transcriptional changes with both FL cells and BM cells; however, the majority of these changes were shared with either BM or FL individually, rather than with both developmental stages. (**Figure 9A**). Among these, several genes are known transcriptional and epigenetic regulators of HSCs stemness, in particular the HOXA9-MEIS1-PBX1 axis and genes involved in myeloid differentiation, such as IRF8 and CEBPB. Interestingly, the expression of DNMT3A, TET2 and ASXL1, epigenetic regulators and drivers of clonal hematopoiesis, is consistently maintained across murine FL, murine BM and JMML patients. Their expression patterns show similar signature in BM-derived murine cells and human HSPCs, whereas FL cells express only a subset of these regulators (**Figure 9B, C**).

Notably, the inflammatory signature observed in murine BM closely mirror the one expressed in human HSPCs, confirming its clinical relevance (**Figure 9D**).

Overall, the overlap of transcriptional patterns expression between murine BM cells and human HSPCs demonstrates that the murine model faithfully recapitulates key aspects of the human disease, supporting the translational relevance and consistency of the model.

Discussion

Accumulating evidence suggests that disease-propagating cells in JMML are highly heterogeneous and are not restricted to the HSC compartment[16,17]. However, the role of embryonic HSPCs as cells of origin for JMML is unclear. This is particularly relevant given the frequent pre-natal origin of the disease[8,16] and the emerging complexity of embryonic hematopoiesis that has recently surfaced, which has pointed to a significantly higher and long-lived contribution of HSC-independent hematopoietic progenitors than previously thought[39]. To investigate the potential of embryonic/fetal HSPCs to act as cells of origin of the disease, we generated a new murine model of $Kras^{G12D}$-driven JMML with prenatal origin, based on a lineage tracing strategy[30] that allows the selective targeting of $Kras^{G12D}$ to discrete subsets of hemogenic endothelium, the direct precursor of embryonic HSPCs emerging at different timepoints during fetal development.

Our data demonstrate that embryonic HSPCs show differential susceptibility to JMML-associated mutation. In particular, EMPs, emerging from E7.5, are not affected by $Kras^{G12D}$ mutations, and therefore the Ras pathway does not appear to be involved in their differentiation, suggesting that they are likely not the main cell of origin of JMML. On the contrary, we show that both HSC-independent fetal-restricted hematopoietic progenitors emerging from HE at E8.5 and embryonic HSCs emerging at E10.5, are highly susceptible to $Kras^{G12D}$ mutation. Surprisingly, our data demonstrate a strong involvement of Ras pathway in fetal erythroid differentiation of HSC-independent progenitors and HSCs, as activation of $Kras^{G12D}$ in these progenitors leads to lethal fetal anemia, due to a delay in erythroid differentiation (**Figure 2F, 3E, 3G**).

To circumvent the embryonic lethality caused by the erythroid differentiation block, a mosaic-induction strategy was applied by reducing the dose of 4-OHT to the 10% of the standard dose. This activation mode, in addition to allowing survival of the offspring (**Figure 4A**), brings the model closer to the human disease by activating the mutation in a reduced subset of the targeted population, which can eventually undergo clonal expansion. Our results on postnatal disease development revealed that both HSC-independent progenitors and HSCs can undergo leukemic transformation upon induction of $Kras^{G12D}$, although the resulting phenotype exhibited different severity. Targeting $Kras^{G12D}$ to embryonic HSCs induced the most aggressive leukemic phenotype in adult mice, faithfully recapitulating the main clinical features of human disease. Nevertheless, as targeting the mutation to HSCs-independent progenitors resulted in a comparable, although milder, phenotype in adult mice, these progenitors could also potentially act as cell of origin of JMML. Notably, in both experimental conditions, a reduction of HSCs in BM was observed (**Figure 6H**). This defect could be attributed to

a cell-autonomous effect of the mutation when $Kras^{G12D}$ was targeted to fetal HSCs. On the other hand, when the mutation was targeted to HSCs-independent progenitors, the observed reduction in the immunophenotypic HSCs compartment could be due to a non-cell autonomous effect, in which $Kras^{G12D}$-induced HSCs-independent progenitors may alter the hematopoietic microenvironment, indirectly leading to a depletion of HSCs. Supporting this notion, we observed that the activation of $Kras^{G12D}$ mutation at E8.5 also targets tissue-resident macrophages, potentially contributing to microenvironment impairments (data not shown). As increasing evidence are demonstrating that leukemic progression is sustained by an altered microenvironment[20,40], it is possible to hypothesize that the impaired microenvironment of the BM might induce this defect in HSCs compartment. Further studies will aim to define better the correlation between the BM microenvironment and the $Kras^{G12D}$ mutation.

KRAS mutations are known to be involved in leukemogenesis and they are typically detected at late stages of AML[41], most often during disease progression or relapse. Previous studies have also shown that these mutations contribute to clonal expansion by accelerating the contribution of mutated HSCs to all hematopoietic lineages[23]. In this context, Kras mutations seem to be associated to disease progression, rather than as initiating mutation in postnatal leukemia, which is necessary for disease advancement but insufficient for leukemogenesis on their own. In contrast, our model demonstrates that when Kras mutation is induced during embryogenesis, it is sufficient to induce JMML at postnatal stage, highlighting the critical importance of the developmental stage and cellular context in which the mutation arises in disease onset.

The early pathogenic events taking place *in utero* were analyzed following the mosaic induction of the mutation in fetal-restricted HSCs-independent progenitors and embryonic HSCs. At E14.5, FLs showed a developmental delay, most likely driven by the erythroid defect, that occurred with both full and mosaic activation, despite not being embryonic lethal in the latter case (**Figure 4C**). Interestingly, at this stage we could detect an expansion of the HSCs compartment, more marked upon acquisition of the mutation in embryonic HSCs. This suggests a cell-autonomous effect of the mutation on the transcriptional reprogramming, which leads to an increase in proliferation and metabolic activity (**Figure 5C, E, F**). Interestingly, the mutation also induced a myeloid-biased transcriptional program in fetal HSPCs already at E14.5 (**Figure 5G**). Taken together, these data identify a pre leukemic state of E14.5 FL cells, with defining features of JMML. Interestingly, Nuclear factor kappa B subunit 1 (Nfkb1), a key regulator of inflammation, was found to be upregulated in transgenic FL cells compared to wild-type counterpart, with a higher expression observed in the HSPCs subset (**Figure 5G, H**), indicating a possible role of inflammation in disease onset. In contrast, activation of the mutation in fetal HSPCs led to a more marked defect in fetal erythropoiesis,

resulting in lethality at earlier stage of fetal development, whereas the expansion of the progenitor compartment was comparatively limited. Further analyses will be needed in order to gain insight into how the mutation differentially affects the transcriptional landscape when targeted to distinct subsets of fetal HSPCs.

While $Kras^{G12D}$ acquisition in embryonic HSCs resulted in an expansion of the fetal liver phenotypic HSCs compartment, in the adult BM it was associated with a reduction. Notably, the extent of HSC reduction in adulthood was inversely proportional to their expansion in FL, suggesting that the fetal HSC expansion may be driven by an uncontrolled increase in self-renewal activity, ultimately leading to an exhaustion of the HSCs pool in adult BM. In contrast, $Kras^{G12D}$ activation in HSC-independent progenitors led to a milder fetal expansion of the phenotypic HSCs compartment and a correspondingly milder reduction in adulthood. This difference is likely due to a non-cell-autonomous effect, potentially mediated by alteration in the microenvironment.

As different features suggested a leukemic potential of FL cells carrying the mutation when it was induced in either fetal-restricted HSC-independent progenitors and embryonic HSCs, they were both tested to verify their transplantability. HSC-induced FL cells showed the capability of fully recapitulating the leukemic phenotype when transplanted in adult recipients. This data confirms that HSC-induced FL cells can be considered as disease-propagating cells. In contrast, FL cells carrying the $Kras^{G12D}$ mutation in HSC-independent progenitors did not lead to any manifestation of leukemic phenotype upon transplantation. Despite the high level of donor chimerism observed in all the transplanted mice, the recombination of the tdTomato reporter gene shows a limited contribution of the HSC-independent progenitors compartment to the recipient repopulation in both wild-type and transgenic samples activated at E8.5, as the expression of the reporter gene in these mice is lower compared to those transplanted with wild-type or transgenic samples activated at E10.5. For this reason, the reduced expansion of this subpopulation does not appear to be caused by the mutation itself, but rather by the intrinsic properties of HSC-independent progenitors. In fact, it is important to consider the physiological features of these progenitors, which are known to have a reduced long-term repopulation capability compared to fetal HSCs[30]. Although these progenitors are expected to display higher repopulation potential than what we observed[30], their capability may have been further compromised by the sublethal irradiation of the recipient mice instead of the lethal irradiation, creating a less favorable environment for their expansion. In addition, targeting the mutation to these progenitors resulted in a delayed onset of disease in adult mice compared to fetal HSCs-induced adult mice (**Figure 6B**), suggesting that they may require a longer latency to acquire pre-leukemic features, including transplantability.

Given the JMML-like phenotype induced by transplantation of embryonic HSCs-induced FL cells (4-OHT E10.5) was the most aggressive, this model was analyzed to further investigate the transcriptional and epigenetic changes occurring at the postnatal stage. Surprisingly, in the transgenic BM, an enriched subset of aberrant progenitors was detected (**Figure 8B, C**). This subpopulation shows marked expression of metabolic and proliferation-related genes, along with a significant cell cycle dysregulation, suggesting a potential involvement in disease progression. Fetal activation of $Kras^{G12D}$ mutation also induced a marked profound transcriptional and epigenetic reprogramming of HSPCs in adult BM, characterized by the altered expression of key regulatory genes controlling proliferation, lineage commitment, and HSC maintenance. While $Kras^{G12D}$-induced FL cells showed enhanced metabolic activity, $Kras^{G12D}$-induced BM cells showed a reduction of metabolic activity compared to both wild-type BM and $Kras^{G12D}$-induced FL cells. Moreover, $Kras^{G12D}$-induced BM cells showed an increased chromatin accessibility at enhancers associated with RAS downstream transcription factors such as FOS and FOSL2. These transcriptional factors have also been associated with inflammation, therefore confirming the broad inflammatory signature observed at transcriptional level, particularly marked in myeloid compartment, with elevated Tnf, Il6, Il1b, and Tbk expression. Notably, this inflammatory signature appears progressively enhanced from FL to adult BM, suggesting an increase in inflammatory signature during disease development. These findings further support the role of inflammation in disease pathogenesis, as recently demonstrated in other JMML models[32–34].

In summary, we have generated a faithful model of $Kras^{G12D}$ driven JMML with prenatal origin, which is able to recapitulate the main features of the human disease. This model represents a powerful platform to investigate cellular and molecular events along disease onset and progression. Our results indicate the crucial role of the cellular context in which JMML-associated mutations arise, underscoring the correlation between the clinical heterogeneity and the cell of origin of JMML. Furthermore, the potential contribution of inflammation to disease onset and progression suggests that inflammatory pathway may be amenable to therapeutic targeting.

Materials and methods

Mice and embryos

Cdh5-CreER^{T2} [25], LSL-Kras^{G12D} [27]and R26^{tdTomato} [28] transgenic mice were previously described and were genotyped according to reported protocols. Cdh5-CreER^{T2} females aged from 6 to 16 weeks were subjected to overnight timed matings with LSL-Kras^{G12D}::R26^{tdTomato} males. Successful mating was judged by the presence of vaginal plugs the morning after, which was considered 0.5 days post conception (E0.5). Embryos were collected and dissected between E12.5 and E18.5 and evaluated on morphological criteria. For inducing Kras^{G12D}, a single dose of 37.5 mg/kg of 4-hydroxytamoxifen (4-OHT) dissolved in corn oil was administered by intra-peritoneal (i.p.) injections to pregnant females at E7.5, E8.5 or E10.5. To counteract adverse effects of 4-OHT on pregnancies, 4-OHT solutions were supplemented with progesterone (18.75 mg/kg). All transgenic mouse lines were maintained on a CD45.2 C57BL/6 genetic background, except for females used for timed matings to generate adult mice with 4-OHT activation during embryogenesis, which were instead of C57BL/6/FVB mixed background (F1). Mice were housed with free access to food and water at the San Raffaele Scientific Institute and University of Milan Institutional mouse facilities. All experiments were performed in accordance with experimental protocols approved by San Raffaele Scientific Institute and University of Milan Institutional Animal Care and Use Committees (IACUC).

Flow cytometric analysis and cell sorting

Single-cell suspensions were obtained from fetal liver by enzymatic digestion with calcium/magnesium free PBS supplemented with FBS 10%, Penicillin-Streptomycin 1%, EDTA 2mM and collagenase type I (Sigma) 0.12% (w/v) for 15 minutes at 37°C with gentle shaking, followed by a mechanical dissociation performed by pipetting. Peripheral blood (PB) samples of adult mice were collected from tail vein by bleeding using a scalpel. Adult bone marrow (BM) was obtained by flushing long bones using a 21G syringe and filtered in 40µm strainers. Spleen samples were obtained by smashing with a 70µm strainers. FL, PB, BM and spleen samples were treated with the appropriate amount of RBC Lysis Buffer. Single cell suspensions were incubated with conjugated antibodies and processed for flow cytometry. Dead cells were excluded based on Hoechst 33258 (Hellobio) or 7-AAD (Sigma) incorporation. Voltages, compensation and gates were set using unstained, single stained and fluorescence-minus-one (FMO) controls. Flow cytometry data acquisition was carried out using a LSR Fortessa X-20 (BD) analyzer and BD FACSDiva software (version 8.0.2). Cell sorting was performed with BD FACSDiscover S8 Cell Sorter or Invitrogen

Bigfoot Spectral Cell Sorter. An average sorting rate of 500-1000 events per second at a sorting pressure of 25 psi with a 100µm nozzle was maintained. Flow cytometric data was analyzed using FlowJo 758 software version 10 (BD).

In vivo transplantation

For transplantation experiments, syngeneic C57BL/6 (CD45.1) recipient mice were sublethally irradiated (4,5 Gy) before intra-venous transplantation of $1x10^6$ unfractionated FL cells obtained from E14.5 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos (4-OHT activation at E8.5 or E10.5). Donor-derived chimerism, cellular proportions and percentage of labeling (tdTomato+) was determined by flow cytometry in PB every 2 weeks, from 3 weeks post transplantation. The final analysis (6-15 weeks after transplant) included also flow cytometric analysis of BM and spleen cells; histological analysis of spleen, liver and lungs using hematoxylin-eosin staining; May-Grunwald Giemsa staining on blood smears and cytospin preparations of BM and PB cells.

GM-CSF-dependent colony forming unit assay

FL were dissected from E14.5 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos, processed in a single cells suspension as previously described and seeded in 30% M3120 Methocult (Stem cell technologies), 30% FBS, 2,45mM L-Glutamine, 440nM β-Mecaptoethanol in IMDM (Gibco), supplementer with serial diluitions of GM-CSF. Cells were plated in duplicate Greiner 35mm vented plates (10000 cells/dish) and cultured at 37°C, 5% CO2 in a humidified chamber. Colonies were scored after 7 days.

Single-cell RNA sequencing (scRNAseq)

FL were dissected from E14.5 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos, processed in a single cells suspension as previously reported. By FACS sorting, live Lineage$^-$ Kit$^+$ cells were isolated. BM cells were isolated as previously described from adult C57BL/6 (CD45.1) recipient mice after FL cells transplant. By FACS sorting, live Lineage$^-$ CD45.1$^-$ CD45.2$^+$ Kit$^+$ cells were isolated. Single-cell scRNA libraries from all the samples were generated using a Chromium instrument (10x Genomics) with a Chromium Next GEM Single Cell 3' HT. Libraries were quantified using a Qubit fluorometer (Thermo Fisher) and their profile was analyzed using a TapeStation instrument (Agilent). NGS sequences were generated using a Novaseq 6000 instrument (Illumina) with a target of 25000 reads/cell. Following multiplexing, raw fastq reads were processed using cellranger v9.0.1 and

aligned against the mm10 mouse genome (GENCODE vM23/Ensembl 98). Filtered count matrices generated with cellranger were processed with Seurat v5.3.0[42] package implemented in R (versions 4.5.0). Cells were filtered using the Scran package, excluding those with outlier gene counts and a mitochondrial read count fraction greater than 0.50 from downstream processing. The dataset in **Figure 5** included 30039 cells (20541 FL WT, 9498 FL TG); The dataset in **Figure 8** included 14568 cells (8408 BM WT, 6160 BM TG). After converting individual matrices in Seurat objects via the Read10X function, data were normalized and transformed using the SCTransform Variance Stabilizing Transformation method, while also regressing-out for feature count, percent of mitochondrial counts and cell phase. Data generated from all the samples were subsequently integrated with a Canonical Correlation Analysis using the PrepSCTIntegration, FindIntegrationAnchors and IntegrateData commands, by using SCT as normalization method. Dimensionality reduction of the integrated data was initially carried-on using Principal Component Analysis (PCA) and subsequently with Uniform Manifold Approximation and Projection (UMAP) algorithms, by retaining the first 30 principal components of the PCA. Clusters were identified with the Louvain algorithm; their number was selected using the Findcluster tool. Individual cell types were identified with manual data curation.

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)

FL were dissected from E14.5 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos, processed in a single cells suspension as previously reported. By FACS sorting, live Lineage- Kit+ cells were isolated. BM cells were isolated as previously described from adult C57BL/6 (CD45.1) recipient mice after FL cells transplant. By FACS sorting, live Lineage- CD45.1- CD45.2+ Kit+ cells were isolated. For each sample, chromatin tagmentation was performed as described in Diagenode ATAC Kit. Unique Dual-Indexed primers (UDIs, Diagenode) were used to construct the libraries by amplifying tagmented DNA. Each library was quantified by Bioanalyzer (4200 TapeStation, Agilent) and pooled together in equimolar concentrations. Pooled libraries were sequenced on four lanes of Illumina NovaSeq using paired-end 150-bp reads. ATAC-seq data processing and alignment was completed following the ENCODE pipeline standards. The mm10 genome build used for alignment was obtained from UCSC. Briefly, all FASTQ files were trimmed to remove Illumina Nextera adapter sequences using fastp with “-l 50 h” options. FastQC was used post-trimming to evaluate sequence quality and confirm successful trimming. Bowtie2was used to align reads to the mm10 reference genome . Samtools was employed to sort BAM files and filter for uniquely mapped reads. Mitochondrial reads and blacklist regions were removed using samtools idxstats and bedtools,

respectively. Picard was then used to remove duplicates using the MarkDuplicates tool. Bedtools was used to convert de-duplicated BAM files into BED files, to shift the Tn5 cutting sites. Subsequently, MACS2 was employed to identify chromatin-accessible regions (peaks) from the BED files. Bedtools intersect was used to identify putative enhancers overlapping chromatin-accessible regions by using FANTOM5 mouse enhancer annotation database[43]. An in-house pipeline was then developed to identify differentially activated enhancers between transgenic and wild-type mice across FL and BM samples. HOMER2 was then employed to perform transcription factor binding site (TFBS) enrichment analysis on the sets of TG versus WT enhancers.

Quantification and statistical analysis

Statistical analyses were performed using GraphPad Prism v10.2.1. No specific randomization method was used. Animals were allocated into experimental groups according to their genotype. No specific methods were used for blinding, but in general samples were collected from mice by one individual and then processed and analyzed by different individuals, at which time genotypes or experimental conditions of each sample were not known. To determine the level of significance, unpaired two-tailed Student t-test, one-way and two-way ANOVA followed by Tukey's multiple comparisons test were used as indicated in figure legends. *$p< 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

Figure 1

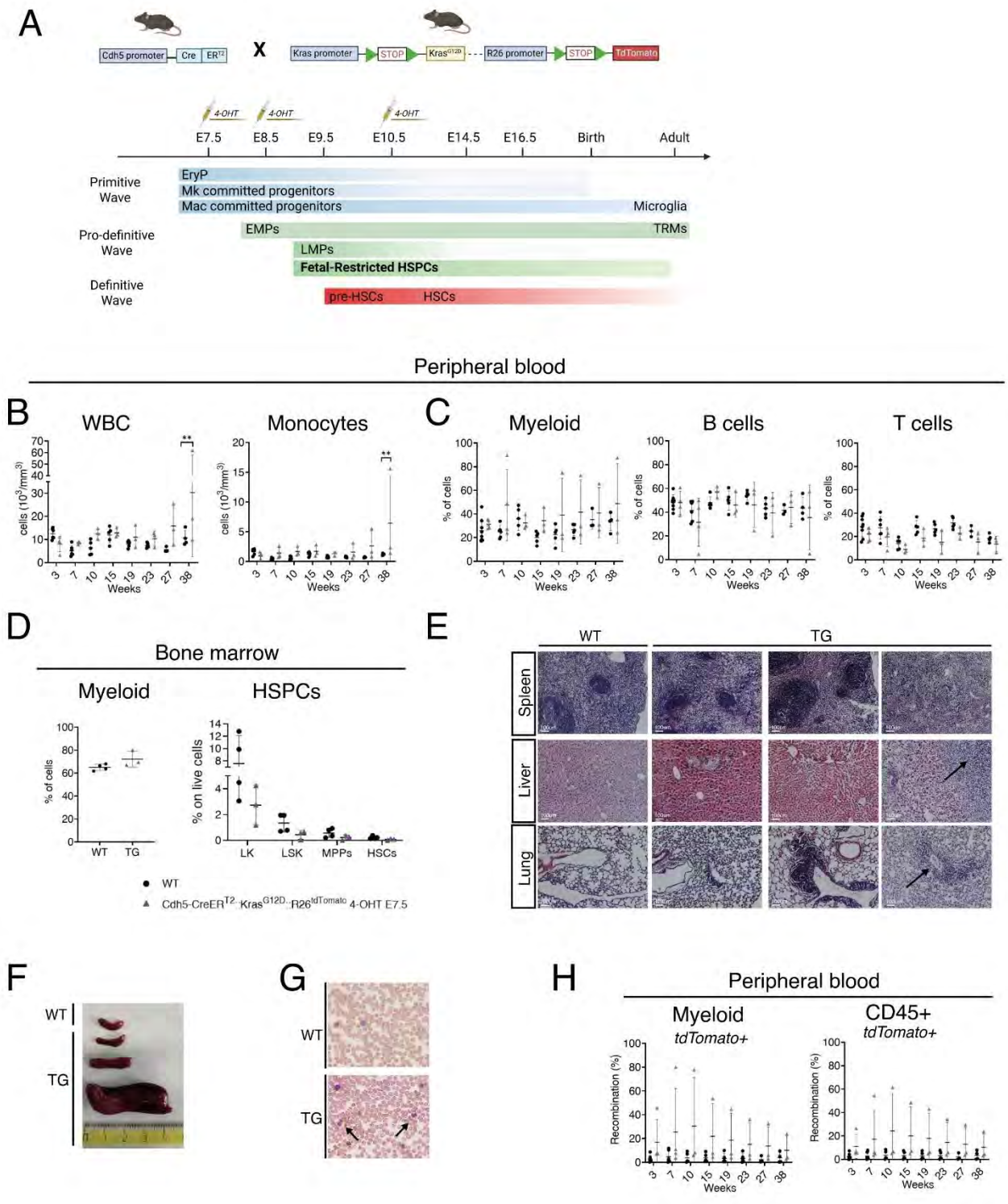

A
Cdh5 promoter
Cre
ER^T2
X
Kras promoter
STOP
Kras^G12D
R26 promoter
STOP
TdTomato
4-OHT
4-OHT
4-OHT
E7.5
E8.5
E9.5
E10.5
E14.5
E16.5
Birth
Adult
Primitive Wave
EryP
Mk committed progenitors
Mac committed progenitors
Microglia
Pro-definitive Wave
EMPs
TRMs
LMPs
Fetal-Restricted HSPCs
Definitive Wave
pre-HSCs
HSCs
Peripheral blood
B
WBC
Monocytes
C
Myeloid
B cells
T cells
cells (10^3/mm^3)
% of cells
Weeks
D
Bone marrow
Myeloid
HSPCs
% of cells
% on live cells
WT
TG
LK
LSK
MPPs
HSCs
WT
Cdh5-CreER^T2::Kras^G12D::R26^tdTomato 4-OHT E7.5
E
WT
TG
Spleen
Liver
Lung
F
WT
TG
G
WT
TG
H
Peripheral blood
Myeloid
tdTomato+
CD45+
tdTomato+
Recombination (%)
Weeks

Figure 1: Phenotypic characterization of slow-progressing myeloproliferative disease upon activation of Kras^{G12D} in EMPs (4-OHT E7.5)

(A) Schematic illustration of the model Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} and the experimental strategy.

(B) Longitudinal analysis of WBC and monocytes in PB of Cdh5CreER^{T2}::R26^{TdTomato} and Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} adult mice activated at E7.5. WT (n= 7), Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} 4-OHT E7.5 (n= 3). Error bars represent mean ± SD. ** $p < 0.01$ (two-way ANOVA followed by Tukey's multiple comparisons test). Samples were analyzed individually; mice were derived from one litter.

(C) Quantification of flow cytometric analysis of myeloid (CD45$^+$ CD11b$^+$), B cells (CD45$^+$ B220$^+$ CD3e$^-$) and T cells (CD45$^+$ B220$^-$ CD3e$^+$) on CD45$^+$ cells in PB. WT (n= 7), Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} 4-OHT E7.5 (n= 3). Error bars represent mean ± SD. Samples were analyzed individually; mice were derived from one litter.

(D) Quantification of flow cytometric analysis of myeloid (CD45$^+$ CD11b$^+$), LK (Lineage$^-$ Kit$^+$ Sca1$^-$), LSK (Lineage$^-$ Kit$^+$ Sca1$^+$), MPPs (Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^+$ CD150$^-$) and HSCs Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^+$ CD150$^+$) in the BM of 38-week-old adult mice. WT (n= 7), Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} 4-OHT E7.5 (n= 3). Error bars represent mean ± SD. Samples were analyzed individually; mice were derived from one litter.

(E) Hematoxylin-eosin staining of spleen, liver and lungs of 38-week-old Cdh5CreER^{T2}::R26^{TdTomato} (n=1) and Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} (n=3) adult mice activated at E7.5. Arrowheads indicate the presence of pathological infiltrations in liver and lungs.

(F) Gross morphology of spleen from Cdh5CreER^{T2}::R26^{TdTomato} and three Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} adult mice at 38 weeks.

(G) Blood smears from WT (Cdh5CreER^{T2}::R26^{TdTomato}) and TG (Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato}). Arrowheads indicate the presence of blast.

(H) Quantification of flow cytometric analysis of TdTomato$^+$ cells in CD45$^+$ and myeloid cells (CD45$^+$ CD11b$^+$) in PB. WT (n= 7), Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} 4-OHT E7.5 (n= 3). Error bars represent mean ± SD. Samples were analyzed individually; mice were derived from one litter.

Figure 2

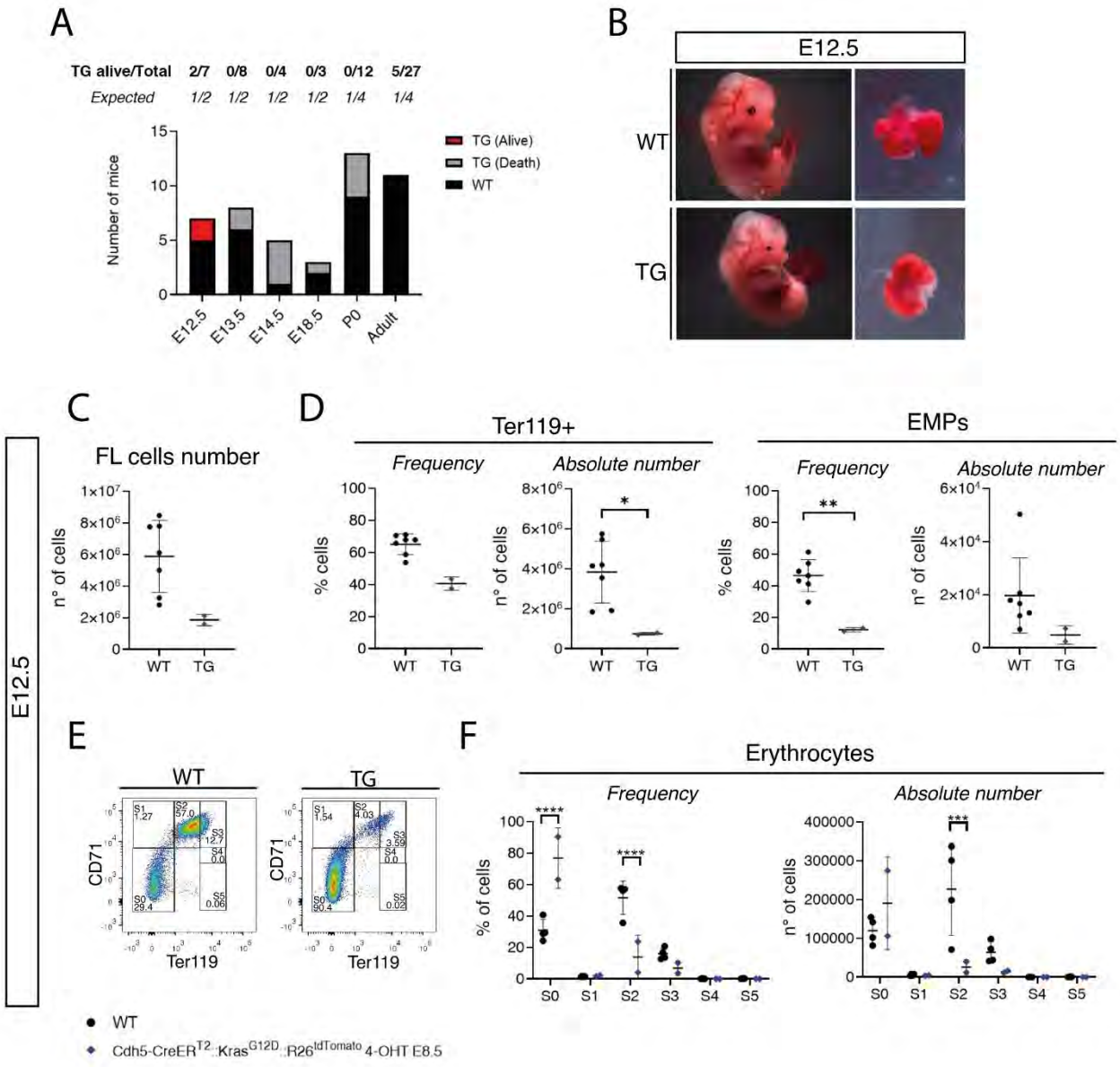


Figure 2: Lethal FL anemia in E12.5 Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} embryos upon activation of Kras^{G12D} in fetal-restricted HSPCs (4-OHT E8.5)

(A) Viability of E12.5-P0 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos compared to expected Mendelian ratio.

(B) Developmental defect in E12.5 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (TG) embryos and FL compared to Cdh5-CreER^{T2}::R26^{tdTomato} (WT) activated at E8.5.

(C) Absolute number of E12.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=7) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (n=2). Error bars represent mean ± SD. FLs were analyzed individually; embryos were derived from one litter.

(D) Quantification of flow cytometric analysis and absolute count of $Ter119^+$ cells and EMPs ($Ter119^-$ Kit^+ $CD41^+$ $CD16/32^+$) in E12.5 FL from Cdh5-$CreER^{T2}$::$R26^{tdTomato}$ (n=7) and Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$ (n=2). Error bars represent mean ± SD. FLs were analyzed individually; embryos were derived from one litter.

(E) Representative flow cytometric analysis of erythrocytes in E12.5 unfractionated FL cells from WT (Cdh5-$CreER^{T2}$::$R26^{tdTomato}$) and TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$). S0, S1, S2, S3, S4, S5 subpopulations were identified based on the expression of CD71 and Ter119.

(F) Quantification of flow cytometric analysis displayed in (E) and absolute count of S0, S1, S2, S3, S4, S5 erythroid subpopulations in E12.5 FL cells from Cdh5-$CreER^{T2}$::$R26^{tdTomato}$ (n=7) and Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$ (n=2). Error bars represent mean ± SD. Data are shown as mean ± SD. *** $p < 0.001$; **** $p < 0.0001$ (two-way ANOVA followed by Tukey's multiple comparisons test). FLs were analyzed individually; embryos were derived from one litter.

Figure 3

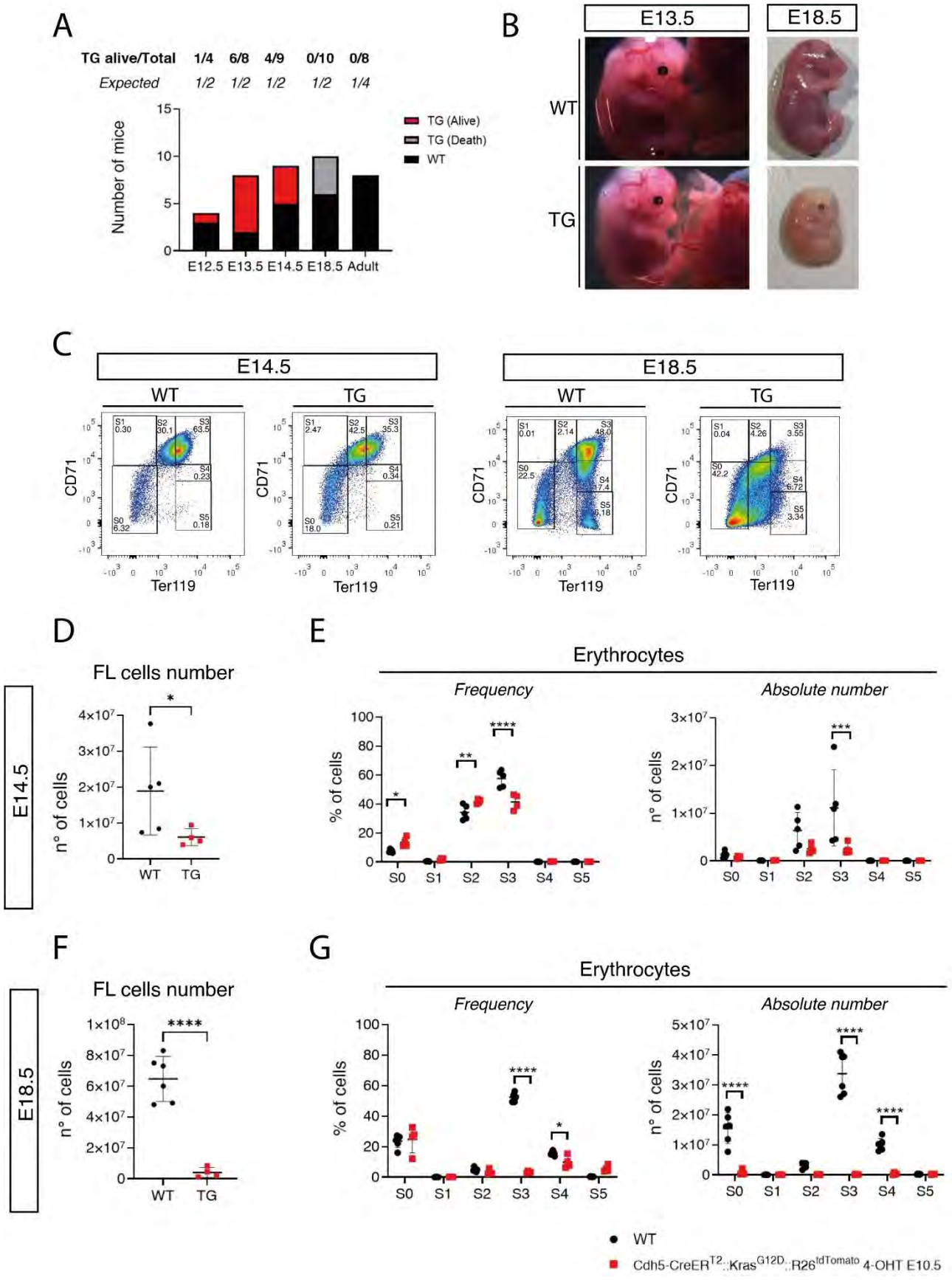

A
TG alive/Total 1/4 6/8 4/9 0/10 0/8
Expected 1/2 1/2 1/2 1/2 1/4
Number of mice
TG (Alive)
TG (Death)
WT
E12.5 E13.5 E14.5 E18.5 Adult
B
E13.5
E18.5
WT
TG
C
E14.5
E18.5
WT
TG
CD71
Ter119
D
FL cells number
n° of cells
WT
TG
E14.5
E
Erythrocytes
Frequency
Absolute number
% of cells
n° of cells
S0 S1 S2 S3 S4 S5
F
FL cells number
E18.5
G
Erythrocytes
Frequency
Absolute number
WT
Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} 4-OHT E10.5

Figure 3: Erythroid differentiation defect in Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} embryo upon activation of Kras^{G12D} in embryonic HSCs (4-OHT E10.5)

(A) Viability of E12.5-P0 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} embryos compared to expected Mendelian ratio.

(B) Developmental defect in E13.5 and E18.5 Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (TG) embryos and FL compared to Cdh5-CreER^{T2}::R26^{tdTomato} (WT) activated at E10.5.

(I) Representative flow cytometric analysis of erythrocytes in E14.5 and E18.5 unfractionated FL cells from WT (Cdh5-CreER^{T2}::R26^{tdTomato}) and TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}). S0, S1, S2, S3, S4, S5 subpopulations were identified based on the expression of CD71 and Ter119.

(C) Absolute number of E14.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=5) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (n=4). Error bars represent mean ± SD. *p< 0.05; (two-tailed unpaired Student's t-test).

(D) Quantification of flow cytometric analysis displayed in (C) and absolute count of S0, S1, S2, S3, S4, S5 erythroid subpopulations in E14.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=5) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (n=4). Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test). FLs were analyzed individually; embryos were derived from one litter.

(E) Absolute number of E18.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=6) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (n=4). Error bars represent mean ± SD. **** p < 0.0001; (two-tailed unpaired Student's t-test).

(F) Quantification of flow cytometric analysis displayed in (C) and absolute count of S0, S1, S2, S3, S4, S5 erythroid subpopulations in E18.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=6) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} (n=4). Error bars represent mean ± SD. *p< 0.05; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test). FLs were analyzed individually; embryos were derived from one litter.

Figure 4

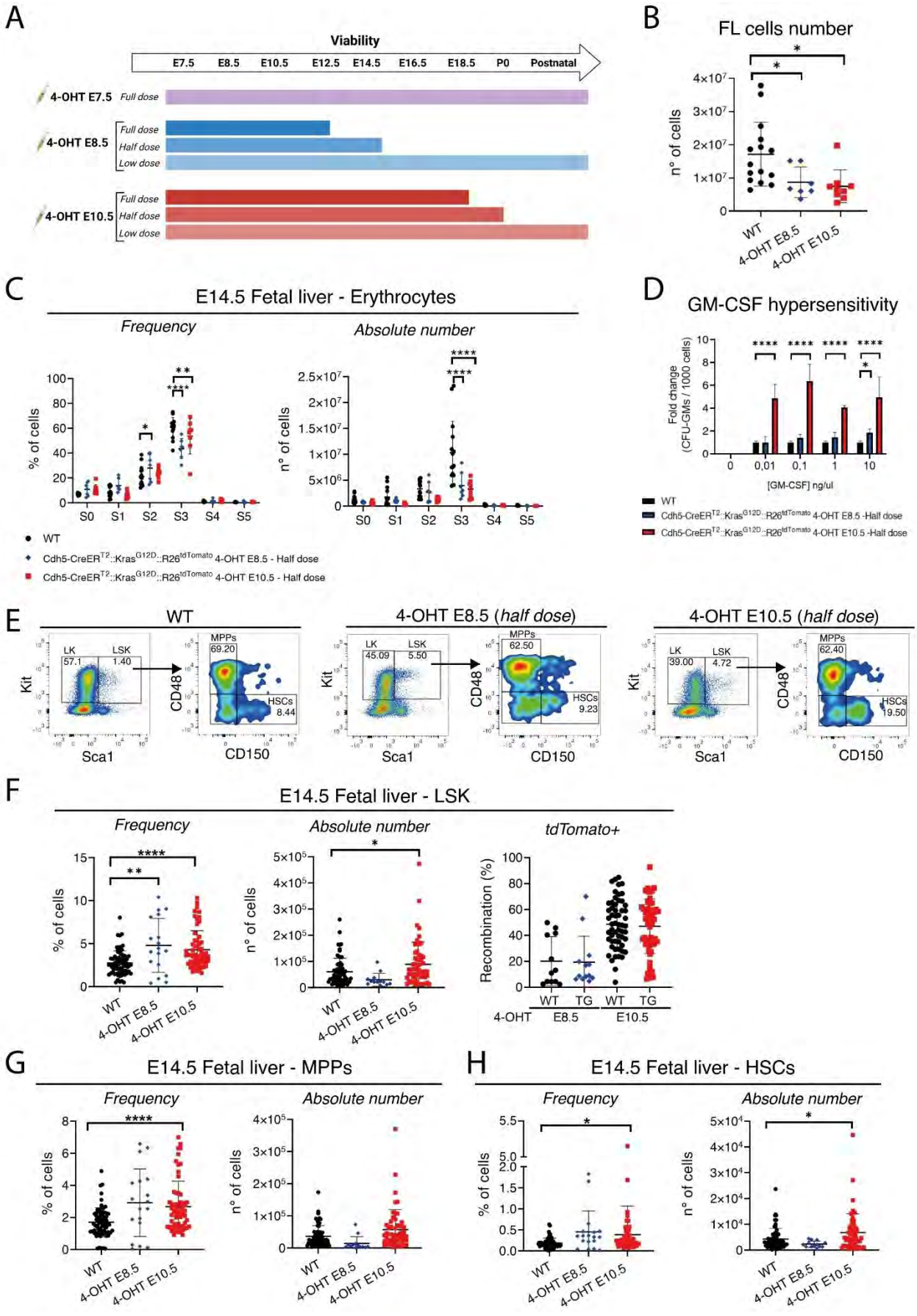

A
Viability
E7.5 E8.5 E10.5 E12.5 E14.5 E16.5 E18.5 P0 Postnatal
4-OHT E7.5 Full dose
4-OHT E8.5 Full dose Half dose Low dose
4-OHT E10.5 Full dose Half dose Low dose
B
FL cells number
n° of cells
WT
4-OHT E8.5
4-OHT E10.5
C
E14.5 Fetal liver - Erythrocytes
Frequency
Absolute number
% of cells
n° of cells
S0 S1 S2 S3 S4 S5
WT
Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} 4-OHT E8.5 - Half dose
Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} 4-OHT E10.5 - Half dose
D
GM-CSF hypersensitivity
Fold change (CFU-GMs / 1000 cells)
[GM-CSF] ng/ul
WT
Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} 4-OHT E8.5 -Half dose
Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} 4-OHT E10.5 -Half dose
E
WT
4-OHT E8.5 (half dose)
4-OHT E10.5 (half dose)
Kit
Sca1
CD48
CD150
LK
LSK
MPPs
HSCs
F
E14.5 Fetal liver - LSK
Frequency
Absolute number
tdTomato+
% of cells
n° of cells
Recombination (%)
WT TG WT TG
4-OHT E8.5 E10.5
G
E14.5 Fetal liver - MPPs
H
E14.5 Fetal liver - HSCs

Figure 4: JMML features in mosaic-induced E14.5 FL cells (4-OHT E8.5, E10.5)

(A) Schematic representation of viability of Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} mice following activation of Kras^{G12D} at E7.5, E8.5, or E10.5 with full dose of 4-OHT (1 mg), half dose (0,5 mg) and low dose (0,1 mg).

(B) Absolute number of E14.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=15) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=7) or E10.5 (n=9). Error bars represent mean ± SD. *p< 0.05; (one-way ANOVA followed by Tukey's multiple comparisons test).

(C) Quantification of flow cytometric analysis and absolute count of S0, S1, S2, S3, S4, S5 erythroid subpopulations in E14.5 unfractionated FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=15, analyzed individually in three independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=7, analyzed individually in two independent experiments) or E10.5 (n=8, analyzed individually in one experiment). Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test). FL were analyzed individually; embryos were derived from three litters and analyzed in independent experiments.

(D) Colony Forming Unit–Granulocyte–Macrophage (CFU-GM) count after 7 days of culture of E14.5 FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=4, analyzed in two independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=4, analyzed in one experiment) or E10.5 (n=2, analyzed in one experiment). Data are shown as fold change relative to the wild-type controls of the same experiment. *p< 0.05; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test). Samples were derived from two litters and analyzed in two independent experiments.

(E) Representative flow cytometric analysis of LK (Lineage$^-$ Kit$^+$ Sca1$^-$), LSK (Lineage$^-$ Kit$^+$ Sca1$^+$), MPPs (Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^+$ CD150$^-$) and HSCs (Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^-$ CD150$^+$) in E14.5 FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (WT) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 or E10.5. Lineage cocktail: B220, CD19, CD3e, F4/80, Gr1, Nk1.1, Ter119, 7-AAD.

(F) Quantification of flow cytometric analysis displayed in (E) and absolute count of LSK (Lineage$^-$ Kit$^+$ Sca1$^+$) (left panel) in E14.5 FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=71, derived from 25 litters and analyzed individually in 25 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=18, derived from 5 litters and analyzed individually in 5 independent experiments) and E10.5 (n=60, derived from 20 litters and analyzed individually in 20 independent experiments). Expression of TdTomato reporter

gene on LSK (right panel) from Cdh5-CreER^{T2}::R26^{tdTomato} (n=12, derived from 4 litters and analyzed individually in 4 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=14, derived from 5 litters and analyzed individually in 5 independent experiments); Cdh5-CreER^{T2}::R26^{tdTomato} (n=57, derived from 19 litters and analyzed individually in 19 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E10.5 (n=60, derived from 19 litters and analyzed individually in 19 independent experiments). Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test).

(G) Quantification of flow cytometric analysis displayed in (E) and absolute count of MPPs (Lineage^{-} Kit^{+} Sca1^{+} CD48^{+} CD150^{-}) in E14.5 FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=71, derived from 25 litters and analyzed individually in 25 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=18, derived from 5 liters and analyzed individually in 5 independent experiments) and E10.5 (n=60, derived from 20 litters and analyzed individually in 20 independent experiments). Error bars represent mean ± SD. **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test).

(H) Quantification of flow cytometric analysis displayed in (E) and absolute count of HSCs (Lineage^{-} Kit^{+} Sca1^{+} CD48^{-} CD150^{+}) in E14.5 FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} (n=71, derived from 25 litters and analyzed individually in 25 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=18, derived from 5 litters and analyzed individually in 5 independent experiments) and E10.5 (n=60, derived from 20 litters and analyzed individually in 20 independent experiments). Error bars represent mean ± SD. **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test).

Figure 5

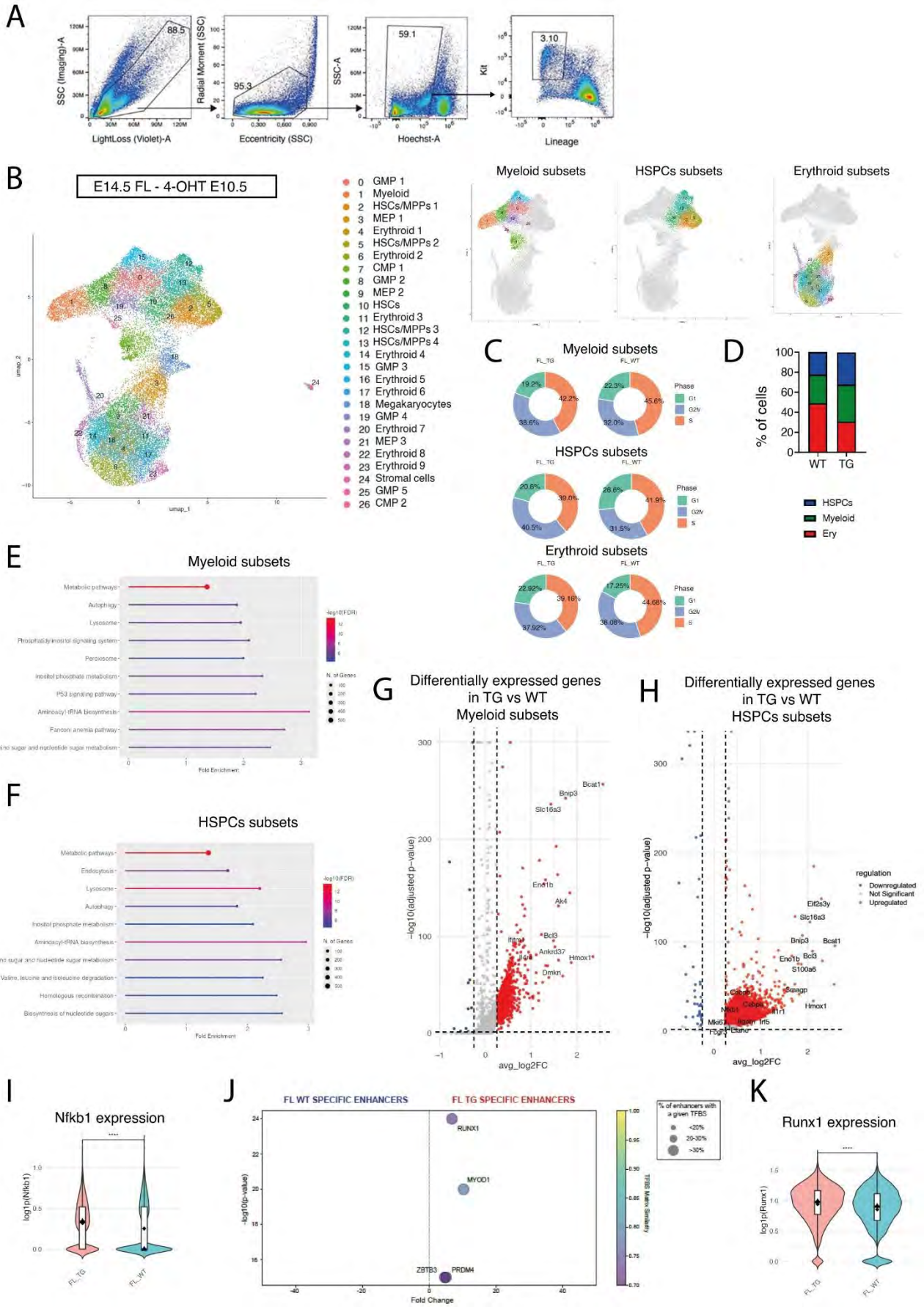

A
SSC (Imaging)-A
LightLoss (Violet)-A
88.5
Radial Moment (SSC)
Eccentricity (SSC)
95.3
SSC-A
Hoechst-A
59.1
Kit
Lineage
3.10
B
E14.5 FL - 4-OHT E10.5
0 GMP 1
1 Myeloid
2 HSCs/MPPs 1
3 MEP 1
4 Erythroid 1
5 HSCs/MPPs 2
6 Erythroid 2
7 CMP 1
8 GMP 2
9 MEP 2
10 HSCs
11 Erythroid 3
12 HSCs/MPPs 3
13 HSCs/MPPs 4
14 Erythroid 4
15 GMP 3
16 Erythroid 5
17 Erythroid 6
18 Megakaryocytes
19 GMP 4
20 Erythroid 7
21 MEP 3
22 Erythroid 8
23 Erythroid 9
24 Stromal cells
25 GMP 5
26 CMP 2
umap_1
umap_2
Myeloid subsets
HSPCs subsets
Erythroid subsets
C
Myeloid subsets
FL_TG
FL_WT
19.2%
42.2%
38.6%
22.3%
45.6%
32.0%
Phase
G1
G2M
S
HSPCs subsets
20.6%
39.0%
40.5%
28.6%
41.9%
31.5%
Erythroid subsets
22.92%
39.16%
37.92%
17.25%
44.68%
38.08%
D
% of cells
WT
TG
HSPCs
Myeloid
Ery
E
Myeloid subsets
Metabolic pathways
Autophagy
Lysosome
Phosphatidylinositol signaling system
Peroxisome
Inositol phosphate metabolism
P53 signaling pathway
Aminoacyl-tRNA biosynthesis
Fanconi anemia pathway
Amino sugar and nucleotide sugar metabolism
-log10(FDR)
N. of Genes
Fold Enrichment
F
HSPCs subsets
Metabolic pathways
Endocytosis
Lysosome
Autophagy
Inositol phosphate metabolism
Aminoacyl-tRNA biosynthesis
Amino sugar and nucleotide sugar metabolism
Valine, leucine and isoleucine degradation
Homologous recombination
Biosynthesis of nucleotide sugars
G
Differentially expressed genes in TG vs WT
Myeloid subsets
−log10(adjusted p-value)
avg_log2FC
Bcat1
Bnip3
Slc16a3
Eno1b
Ak4
Bcl3
Ifitm3
Ankrd37
Hmox1
Jun
Dmkn
H
Differentially expressed genes in TG vs WT
HSPCs subsets
regulation
Downregulated
Not Significant
Upregulated
Eif2s3y
Slc16a3
Bnip3
Bcat1
Eno1b
Bcl3
S100a6
Smagp
Cebpa
Hmox1
Nkd1
Mki67
Ifi5
I
Nfkb1 expression
log1p(Nfkb1)
FL_TG
FL_WT
J
FL WT SPECIFIC ENHANCERS
FL TG SPECIFIC ENHANCERS
RUNX1
MYOD1
ZBTB3
PRDM4
-log10(p-value)
Fold Change
% of enhancers with a given TFBS
TFBS Matrix Similarity
K
Runx1 expression
log1p(Runx1)
FL_TG
FL_WT

Figure 5: scRNAseq of E14.5 FL cells following the acquisition of the Kras^{G12D} mutation in embryonic HSCs (4-OHT E10.5)

(A) Representative flow cytometric gating strategy for the sorting of live Lineage^{-} Kit^{+} cells for scRNAseq analysis of E14.5 FL cells from Cdh5-CreER^{T2}::R26^{tdTomato} and Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato} embryos.

(B) Uniform Manifold Approximation and Projection (UMAP) of 30039 cells isolated as displayed in (A), from 11 E14.5 FL WT (Cdh5-CreER^{T2}::R26^{tdTomato}) and 15 E14.5 FL TG (Cdh5-CreER^{T2}::Kras^{G12D}::R26^{tdTomato}) activated at E10.5. FL WT cells (20541) and FL TG cells (9498) were separately sequenced. Cells are colored according to individual cluster identities. UMAPs on the right highlight myeloid subsets (clusters 0,1, 7, 8, 15, 19, 25, 26), HSPCs subsets (clusters 2, 5, 10, 12, 13) and erythroid subsets (clusters 3, 4, 6, 9, 11, 14, 16, 17, 20, 21, 22, 23).

(C) Donut plots reporting the percentage of cells in each phase of the cell cycle (G1, S, G2/M) for myeloid, erythroid and HSPCs subset for WT and TG samples. Cell cycle phases were inferred from single-cell transcriptomic data using canonical marker genes.

(D) Proportion of different cell types in WT and TG samples. Percentages were calculated relative to the total number of cells sequenced per sample.

(E) –(F) KEGG pathways enriched among genes upregulated in TG vs WT in myeloid and HSPCs subsets. Terms are ranked by the number of genes associated with each term. Dot size indicates the number of genes, and color intensity corresponds to statistical significance (-log(FDR)).

(G) – (H) Volcano plot of differentially expressed genes in FL TG vs FL WT in myeloid and HSPCs subsets. Each point represents a single gene, x-axis shows log2 fold change, and the y-axis shows statistical significance (–log10 adjusted p-value). Genes significantly upregulated FL TG cells are shown in red, and downregulated genes in blue. Selected genes of interest are labeled.

(I) Violin plot of the distribution of Nfkb1 expression in FL TG and FL WT, calculated by pseudobulk analysis. **** $p < 0.0001$ (Wilcoxon rank-sum test)

(J) TFBS enrichment analysis in the aberrantly inactive and active enhancers in FL WT and FL TG. The color scale represents the similarity of the sequences found in these enhancers with respect to specific TFBS matrixes. The dot sizes indicate the percentage of aberrantly active or inactive enhancers which harbor that TFBS.

(K) Violin plot of the distribution of Runx1 expression in FL TG and FL WT, calculated by pseudobulk analysis. **** $p < 0.0001$ (Wilcoxon rank-sum test)

Figure 6

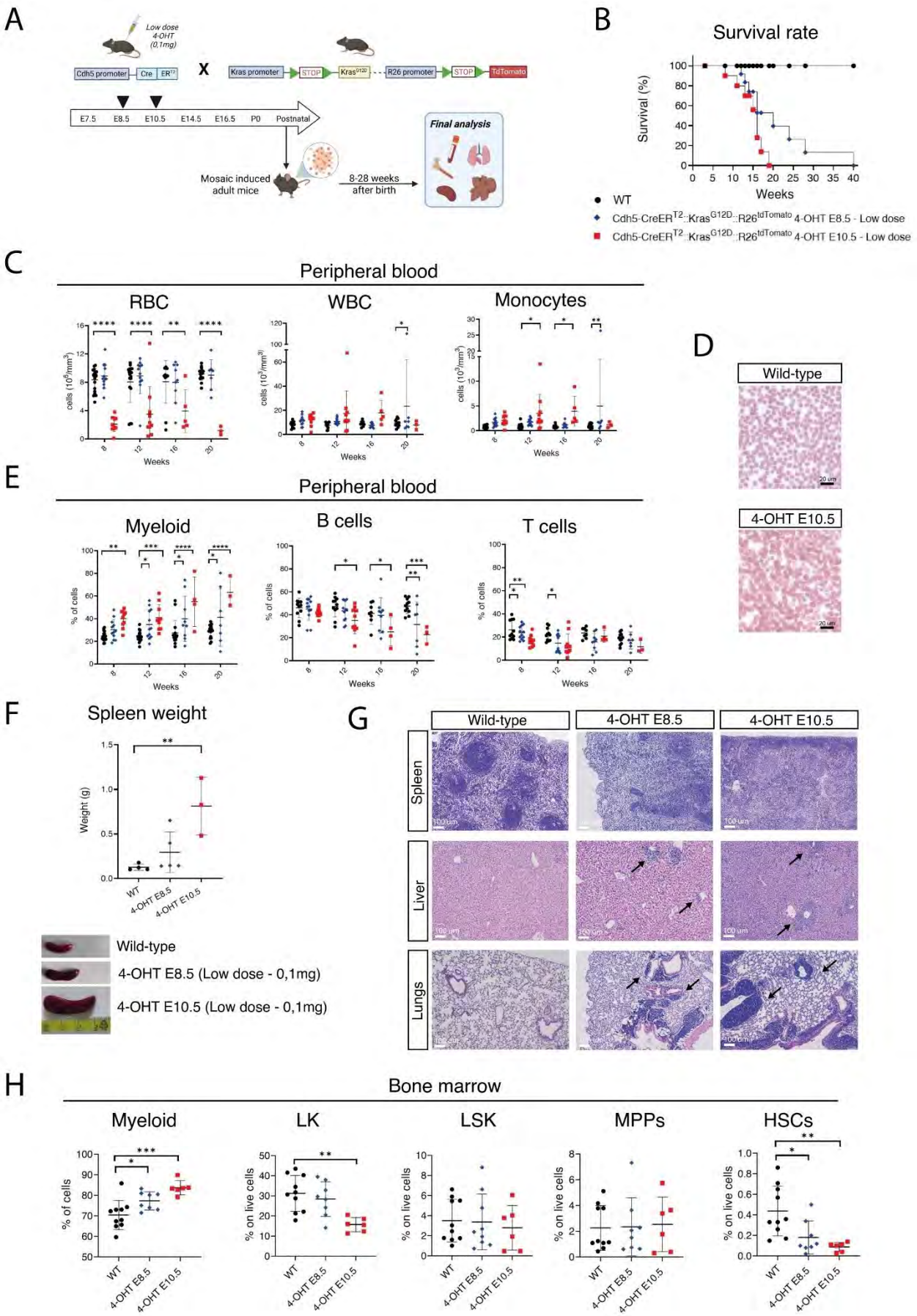

A
Low dose 4-OHT (0,1mg)
Cdh5 promoter
Cre
ER^T2
X
Kras promoter
STOP
Kras^G12D
R26 promoter
STOP
TdTomato
E7.5
E8.5
E10.5
E14.5
E16.5
P0
Postnatal
Mosaic induced adult mice
8-28 weeks after birth
Final analysis
B
Survival rate
Survival (%)
Weeks
WT
Cdh5-CreER^T2::Kras^G12D::R26^tdTomato 4-OHT E8.5 - Low dose
Cdh5-CreER^T2::Kras^G12D::R26^tdTomato 4-OHT E10.5 - Low dose
C
Peripheral blood
RBC
WBC
Monocytes
cells (10^6/mm^3)
cells (10^3/mm^3)
Weeks
D
Wild-type
4-OHT E10.5
E
Peripheral blood
Myeloid
B cells
T cells
% of cells
Weeks
F
Spleen weight
Weight (g)
WT
4-OHT E8.5
4-OHT E10.5
Wild-type
4-OHT E8.5 (Low dose - 0,1mg)
4-OHT E10.5 (Low dose - 0,1mg)
G
Wild-type
4-OHT E8.5
4-OHT E10.5
Spleen
Liver
Lungs
H
Bone marrow
Myeloid
LK
LSK
MPPs
HSCs
% of cells
% on live cells
WT
4-OHT E8.5
4-OHT E10.5

Figure 6: JMML-resembling phenotype in mosaic-induced Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} adult mice (4-OHT E8.5, E10.5 – Low dose)

(A) Schematic representation of experimental strategy applied for the mosaic activation of Kras^{G12D} in Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} adult mice

(B) Kaplan-Meier survival curve comparing WT (Cdh5-CreER^{T2}::::R26^{tdTomato} , n=11) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n=12) and E10.5 (n=6) with low dose of 4-OHT (0,1 mg)

(C) Longitudinal analysis of RBC, WBC and monocytes in WT (Cdh5-CreER^{T2}::::R26^{tdTomato}) (n=19, derived from 8 litters and analyzed individually in 8 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n= 12, derived from 6 litters and analyzed individually in 6 independent experiments) and E10.5 (n=11, derived from 3 litters and analyzed individually in 3 independent experiment). Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test).

(D) Blood smears from WT (Cdh5CreER^{T2}::R26^{TdTomato}) and TG (Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato}).

(E) Quantification of flow cytometric analysis of myeloid (CD45$^+$ CD11b$^+$), B cells (CD45$^+$ B220$^+$ CD3e$^-$) and T cells (CD45$^+$ B220$^-$ CD3e$^+$) on CD45$^+$ cells in PB of WT (Cdh5-CreER^{T2}::::R26^{tdTomato}) (n=19, derived from 8 litters and analyzed individually in 8 independent experiments) and Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato} activated at E8.5 (n= 12, derived from 6 litters and analyzed individually in 6 independent experiments) and E10.5 (n=11, derived from 3 litters and analyzed individually in 3 independent experiment). Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001 ; (two-way ANOVA followed by Tukey's multiple comparisons test).

(F) Gross morphology of spleen (top) and spleen weight (bottom) from Cdh5CreER^{T2}::R26^{TdTomato} (n=4) and Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} activated at E8.5 (n=5) and E10.5 (n=3). Error bars represent mean ± SD. ** p < 0.01 (two-tailed unpaired Student's t-test).

(G) Hematoxylin-eosin staining of spleen, liver and lungs of Cdh5CreER^{T2}::R26^{TdTomato} (WT) and Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} adult mice activated at E8.5 and E10.5. Arrowheads indicate the presence of pathological infiltrations in liver and lungs.

(H) Quantification of flow cytometric analysis of myeloid (CD45$^+$ CD11b$^+$), LK (Lineage$^-$ Kit$^+$ Sca1$^-$), LSK (Lineage$^-$ Kit$^+$ Sca1$^+$), MPPs (Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^+$ CD150$^-$) and HSCs Lineage$^-$ Kit$^+$ Sca1$^+$ CD48$^+$ CD150$^+$) in the BM of adult mice. WT (n= 10, derived from 6

litters and analyzed individually in 6 independent experiments), Cdh5CreER^{T2}::LSL-Kras^{G12D}::R26^{TdTomato} 4-OHT E8.5 (n= 8, derived from 5 litters and analyzed individually in 5 independent experiments) and 4-OHT E10.5 (n=6, derived from 3 litters and analyzed individually in 3 independent experiments. Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; *** p < 0.001; (one-way ANOVA followed by Tukey's multiple comparisons test).

Figure 7

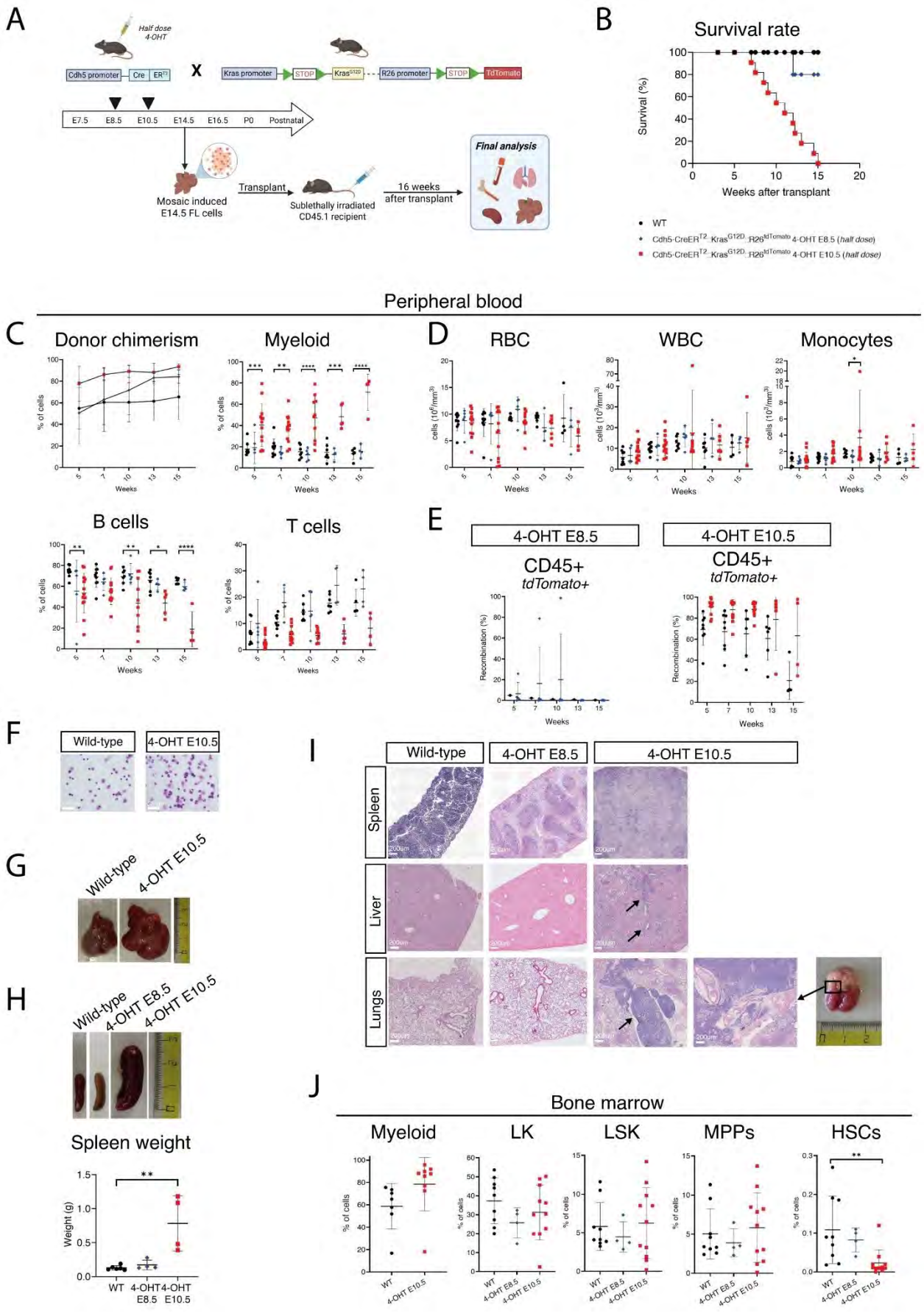

A
Half dose 4-OHT
Cdh5 promoter
Cre
ER^T1
X
Kras promoter
STOP
Kras^G12D
R26 promoter
STOP
TdTomato
E7.5
E8.5
E10.5
E14.5
E16.5
P0
Postnatal
Mosaic induced E14.5 FL cells
Transplant
Sublethally irradiated CD45.1 recipient
16 weeks after transplant
Final analysis
B
Survival rate
Survival (%)
Weeks after transplant
WT
Cdh5-CreER^T2:Kras^G12D:R26^tdTomato 4-OHT E8.5 (half dose)
Cdh5-CreER^T2:Kras^G12D:R26^tdTomato 4-OHT E10.5 (half dose)
Peripheral blood
C
Donor chimerism
Myeloid
B cells
T cells
% of cells
Weeks
D
RBC
WBC
Monocytes
cells (10^6/mm^3)
cells (10^3/mm^3)
Weeks
E
4-OHT E8.5
4-OHT E10.5
CD45+ tdTomato+
Recombination (%)
Weeks
F
Wild-type
4-OHT E10.5
G
Wild-type
4-OHT E10.5
H
Wild-type
4-OHT E8.5
4-OHT E10.5
Spleen weight
Weight (g)
WT
4-OHT E8.5
4-OHT E10.5
I
Wild-type
4-OHT E8.5
4-OHT E10.5
Spleen
Liver
Lungs
J
Bone marrow
Myeloid
LK
LSK
MPPs
HSCs
% of cells
WT
4-OHT E8.5
4-OHT E10.5

Figure 7: JMML-resembling phenotype with prenatal origin in adult mice transplanted with $Kras^{G12D}$-induced FL cells (4-OHT E8.5, E10.5 – Half dose)

(A) Schematic representation of experimental strategy applied for transplantation of $Kras^{G12D}$ -induced E14.5 FL cells in sublethally irradiated adult mice.

(B) Kaplan-Meier survival curve comparing adult mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$, n = 10) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 (n= 6) and E10.5 (n=13) with half dose of 4-OHT (0,5 mg)

(C) Quantification of flow cytometric analysis of donor chimerism (reported as $CD45.1^{-}$ $CD45.2^{+}$ cells on total $CD45^{+}$ cells), myeloid ($CD45^{+}$ $CD11b^{+}$), B cells ($CD45^{+}$ $B220^{+}$ $CD3e^{-}$) and T cells ($CD45^{+}$ $B220^{-}$ $CD3e^{+}$) on $CD45^{+}$ cells in PB of mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) (n=10, analyzed individually in 4 independent experiments) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 (n= 6, analyzed individually in 2 independent experiments) and E10.5 (n=13, analyzed individually in 4 independent experiment). Error bars represent mean ± SD. *p< 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test).

(D) Longitudinal analysis of RBC, WBC and monocytes in PB of mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) (n=10, analyzed individually in 4 independent experiments) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 (n= 6, analyzed individually in 2 independent experiments) and E10.5 (n=13, analyzed individually in 4 independent experiment). Error bars represent mean ± SD. *p< 0.05; (two-way ANOVA followed by Tukey's multiple comparisons test).

(E) Quantification of flow cytometric analysis of $TdTomato^{+}$ cells in $CD45^{+}$ cells in PB of mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) (n=10, analyzed individually in 4 independent experiments) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 (n= 6, analyzed individually in 2 independent experiments) and E10.5 (n=13, analyzed individually in 4 independent experiment). Error bars represent mean ± SD.

(F) Representative cytospin preparation of PB stained with May–Grünwald–Giemsa from mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E10.5.

(G) Gross morphology of liver from mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E10.5.

(H) Gross morphology of spleen (top) and spleen weight (bottom) from mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) and FL TG (Cdh5-$CreER^{T2}$::LSL-

$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 and E10.5. ** $p < 0.01$; (two-tailed unpaired Student's t-test).

(I) Hematoxylin-eosin staining of spleen, liver and lungs of mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 and E10.5. Arrowheads indicate the presence of pathological infiltrations in liver and lungs.

(J) Quantification of flow cytometric analysis of myeloid ($CD45^+$ $CD11b^+$), LK ($Lineage^-$ Kit^+ $Sca1^-$), LSK ($Lineage^-$ Kit^+ $Sca1^+$), MPPs ($Lineage^-$ Kit^+ $Sca1^+$ $CD48^+$ $CD150^-$) and HSCs $Lineage^-$ Kit^+ $Sca1^+$ $CD48^+$ $CD150^+$) in the BM of mice transplanted with FL WT (Cdh5-$CreER^{T2}$::::$R26^{tdTomato}$) (n=9, analyzed individually in 4 independent experiments) and FL TG (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{tdTomato}$) activated at E8.5 (n= 4, analyzed individually in 2 independent experiments) and E10.5 (n=11, analyzed individually in 4 independent experiment). Error bars represent mean ± SD. ** $p < 0.01$ (one-way ANOVA followed by Tukey's multiple comparisons test).

Figure 8

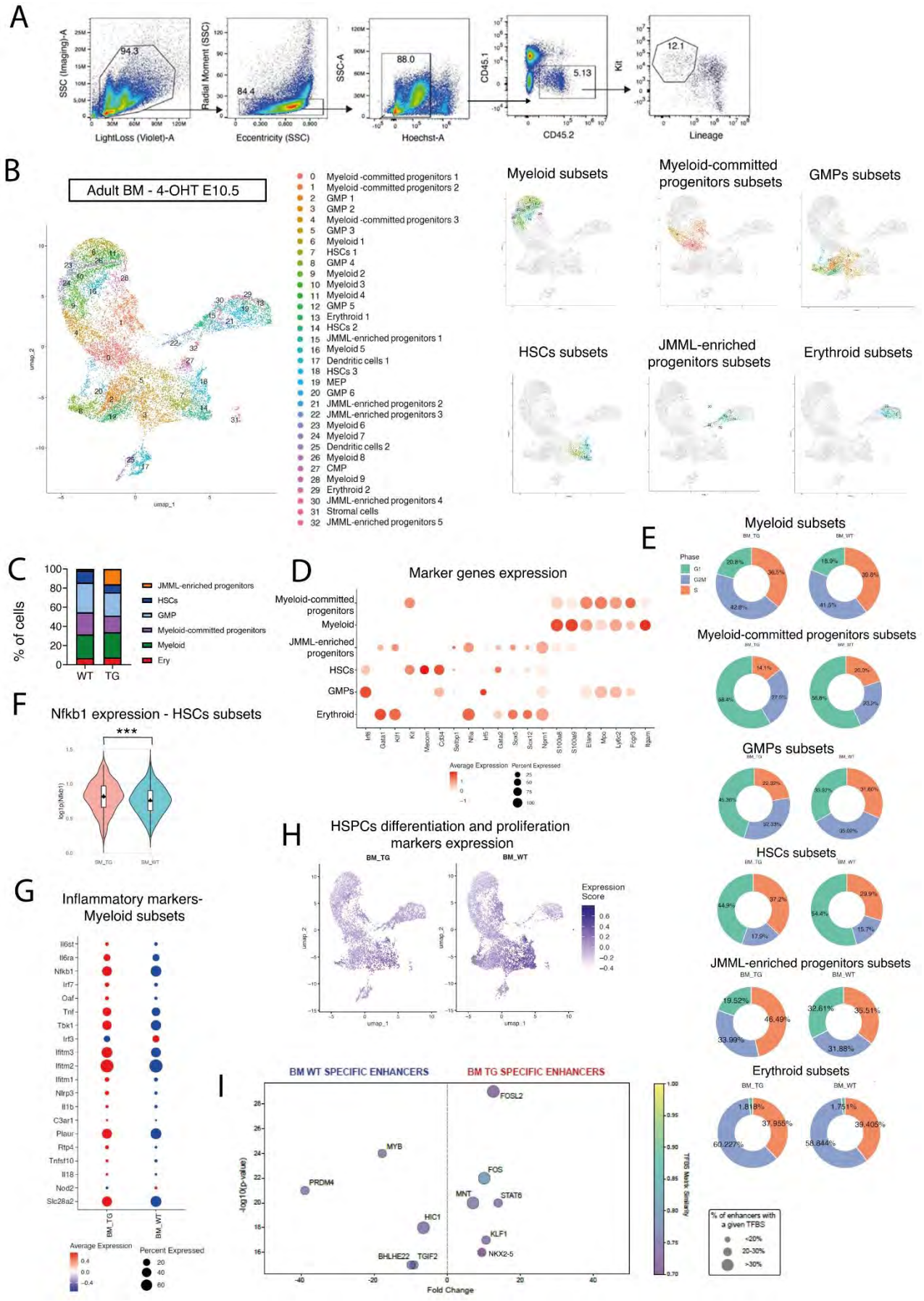

A
SSC (Imaging)-A
LightLoss (Violet)-A
94.3
Radial Moment (SSC)
Eccentricity (SSC)
84.4
SSC-A
Hoechst-A
88.0
CD45.1
CD45.2
5.13
Kit
Lineage
12.1
B
Adult BM - 4-OHT E10.5
0 Myeloid -committed progenitors 1
1 Myeloid -committed progenitors 2
2 GMP 1
3 GMP 2
4 Myeloid -committed progenitors 3
5 GMP 3
6 Myeloid 1
7 HSCs 1
8 GMP 4
9 Myeloid 2
10 Myeloid 3
11 Myeloid 4
12 GMP 5
13 Erythroid 1
14 HSCs 2
15 JMML-enriched progenitors 1
16 Myeloid 5
17 Dendritic cells 1
18 HSCs 3
19 MEP
20 GMP 6
21 JMML-enriched progenitors 2
22 JMML-enriched progenitors 3
23 Myeloid 6
24 Myeloid 7
25 Dendritic cells 2
26 Myeloid 8
27 CMP
28 Myeloid 9
29 Erythroid 2
30 JMML-enriched progenitors 4
31 Stromal cells
32 JMML-enriched progenitors 5
umap_1
umap_2
Myeloid subsets
Myeloid-committed progenitors subsets
GMPs subsets
HSCs subsets
JMML-enriched progenitors subsets
Erythroid subsets
C
% of cells
100
80
60
40
20
0
WT
TG
JMML-enriched progenitors
HSCs
GMP
Myeloid-committed progenitors
Myeloid
Ery
D
Marker genes expression
Myeloid-committed progenitors
Myeloid
JMML-enriched progenitors
HSCs
GMPs
Erythroid
Irf8
Gata1
Kit
Meoom
Cd34
Selpg1
Nfia
Irf5
Gata2
Sox5
Sox12
Npm1
S100a8
S100a9
Elane
Mpo
Ly6c2
Fcgr3
Itgam
Average Expression
Percent Expressed
E
Myeloid subsets
Phase
G1
G2M
S
BM_TG
BM_WT
20.8%
36.5%
42.8%
18.9%
39.6%
41.5%
Myeloid-committed progenitors subsets
14.1%
58.4%
27.5%
26.0%
56.8%
23.2%
GMPs subsets
22.33%
45.36%
32.33%
33.67%
31.60%
35.02%
HSCs subsets
37.2%
44.9%
17.9%
29.9%
54.4%
15.7%
JMML-enriched progenitors subsets
19.52%
46.49%
33.99%
32.61%
35.51%
31.88%
Erythroid subsets
1.818%
37.955%
60.227%
1.751%
39.405%
58.844%
F
Nfkb1 expression - HSCs subsets
***
log1p(Nfkb1)
BM_TG
BM_WT
G
Inflammatory markers-
Myeloid subsets
Il6st
Il6ra
Nfkb1
Irf7
Oaf
Tnf
Tbk1
Irf3
Ifitm3
Ifitm2
Ifitm1
Nlrp3
Il1b
C3ar1
Plaur
Rtp4
Tnfsf10
Il18
Nod2
Slc28a2
BM_TG
BM_WT
Average Expression
Percent Expressed
H
HSPCs differentiation and proliferation
markers expression
BM_TG
BM_WT
Expression
Score
umap_1
umap_2
I
BM WT SPECIFIC ENHANCERS
BM TG SPECIFIC ENHANCERS
-log10(p-value)
Fold Change
FOSL2
MYB
FOS
PRDM4
MNT
STAT6
HIC1
KLF1
BHLHE22
TGIF2
NKX2-5
TFBS Matrix Similarity
% of enhancers with
a given TFBS
<20%
20-30%
>30%

Figure 8: scRNAseq of adult BM cells isolated from FL-transplanted mice (4-OHT E10.5)

(A) Representative flow cytometric gating strategy for the sorting of live Lineage$^-$ CD45.1$^-$ CD45.2$^+$ Kit$^+$ cells for scRNAseq analysis of BM cells cells from adult mice transplanted with WT (Cdh5-CreER^{T2}::R26^{tdTomato}) and TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}) E14.5 FL cells.

(B) UMAP of 14568 cells isolated as displayed in (A), from BM of 2 mice transplanted with WT (Cdh5-CreER^{T2}::R26^{tdTomato}) FL cells (8408) and BM of 2 mice transplante with TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}) FL (6160) activated at E10.5. BM WT and BM TG cells were separately sequenced. Cells are colored according to individual cluster identities. UMAPs on the right highlight myeloid subsets (clusters 6, 9, 10, 11, 23, 24, 26, 28), myeloid-committed progenitors subsets (clusters 0, 1, 4, 16), GMP subsets (2, 3, 5, 8, 12, 20), JMML-enriched progenitors subsets (clusters 15, 21, 22, 30, 32), HSCs subsets (clusters 7, 14, 18) and erythroid subsets (clusters 13, 19, 29).

(C) Proportion of different cell types in WT and TG samples. Percentages were calculated relative to the total number of cells sequenced per sample.

(D) Bubble plot showing the relative expression of key hematopoietic genes among different cellular subsets.

(E) Donut plots reporting the percentage of cells in each phase of the cell cycle (G1, S, G2/M) for myeloid, erythroid and HSPCs subset for WT and TG samples. Cell cycle phases were inferred from single-cell transcriptomic data using canonical marker genes.

(F) Violin plot of the distribution of Nfkb1 expression in HSCs subset of BM TG and BM

(G) WT, calculated by pseudobulk analysis. **** $p < 0.001$ (Wilcoxon rank-sum test)

(H) Bubble plot showing the relative expression of key inflammatory marker genes among BM WT and BM TG of myeloid subsets.

(I) UMAP highlighting the expression of a panel of key regulatory HSPCs genes (Myc, Ccnd1, Cdk4, Cdk6, Kit, Flt3, Gata2, Hoxa9, Runx1, Notch1, Jak2, Bcl2, Mpl, Socs3, Cxcr4).

(J) TFBS enrichment analysis in the aberrantly inactive and active enhancers in BM WT and BM TG. The color scale represents the similarity of the sequences found in these enhancers with respect to specific TFBS matrixes. The dot sizes indicate the percentage of aberrantly active or inactive enhancers which harbor that TFBS.

Figure 9

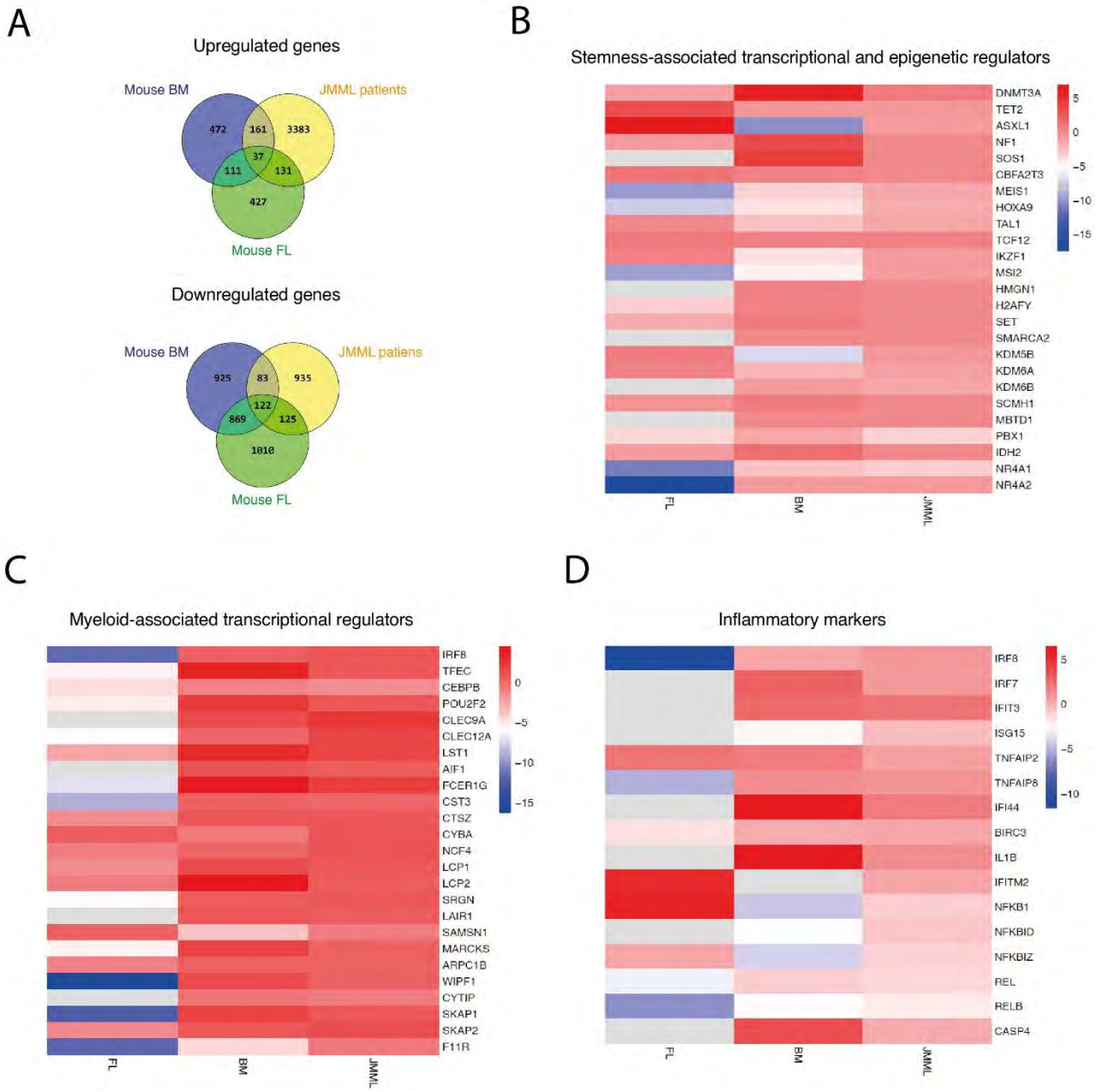


Figure 9: Comparative analysis of murine and human JMML HSPCs

(A) Venn diagrams showing overlap of upregulated and downregulated genes between murine bone marrow (BM), fetal liver (FL), and human JMML HSPCs, derived from pediatric BM of two JMML patients[16].

(B) Heatmap of stemness-associated transcriptional and epigenetic regulators in BM and FL murine cells, and human HSPCs. The heatmap reports the average fold change of each sample relative to its respective control.

(C) Heatmap of myeloid lineage-associated genes in BM and FL murine cells, and human HSPCs. The heatmap reports the average fold change of each sample relative to its respective control.

(D) Heatmap of inflammatory genes in BM and FL murine cells, and human HSPCs. The heatmap reports the average fold change of each sample relative to its respective control.

Figure S1

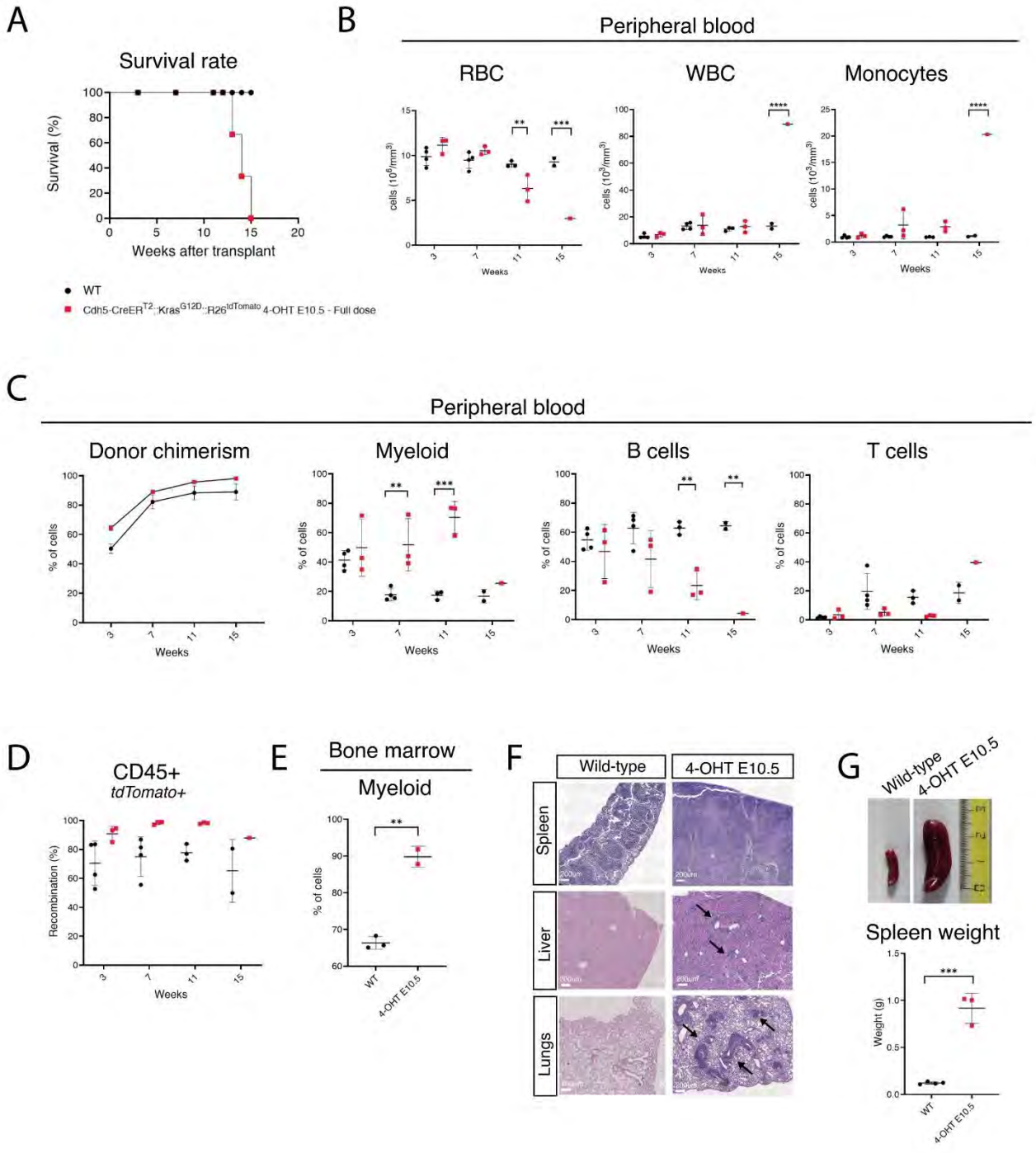


Figure S1: Aggressive JMML-like disease in adult mice transplanted with Kras^{G12D}-induce d FL cells (4-OHT E10.5, full dose)

(A) Kaplan-Meier curve comparing survival of adult mice transplanted with FL WT (Cdh5-CreER^{T2}::::R26^{tdTomato} , n = 4) and FL TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}, n=3)

(B) Longitudinal analysis of RBC, WBC and monocytes in PB of mice transplanted with FL WT (Cdh5-CreER^{T2}::::R26^{tdTomato}, n=4, analyzed in one experiment) and FL TG (Cdh5-

CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}, n=3, analyzed in one experiment). Error bars represent mean ± SD. ** p < 0.01; *** p < 0.001; **** p < 0.0001; (two-way ANOVA followed by Tukey's multiple comparisons test).

(C) Quantification of flow cytometric analysis of donor chimerism (reported as CD45.1^{-} CD45.2^{+} cells on total CD45^{+} cells), myeloid (CD45^{+} CD11b^{+}), B cells (CD45^{+} B220^{+} CD3e^{-}) and T cells (CD45^{+} B220^{-} CD3e^{+}) on CD45^{+} cells in PB of mice transplanted with FL WT (Cdh5-CreER^{T2}::::R26^{tdTomato}, n=4, analyzed in one experiment) and FL TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}, n=3, analyzed in one experiment). ** p < 0.01; *** p < 0.001; (two-way ANOVA followed by Tukey's multiple comparisons test).

(D) Quantification of flow cytometric analysis of TdTomato^{+} cells in CD45^{+} cells in PB of mice transplanted with FL WT (Cdh5-CreER^{T2}::::R26^{tdTomato}, n=4, analyzed in one experiment) and FL TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}, n=3, analyzed in one experiment)

(E) Quantification of flow cytometric analysis of myeloid (CD45^{+} CD11b^{+}) in the BM of transplanted mice (WT n=4, TG n=3, analyzed in one experiment) ** p < 0.01; (two-tailed unpaired Student's t-test).

(F) Hematoxylin-eosin staining of spleen, liver and lungs of mice transplanted with FL WT (Cdh5-CreER^{T2}::::R26^{tdTomato}) and FL TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}). Arrowheads indicate the presence of pathological infiltrations in liver and lungs.

(G) Gross morphology of spleen (top) and spleen weight (bottom) from mice transplanted with FL WT (Cdh5-CreER^{T2}::::R26^{tdTomato}) and FL TG (Cdh5-CreER^{T2}::LSL-Kras^{G12D}::R26^{tdTomato}). *** p < 0.001; (two-tailed unpaired Student's t-test).

References

1. Locatelli, F. & Niemeyer, C. M. How I treat juvenile myelomonocytic leukemia. *Blood* 125, 1083–1090 (2015).

2. Chronic Myelomonocytic Leukemia in Childhood: A Retrospective Analysis of 110 Cases. *Blood* 89, 3534–3543 (1997).

3. Meynier, S. & Rieux-Laucat, F. After 95 years, it's time to eRASe JMML. *Blood Reviews* vol. 43 Preprint at https://doi.org/10.1016/j.blre.2020.100652 (2020).

4. Niemeyer, C. M. & Flotho, C. *Review Series MYELODYSPLASTIC SYNDROMES Juvenile Myelomonocytic Leukemia: Who's the Driver at the Wheel?* (2019).

5. Rudelius, M., Weinberg, O. K., Niemeyer, C. M., Shimamura, A. & Calvo, K. R. The International Consensus Classification (ICC) of hematologic neoplasms with germline predisposition, pediatric myelodysplastic syndrome, and juvenile myelomonocytic leukemia. *Virchows Archiv* 482, 113–130 (2023).

6. Wintering, A., Dvorak, C. C., Stieglitz, E. & Loh, M. L. Juvenile myelomonocytic leukemia in the molecular era: a clinician's guide to diagnosis, risk stratification, and treatment. *Blood Adv.* 5, 4783–4793 (2021).

7. Locatelli, F. *et al.* Hematopoietic stem cell transplantation (HSCT) in children with juvenile myelomonocytic leukemia (JMML): results of the EWOG-MDS/EBMT trial. *Blood* 105, 410–419 (2005).

8. Stieglitz, E. *et al.* The genomic landscape of juvenile myelomonocytic leukemia. *Nat. Genet.* 47, 1326–1333 (2015).

9. Fiñana, C., Gómez-Molina, N., Alonso-Moreno, S. & Belver, L. Genomic and Epigenomic Landscape of Juvenile Myelomonocytic Leukemia. *Cancers* vol. 14 Preprint at https://doi.org/10.3390/cancers14051335 (2022).

10. Murakami, N. *et al. Integrated Molecular Profiling of Juvenile Myelomonocytic Leukemia*. http://genomon.hgc.jp/exome/ (2018).

11. Caye, A. *et al.* Juvenile myelomonocytic leukemia displays mutations in components of the RAS pathway and the PRC2 network. *Nat. Genet.* 47, 1334–1340 (2015).

12. Sakaguchi, H. *et al.* Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. *Nat. Genet.* 45, 937–941 (2013).

13. Wakamatsu, M. *et al.* Detection of subclonal SETBP1 and JAK3 mutations in juvenile myelomonocytic leukemia using droplet digital PCR. *Leukemia* 35, 259–263 (2021).

14. Strullu, M. *et al.* Juvenile myelomonocytic leukaemia and Noonan syndrome. *J. Med. Genet.* 51, 689–697 (2014).

15. Behnert, A. *et al.* Exploring the genetic and epigenetic origins of juvenile myelomonocytic leukemia using newborn screening samples. *Leukemia* 36, 279 (2021).

16. Louka, E. *et al.* Heterogeneous disease-propagating stem cells in juvenile myelomonocytic leukemia. *Journal of Experimental Medicine* 218, (2021).

17. Caye, A. *et al.* Despite mutation acquisition in hematopoietic stem cells, JMML-propagating cells are not always restricted to this compartment. *Leukemia* 34, 1658 (2019).

18. Barone, C., Orsenigo, R., Meneveri, R., Brunelli, S. & Azzoni, E. One Size Does Not Fit All: Heterogeneity in Developmental Hematopoiesis. *Cells* 11, 1061 (2022).

19. Barone, C. *et al.* Hemogenic endothelium of the vitelline and umbilical arteries is the major contributor to mouse fetal lympho-myelopoiesis. *bioRxiv* 2024.07.11.603050 (2024) doi:10.1101/2024.07.11.603050.

20. Cani, A. *et al.* Long-term proliferation of immature hypoxia-dependent JMML cells supported by a 3D in vitro system. *Blood Adv.* 7, 1513–1524 (2023).

21. Gandre-Babbe, S. *et al.* Patient-derived induced pluripotent stem cells recapitulate hematopoietic abnormalities of juvenile myelomonocytic leukemia. *Blood* 121, 4925–4929 (2013).

22. Tarnawsky, S. P., Kobayashi, M., Chan, R. J. & Yode, M. C. Mice expressing KrasG12D in hematopoietic multipotent progenitor cells develop neonatal myeloid leukemia. *J. Clin. Invest.* 127, 3652–3656 (2017).

23. Jang, G. *et al.* Leukemogenic Kras mutation reprograms multipotent progenitors to facilitate its spread through the hematopoietic system. *J. Exp. Med.* 222, (2025).

24. Rogers, J. H. *et al.* Dose-dependent effects of Dnmt3a in an inducible murine model of KrasG12D-driven leukemia. *Exp. Hematol.* 135, 104248 (2024).

25. Alva, J. A. *et al.* VE-cadherin-cre-recombinase transgenic mouse: A tool for lineage analysis and gene deletion in endothelial cells. *Developmental Dynamics* 235, 759–767 (2006).

26. Dzierzak, E. & Bigas, A. Blood Development: Hematopoietic Stem Cell Dependence and Independence. *Cell Stem Cell* 22, 639–651 (2018).

27. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* 15, 3243–3248 (2001).

28. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* 13, 133–140 (2010).

29. Solman, M. *et al.* Inflammatory response in hematopoietic stem and progenitor cells triggered by activating SHP2 mutations evokes blood defects. *Elife* 11, (2022).

30. Barone, C. *et al.* Hemogenic endothelium of the vitelline and umbilical arteries is the major contributor to mouse fetal lympho-myelopoiesis. *bioRxiv* 2024.07.11.603050 (2024) doi:10.1101/2024.07.11.603050.

31. Gupta, A. K., Meena, J. P., Chopra, A., Tanwar, P. & Seth, R. Juvenile myelomonocytic leukemia-A comprehensive review and recent advances in management. *Am. J. Blood Res.* 11, 1 (2021).

32. Zheng, H. *et al.* Innate immune mechanisms hijacked by leukemia-initiating stem cells for selective advantage and immune evasion in Ptpn11-associated JMML. *Cell Rep.* 44, (2025).

33. Solman, M. *et al.* Inflammatory response in hematopoietic stem and progenitor cells triggered by activating SHP2 mutations evokes blood defects. *Elife* 11, (2022).

34. Yan, Y. *et al.* JMML tumor cells disrupt normal hematopoietic stem cells by imposing inflammatory stress through overproduction of IL-1β. *Blood Adv.* 6, 200–206 (2022).

35. Hu, N., Qiu, Y. & Dong, F. Role of Erk1/2 signaling in the regulation of neutrophil versus monocyte development in response to G-CSF and M-CSF. *Journal of Biological Chemistry* 290, 24561–24573 (2015).

36. Huber, R. *et al.* Regulation of monocyte differentiation by specific signaling modules and associated transcription factor networks. *Cell. Mol. Life Sci.* 71, 63 (2013).

37. Yu, X., Wang, Y., Song, Y., Gao, X. & Deng, H. AP-1 is a regulatory transcription factor of inflammaging in the murine kidney and liver. *Aging Cell* 22, (2023).

38. Renoux, F. *et al.* The AP1 Transcription Factor Fosl2 Promotes Systemic Autoimmunity and Inflammation by Repressing Treg Development. *Cell Rep.* 31, (2020).

39. Yokomizo, T. & Suda, T. Development of the hematopoietic system: expanding the concept of hematopoietic stem cell-independent hematopoiesis. *Trends Cell Biol.* 34, 161–172 (2024).

40. Yan, Y. *et al.* JMML tumor cells disrupt normal hematopoietic stem cells by imposing inflammatory stress through overproduction of IL-1b. *Blood Adv.* 6, 200–206 (2022).

41. Sango, J. *et al.* RAS-mutant leukaemia stem cells drive clinical resistance to venetoclax. *Nature* 636, 241–250 (2024).

42. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* 184, 3573-3587.e29 (2021).

43. Noguchi, S. *et al.* FANTOM5 CAGE profiles of human and mouse samples. *Sci. Data* 4, 1–10 (2017).

cancers

MDPI

Article

Hematopoietic Stem Cell (HSC)-Independent Progenitors Are Susceptible to Mll-Af9-Induced Leukemic Transformation

Cristiana Barone [1,†], Roberto Orsenigo [1,†,‡], Anna Cazzola [1,§], Elisabetta D'Errico [1], Arianna Patelli [1], Giulia Quattrini [1], Barbara Vergani [1], Silvia Bombelli [1], Sofia De Marco [1], Cristina D'Orlando [1], Cristina Bianchi [1], Biagio Eugenio Leone [1], Raffaella Meneveri [1], Andrea Biondi [1,2], Giovanni Cazzaniga [1,3], Terence Howard Rabbitts [4], Silvia Brunelli [1] and Emanuele Azzoni [1,*]

[1] School of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy; cristiana.barone@unimib.it (C.B.); roberto.orsenigo@vhir.org (R.O.)
[2] Pediatrics, Fondazione IRCCS San Gerardo dei Tintori, 20900 Monza, Italy
[3] Centro Tettamanti, IRCCS San Gerardo dei Tintori, 20900 Monza, Italy
[4] Division of Cancer Therapeutics, Institute of Cancer Research, 15 Cotswold Road, Sutton, London SM2 5NG, UK
* Correspondence: emanuele.azzoni@unimib.it
† These authors contributed equally to this work.
‡ Current address: Biomedical Research in Melanoma-Animal Models and Cancer Laboratory, Vall d'Hebron Research Institute (VHIR), Vall d'Hebron Hospital Barcelona-UAB, 08035 Barcelona, Spain.
§ Current address: Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark.

Simple Summary: Despite important advances in the understanding of the genetics of infant acute myeloid leukemia (AML), it is still unclear why this blood cancer has features that are distinct to adult AML. Infant AML can often have prenatal origins, but the cell of origin has not been conclusively determined. Using transgenic mice, we found that the introduction of the Mll-Af9 inter-chromosomal rearrangement in hematopoietic stem cell (HSC)-independent progenitors results in a transplantable myeloid leukemia. Thus, we hypothesize that the peculiar characteristics of embryonic hematopoiesis might play a key role in determining the disease phenotype and leukemic progression.

Abstract: Infant acute myeloid leukemia (AML) is a heterogeneous disease, genetically distinct from its adult counterpart. Chromosomal translocations involving the *KMT2A* gene (*MLL*) are especially common in affected infants of less than 1 year of age, and are associated with a dismal prognosis. While these rearrangements are likely to arise in utero, the cell of origin has not been conclusively identified. This knowledge could lead to a better understanding of the biology of the disease and support the identification of new therapeutic vulnerabilities. Over the last few years, important progress in understanding the dynamics of fetal hematopoiesis has been made. Several reports have highlighted how hematopoietic stem cells (HSC) provide little contribution to fetal hematopoiesis, which is instead largely sustained by HSC-independent progenitors. Here, we used conditional Cre-Lox transgenic mouse models to engineer the Mll-Af9 translocation in defined subsets of embryonic hematopoietic progenitors. We show that embryonic hematopoiesis is generally permissive for Mll-Af9-induced leukemic transformation. Surprisingly, the selective introduction of Mll-Af9 in HSC-independent progenitors generated a transplantable myeloid leukemia, whereas it did not when introduced in embryonic HSC-derived cells. Ex vivo engineering of the Mll-Af9 rearrangement in HSC-independent progenitors using a CRISPR/Cas9-based approach resulted in the activation of an aberrant myeloid-biased self-renewal program. Overall, our results demonstrate that HSC-independent hematopoietic progenitors represent a permissive environment for Mll-Af9-induced leukemic transformation, and can likely act as cells of origin of infant AML.

Keywords: infant AML; Mll-Af9; hematopoiesis; hematopoietic stem cell; erythro-myeloid progenitors; cell of origin; CRISPR/Cas9

Citation: Barone, C.; Orsenigo, R.; Cazzola, A.; D'Errico, E.; Patelli, A.; Quattrini, G.; Vergani, B.; Bombelli, S.; De Marco, S.; D'Orlando, C.; et al. Hematopoietic Stem Cell (HSC)-Independent Progenitors Are Susceptible to Mll-Af9-Induced Leukemic Transformation. *Cancers* **2023**, *15*, 3624. https://doi.org/10.3390/cancers15143624

Academic Editors: Yvonne A. Efebera and Nidhi Sharma

Received: 23 May 2023
Revised: 7 July 2023
Accepted: 10 July 2023
Published: 14 July 2023

1. Introduction

Infant leukemias (defined as those with an onset taking place before 12 months of age) are hematological diseases displaying unique characteristics [1,2]. Whereas leukemias are the most common type of childhood cancer, during infancy they are relatively less frequent (5 versus 1.6–1.7 per 100,000) [3,4]. Clinically, childhood and infant leukemias exhibit different features and the latter are usually characterized by a worse prognosis, largely due to resistance to conventional chemotherapy and the lack of targeted approaches [1,3,5]. In childhood, acute lymphoblastic leukemia (ALL) occurs with a significantly higher frequency than acute myeloid leukemia (AML), whereas the incidence rates of ALL and AML are similar in infants [3].

The genetic alterations found in infant leukemia are different to those generally occurring in older children [3,6]. In particular, rearrangements involving the *MLL* (or *KMT2A*) gene, encoding for the histone lysine methyltransferase 2A, are detected in the majority of infant leukemia patients, and specifically in around 75% of ALL and 50% of AML [7,8]. Mutations in genes involved in the RAS pathway, often found as subclonal events, are also more frequent in infants than in children [3,9]. Generally, infant leukemias carry a small number of genetic lesions and MLL rearrangements are known to act as genetic drivers [6,9]. Interestingly, the same specific rearrangements can be found in both infant ALL and AML. One such example is the Mll-Af9 (t9:11) fusion oncogene, which in infants can result in acute leukemias of either myeloid or lymphoid lineage, whereas in adults the same rearrangement is almost invariably associated with AML [10]. It has been suggested that the cell of origin [11] and/or the hematopoietic microenvironment [12] can affect the outcome of Mll-Af9-induced transformation, and therefore dictate the leukemic phenotype and properties. However, a full understanding of the developmental stage-specific origin of infant leukemia is still missing [13]. Obtaining such knowledge will be critical in order to identify fetal specific cellular and/or environmental clues affecting leukemic biology, and in turn will help in developing better preclinical models and will lead to the identification of targetable vulnerabilities.

The large majority of infant leukemias arise in utero [4,7]. Several efforts have been made in the preclinical modeling of Mll-rearranged leukemias with prenatal origin, and the general consensus is that embryonic-derived leukemias have distinct features compared to those initiated postnatally [3]. When the Mll-Af9 translocation was introduced into fetal hematopoietic stem cells (HSCs) with a knockin model, fetal HSCs showed an increased latency compared to bone marrow HSCs [14]. However, these effects may be context- and model-dependent, as human cord blood CD34+ hematopoietic stem/progenitor cells (HSPCs) were instead easier to immortalize than adult HSPCs [11]. Notably, in both humans and mice, the introduction of Mll-Af9 in fetal HSPCs yielded leukemias of both myeloid and lymphoid lineages, paralleling what happens in patients [11,14]. Other translocations seen in infant AML, such as Mll-Enl, behave more aggressively when induced in fetal or neonatal HSPCs [15,16].

Thus, a relatively new concept of "layered leukemogenicity" has emerged over the last few years, which is trying to link the age-dependent features of leukemias with the peculiar characteristics of embryonic and fetal hematopoietic development [13]. Embryonic hematopoiesis is a remarkably complex process which takes place in several consecutive waves [17,18], and our understanding of it is still incomplete. In recent years, a new paradigm has emerged in the field, highlighting how HSCs do not primarily contribute to fetal hematopoiesis, which is instead primarily sustained by HSC-independent progenitors [19–21]. The fetal liver (FL) has long been recognized as the major hematopoietic organ during the second half of gestation, and it is perceived as an important site of HSC expansion [22]. However, a recent study challenged this long-standing view by showing that HSCs fated to adult life do not significantly expand in the FL but are instead kept quiescent [23], which is in agreement with HSC-independent progenitors playing a major role in pre-natal hematopoiesis. At the same time, the FL microenvironment can function as a hematopoietic differentiation hub for fetal-specific progenitors which emerge in other

locations as a "predefined" hierarchy [21,24]. These results evidence a previously unrecognized complexity in FL hematopoiesis, which, as a result, can impact our understanding of leukemias with prenatal origin, and, consequently, their modeling.

Most experimental strategies used so far to model fetal-derived leukemias have considered FL HSPCs "as a whole"; however, HSPCs in the mid-gestation FL are highly heterogeneous and can have various origins and fates. Therefore, in order to generate faithful preclinical models of childhood and infant leukemias, we need different approaches which should take into account the complexity of embryonic and fetal hematopoiesis. To date, it is unclear whether HSC-independent progenitors may be permissive for Mll rearrangement-mediated leukemic transformation, and thus could act as cells of origin for infant AML. Here, we aimed at gaining a better understanding of the ontogeny of infant Mll-Af9-rearranged leukemia. To this scope, we employed different genetic strategies in order to introduce the Mll-Af9 translocation in defined subsets of embryonic HSPCs, including HSC-independent progenitors.

2. Materials and Methods

2.1. Mouse Strains

All mouse strains used in this study have been previously described. To obtain mice bearing the Mll-Af9 translocation under constitutive expression of the Cre recombinase in embryonic hematopoietic progenitors, *Tie2-Cre* transgenic mice [25] were crossed with Mll^{loxP}::$Af9^{loxP}$ "translocator" mice [26,27]. For the generation of a mouse model in which the Mll-Af9 translocation could be induced at specific time-points in embryogenesis, we used *Cdh5-CreER*^{T2} mice [28–30]. For the induction of the Mll-Af9 translocation using the CRISPR/Cas9 system, *Cdh5-CreER*^{T2} mice were crossed with *Rosa26-LSL-Cas9-EGFP* knockin mice [31]. For lineage tracing experiments, *Cdh5-CreER*^{T2} mice were crossed with *Rosa26-LSL-tdTomato* or *Rosa26-LSL-zsGreen* mice [32].

To define the gestational age of mouse embryos, mice were mated in the evening and the presence of a vaginal plug the following morning was defined as embryonic day(E)0.5. Time-specific Cre induction was achieved through an intra-peritoneal (i.p.) injection of 4-hydroxytamoxifen (4-OHT) (37.5 mg/kg) or tamoxifen (75 mg/kg).

All transgenic mouse lines were maintained on a CD45.2 C57BL/6 genetic background, with the exception of females used for timed mating in order to generate adult mice with 4-OHT activation during embryogenesis, which were instead of a C57BL/6/FVB mixed background (F1). For transplantation experiments, CD45.1 C57BL/6 syngeneic mice were used as recipients. In some experiments, donor cells were CD45.1/CD45.2 heterozygotes. Mice were housed with free access to food and water at the San Raffaele Scientific Institute institutional mouse facilities. All experiments were performed in accordance with experimental protocols approved by the local Institutional Animal Care and Use Committees (IACUC).

2.2. Polymerase Chain Reaction (PCR) for Mll-Af9 Detection and Sequencing

The detection of the Mll-Af9 translocation induced with the Cre-loxP system was performed via nested PCR to increase the amplification specificity, as previously reported [26]. DNA was obtained from the fetal liver (FL) or adult bone marrow (BM) of transgenic and wild type embryos. For both the Cre-loxP- and CRISPR/Cas9-induced translocation, a HotStarTaq Master mix kit (Qiagen, Hilden, Germany) was used. PCR primers used for Mll-Af9 detection are reported in Table S1.

Confirmation of the correct inter-chromosomal translocation after CRISPR/Cas9 editing was achieved through Sanger sequencing.

2.3. Flow Cytometry and Hematological Analysis

Flow cytometry analysis was performed as previously described [33]. Bleedings were performed from the mouse tail vein every 8 weeks (unperturbed model) or every 4 weeks (transplants). BM cells were extracted from the femur and tibia. Briefly, samples were

stained and analyzed in CMF (Calcium/Magnesium-Free PBS, 10% FBS, 5% Pen/Strep, 2 mM EDTA), incubated for 15 min with 0.5 mg/mL Fc block (1:500) (BD Biosciences, Franklin Lakes, NJ, USA) and labeled with a combination of PE-conjugated rat anti-CD45 (1:200) (BD Biosciences), PE-conjugated rat anti-CD45.1 (1:200) (BD Biosciences), FITC-conjugated rat anti-CD45.2 (1:400) (BD Biosciences), PE-Cy7-conjugated rat anti-CD11b (1:200) (BD Biosciences), APC-Cy7-conjugated rat anti-B220 (1:200) (BioLegend, San Diego, CA, USA), PE-Cy5-conjugated rat anti-CD3e (1:100) (BioLegend), and APC-conjugated rat anti-Gr1 (1:200) (BioLegend). Appropriate fluorescence gating parameters were established with compensation beads (BD Biosciences), unstained, and fluorescence-minus-one (FMO) staining. In all experiments, doublets were gated out using pulse geometry gates (FSC-H versus FSC-A and SSC-H versus SSC-A), whereas dead cells were excluded using Hoechst 33258 (Hellobio, Bristol, UK). Single-cell suspensions were analyzed using a Fortessa X-20 analyzer (BD Biosciences). FlowJo software v10 (FlowJo/BD, Ashland, OR, USA) was used for subsequent analyses. For hematological analyses, peripheral blood was analyzed using an ABX Pentra 60 C+ analyzer (HORIBA Medical, Kyoto, Japan).

2.4. Primary and Secondary Transplants

For transplantation experiments, syngeneic C57BL/6 (CD45.1) recipient mice were lethally irradiated (9 Gy, split dose) before the intra-venous transplantation of 2×10^6 unfractionated E14.5 FL (primary) or adult BM cells (secondary) from *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* (4-OHT at E8.5 or E10.5). Donor-derived chimerism and hematological parameters within donor cells were determined using flow cytometry in peripheral blood (PB) every 4 weeks post-transplantation, and in PB and BM at 8 (*Tie2-Cre::Mll^{loxP}::Af9^{loxP}*), 28 (*Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* primary transplants), and 12 (*Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* secondary transplants) weeks post-transplantation.

For the transplantation experiments of CRISPR/Cas9-edited cells, C57BL/6 (CD45.1) recipient mice were sub-lethally irradiated (4.5 Gy, single dose) before the intra-venous transplantation of 2–4 $\times$ 10^5 E16.5 CD45+ c-Kit+ GFP+ FL cells from *Cdh5-CreER^{T2}::R26-LSL-Cas9-GFP* embryos (4-OHT at E8.5).

2.5. Histological Analysis

The spleen, liver, and lungs were collected from wild type and leukemic mice. Tissue samples were fixed in a 4% solution of paraformaldehyde (PFA) in PBS, dehydrated with ethanol, cleared with xylenes, and embedded in paraffin. Then, 5 mm organ sections were cut on a rotary microtome and stained with hematoxylin and eosin to assess the presence of extramedullary leukemic infiltrates. The size and weight of spleens were also recorded.

2.6. Fluorescence-Activated Cell Sorting (FACS) for the Isolation of HSC-Independent Hematopoietic Progenitors

For FACS, E14.5 FL from *Cdh5-CreER^{T2}::R26-LSL-Cas9-GFP* were dissected and enzymatically digested with collagenase type I (Sigma-Aldrich, St.Louis, MO, USA) 0.12% (*w*/*v*), followed by mechanical dissociation via pipetting. Cells were incubated with Fc block (1:500) (BD Biosciences), PE-conjugated rat anti-CD45 (1:200) (BD Biosciences), and PE-Cy7-conjugated rat anti-CD117 (c-Kit, 1:200) (BD Biosciences). Live single CD45+ c-Kit+ GFP+ cells were isolated using a MoFlo Astrios Cell Sorter equipped with Summit software version 6.1 (Beckman Coulter, Brea, CA, USA).

2.7. Single Guide RNAs (sgRNA) Design and Nucleofection

sgRNAs were purchased from Synthego (Redwood City, CA, USA) and designed using the bioinformatic tools GT-Scan [34] and ATUM gRNA Design tool (ATUM, Newark, CA, USA) to target sites within intron 10 and intron 8 of the Mll and Af9 genes, respectively. The best guides were chosen based on the highest cutting score and the minimum potential off-targets with sequences containing up to 3 mismatches. The sgRNA sequences are Mll: 5′-AGGTCTGTCTTCTGCTACGC-3′; Af9: 5′-GTTGCACTTTCGGAATGTGT-3′.

The nucleofection of target cells was performed with an Amaxa Nucleofector 4D (Lonza, Basel, Switzerland), following the manufacturer's protocol. The program of nucleofection used here was CM-137. Nucleofected cells were then transferred on fresh medium and co-cultured with OP9 stromal cells for a maximum of 7 days (described below).

2.8. CFU-C Assays

Sorted Cas9-expressing hematopoietic progenitors (CD45+ c-Kit+ GFP+) nucleofected with both sgRNAs against Mll and Af9 were cultured with MethoCult GF M3434 (Stem Cell Technologies, Vancouver, BC, Canada), as per the manufacturer's instructions, to give rise to CFU-E, CFU-G/M/GM, and CFU-GEMM. The number and identity of colonies were scored after 7 days, after which replating with an equal number of starting cells was performed.

2.9. Cell Cultures

All cell cultures were grown at 37 °C, 5% CO_2. Hematopoietic progenitor cells isolated from E14.5 FL were co-cultured with OP9 cells and maintained in Iscove's Modified Dulbecco's Medium Glutamax (IMDM) (GIBCO/ThermoFisher, Paisley, Scotland, UK) supplemented with 10% Fetal Bovine serum (FBS) (HyClone, Logan, UT, USA), 1% Bovine Serum Albumin (BSA), 1% Pen/Strep, 50 ng/mL SCF, 50 ng/mL IL-3, and 50 ng/mL FLT-3L for a maximum of 7 days. All cytokines were purchased from PeproTech (Cranbury, NJ, USA).

2.10. Statistical Analysis

No specific randomization method was used in the present study. Experimental animals were allocated to groups according to their genotype. No specific methods were used for blinding. In general, samples were collected from mice by one individual and then processed and analyzed by different individuals, at a time during which the genotypes or experimental conditions of each sample were not known. To determine the level of significance, an unpaired Student's *t*-test with Welch correction was used through GraphPad Prism software version 9.4.1. $p < 0.05$ was considered statistically significant, and the level of significance is indicated by asterisks: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

3. Results

3.1. In Vivo Introduction of Mll-Af9 Inter-Chromosomal Translocation in Embryonic HSPCs Results in a Transplantable Myeloproliferative Disease

To generate a model in which the Mll-Af9 translocation is introduced in vivo in embryonic HSPCs, we took advantage of the Mll-Af9 "translocator" mouse model, in which *loxP* sites are inserted in *Mll* and *Af9* genes in positions corresponding to the breakpoints observed in patients [26]. In the presence of Cre recombinase, in vivo inter-chromosomal recombination events are induced, sufficient to cause transplantable myeloid leukemia provided the cellular context is permissive [27].

Here, we crossed Mll-Af9 translocator mice with *Tie2-Cre* transgenic mice (Figure 1A), which express a constitutive Cre recombinase under the control of the Tie2 promoter, active in endothelial cells from early embryonic stages, including the hemogenic endothelium (HE), a specialized transient cell type present during ontogeny which is the precursor to most blood cells [35]. Hence, genetic targeting with the *Tie2-Cre* allele results in recombination in embryonic hematopoietic progenitor cells, which eventually give rise to the large majority of postnatal blood cells [36].

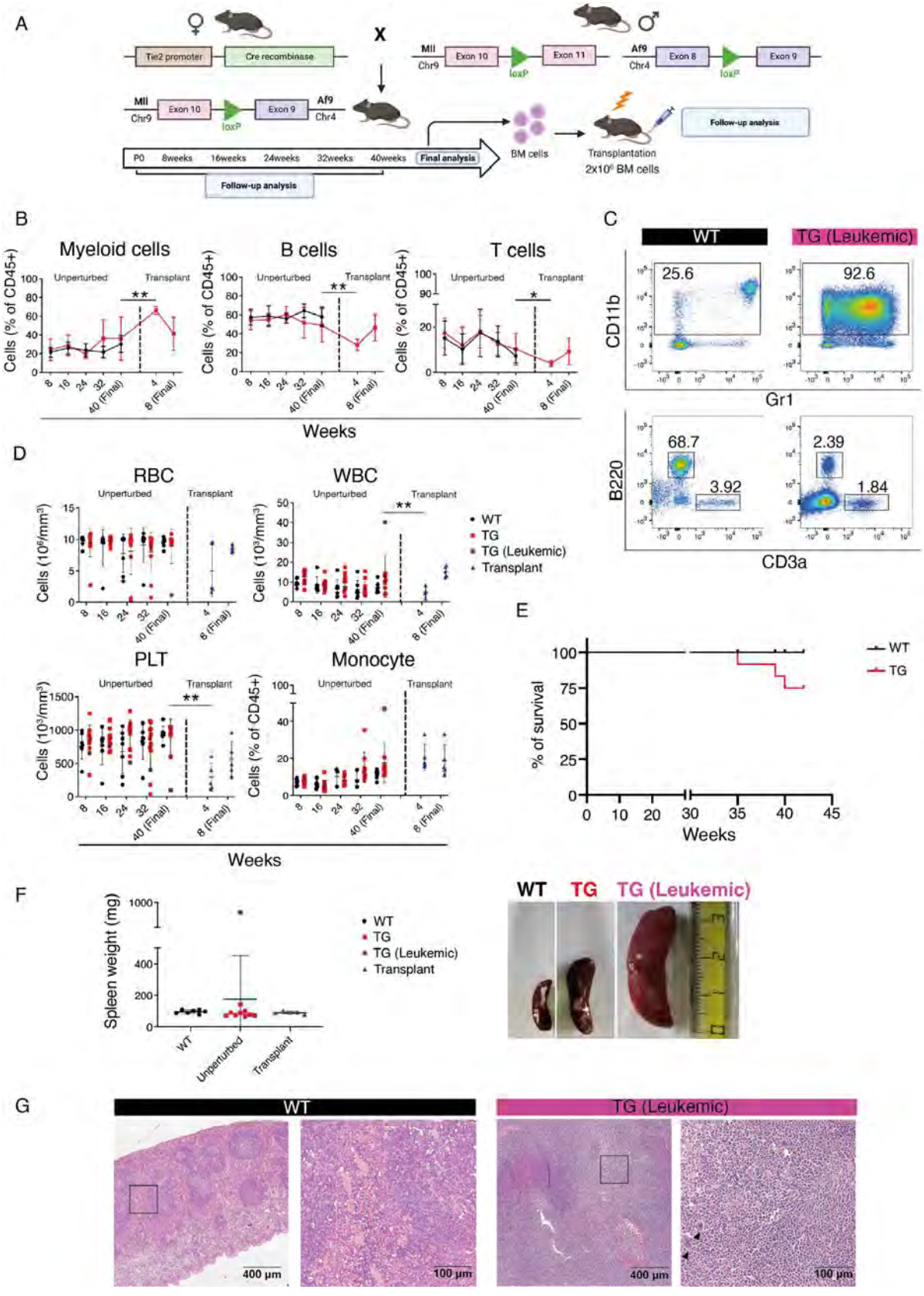


Figure 1. Generation and analysis of *Tie2-Cre::Mll*^{loxP}*::Af9*^{loxP} mice. (**A**) Schematic of crossing strategy and analyses of *Tie2-Cre::Mll*^{loxP}*::Af9*^{loxP} mice. (**B**) Quantification of flow cytometric analysis displayed

in (**C**) of myeloid (left), B (center), and T cells (right) in *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* (TG, red) or wild-type (WT, black) peripheral blood (PB) in unperturbed and transplant settings. Data are shown as mean ± standard deviation (SD). * $p < 0.05$, ** $p < 0.01$. Indicated statistical comparisons were made between transgenic (TG) mice at the indicated time points. (**C**) Representative flow cytometric analysis of PB cells in unperturbed WT (left) or leukemic *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* (TG, right) adult mice at 40 weeks. (**D**) Red blood cell (RBC, top left), white blood cell (WBC, top right), platelets (PLT, bottom left) absolute counts, and monocyte percentage (bottom right) in *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* (TG, red or leukemic TG, purple) or WT mice (black) in unperturbed and transplant settings. Data are mean ± SD. ** $p < 0.01$. (**E**) Kaplan–Meier survival curve of WT and TG (*Tie2-Cre::Mll^{loxP}::Af9^{loxP}*) mice over the course of the analysis. (**F**) (Left) Spleen weight of WT (left), unperturbed TG (center), and transplanted (right) mice. The leukemic TG mouse that was analyzed is shown in purple. (Right) Comparison between the spleen size of unperturbed background WT (top), *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* non-leukemic TG (center), and leukemic TG (right) mice. Data are mean ± SD. (**G**) Spleen histology of unperturbed mice, showing the typical splenic structure in the WT (left) compared to the absence of splenic architecture in the leukemic *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* (right), with the presence of histiocytic infiltrates (black arrow heads). Boxed areas in left panels of WT and TG (Leukemic) are magnified on the right. Scale bar: 400 μm (left panels), 100 μm (right panels).

Tie2-Cre::Mll^{loxP}::Af9^{loxP} mice were born at normal expected Mendelian ratios and we could detect evidence of the Mll-Af9 inter-chromosomal translocation in the bone marrow (BM) of these animals by PCR (Figure S1A). Periodic follow-up by peripheral blood (PB) flow cytometric analysis and blood counts showed signs of overt disease in a few cases out of a cohort of 12 double transgenic mice, in which a large increase in the percentage of myeloid cells and concomitant decrease in B and T lymphocytes was detected already from 32 weeks (Figure 1B,C).

Accordingly, leukocytes, red blood cells (RBC), and platelets (PLT) did not show alterations except in leukemic mice, whereas the percentage of monocytes showed a steady increase in all transgenic mice, apparent from 32 weeks (Figure 1D). Unfortunately, two transgenic mice died before the final analysis, likely of myeloid leukemia (Figure 1E). Mice were sacrificed at 40 weeks, and the spleen of a leukemic mouse, which showed a large increase in myeloid cells, was found to be abnormally enlarged (Figure 1F). A histological analysis found a complete subversion of the splenic architecture composed of white and red pulp, and the presence of large nucleated cells, identifiable as histiocytic infiltrates (arrowheads in Figure 1G). These results showed that the introduction of the Mll-Af9 translocation in embryonic HSPCs could generate a myeloproliferative disease.

To determine whether BM cells from *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* mice contained cells able to propagate the disease, we performed transplants into lethally irradiated syngeneic recipients (Figure 1A and Figure S1B). We transplanted cells from four different transgenic mice into seven recipients (1 to 2 recipient mice for each donor). Notably, the two recipients transplanted with the BM of the primary mouse showing the overt leukemic phenotype died 14 and 15 days post-transplant, probably due to severe anemia; therefore, they were not included in the subsequent analysis. In all other cases, we could detect a large increase in myeloid cells accompanied by significant reduction in B lymphocytes already from 4 weeks post-transplant (Figure 1B and Figure S1B). Because by 8 weeks post-transplantation recipient mice showed clear signs of suffering with large increases in leukocyte counts and monocytes (Figure 1D) and weight loss [37], these animals were suppressed. These data suggest that the myeloid disease resulting from the introduction of Mll-Af9 in embryonic HSPCs is transplantable.

3.2. Transplantation of HSC-Independent Hematopoietic Progenitors Carrying Mll-Af9 Translocation Induces a Myeloid Neoplastic Disease

The *Tie2-Cre* allele directs recombination to all subsets of HE, and therefore targets both HSC-independent progenitors (such as erythro-myeloid progenitors, or EMPs) and HSCs [38]. To generate a model that would allow the selective introduction of the Mll-Af9 translocation into HSC-independent progenitors separate from HSCs, we turned to an

inducible strategy by employing the *Cdh5-CreER*^{T2} model. It was previously shown that the administration of 4-hydroxytamoxifen (4-OHT) at E7.5 and E10.5 in these mice can direct recombination to EMPs and HSCs, respectively [28]. We validated this experimental strategy by performing lineage tracing experiments in which we crossed *Cdh5-CreER*^{T2} mice with *R26-LSL-tdTomato* or *R26-LSL-zsGreen* reporters [32] and induced Cre recombination with a single dose of 4-OHT or tamoxifen at E7.5 or E10.5 (Figure 2A). We confirmed that 4-OHT at E7.5 directs recombination to the majority of EMPs in the E10.5 yolk sac (YS) (Figure 2B,C) and in the E11.5 FL (Figure 2D), while largely avoiding recombination in E14.5 FL HSCs (Figure 2B,C). Conversely, E10.5 tamoxifen administration yields recombination in most E14.5 FL HSCs (Figure 2B,C). Notably, E10.5 Cre induction (performed with 4-OHT, chosen over tamoxifen in this particular experiment because of its more rapid metabolization) [39] avoids targeting the recombination to EMPs in the E11.5 FL (Figure 2D).

Next, we crossed *Cdh5-CreER*^{T2} with *Mll*^{loxP}*::Af9*^{loxP} mice and activated Cre recombination with 4-OHT at E7.5 or E10.5 (Figure S2A). Double transgenic mice were born at normal ratios and were followed up for 48 weeks, without evidence of alterations in blood parameters or in the percentage of myeloid cells, B, and T lymphocytes (Figure S2B,C). As with the *Tie2-Cre* constitutive strategy, the leukemic phenotype was more evident after transplantation than in the unperturbed background (Figure 1); in order to conclusively determine whether HSC-independent progenitors are susceptible to Mll-Af9-mediated transformation, we next decided to combine the inducible Cre-Lox strategy with transplantation. We crossed *Cdh5-CreER*^{T2} with *Mll*^{loxP}*::Af9*^{loxP} lines and activated Cre recombination with 4-OHT at E7.5 or tamoxifen at E10.5 (Figure 3A). For this specific time point, we used tamoxifen as it has a longer half-life than 4-OHT, and therefore yields recombination in a higher percentage of fetal HSCs ([39]; unpublished data). We isolated E14.5 FL from *Cdh5-CreER*^{T2}*::Mll*^{loxP}*::Af9*^{loxP} and Cre-controls ("WT"), and transplanted cells into lethally irradiated recipients (Figure 3A). The 352-bp PCR band corresponding to the Mll-Af9 translocation was detected in 100% of the transgenic FL samples activated at E7.5, whereas in 50% of those activated at E10.5, probably reflecting the fact that in negative samples the number of cells carrying the translocation was under the threshold level of detection at this stage (Figure S3A,B). In all recipient mice, from 12 weeks post-transplant onwards, PB donor chimerism was high (>90%) (Figure 3B). At 24 weeks post-transplant, we detected a significant increase in the percentage of donor-derived PB myeloid cells with a decrease in B and T lymphocytes only in mice transplanted with FL activated at E7.5 (Figure 3B,C). Therefore, at 28 weeks post-transplant, we sacrificed transplanted mice. Absolute counts of leukocytes, RBC, and PLT did not show significant differences at this time point (Figure 3D), and the spleen sizes of transgenic and control animals were comparable (Figure 4A). BM analysis showed very high donor chimerism in all animals, and, again only with the E7.5 activation, a significant increase in myeloid cells was observed, accompanied by a decrease in B and T lymphocytes (Figure 3E), although this was not mirrored in the PB (Figure 3B).

Next, we performed secondary transplantations using the BM of animals transplanted with transgenic FL cells. Secondary transplanted mice quickly developed signs of disease with prominent leukocytosis in all animals (Figure 3D), which were therefore sacrificed at 12 weeks post-transplant. The Mll-Af9 PCR band was detected in the majority of secondary transplanted animals in both experimental conditions, suggesting that the expansion of cells carrying the translocation had occurred (Figure S3). However, only animals transplanted with cells activated at E7.5 showed a significant increase in myeloid cells in the BM, with a concomitant decrease in B lymphocytes, whereas with the E10.5 activation we observed a slightly increased percentage of B lymphocytes and no change in myeloid cells (Figure 3E). Secondary transplanted mice showed significantly enlarged spleens, particularly those transplanted with cells activated at E7.5 (Figure 4A). A histological examination of the liver and lungs from these animals showed clear signs of extramedullary hematopoiesis in both organs, with leukocyte infiltrates often found adjacent to vessels (Figure 4B).

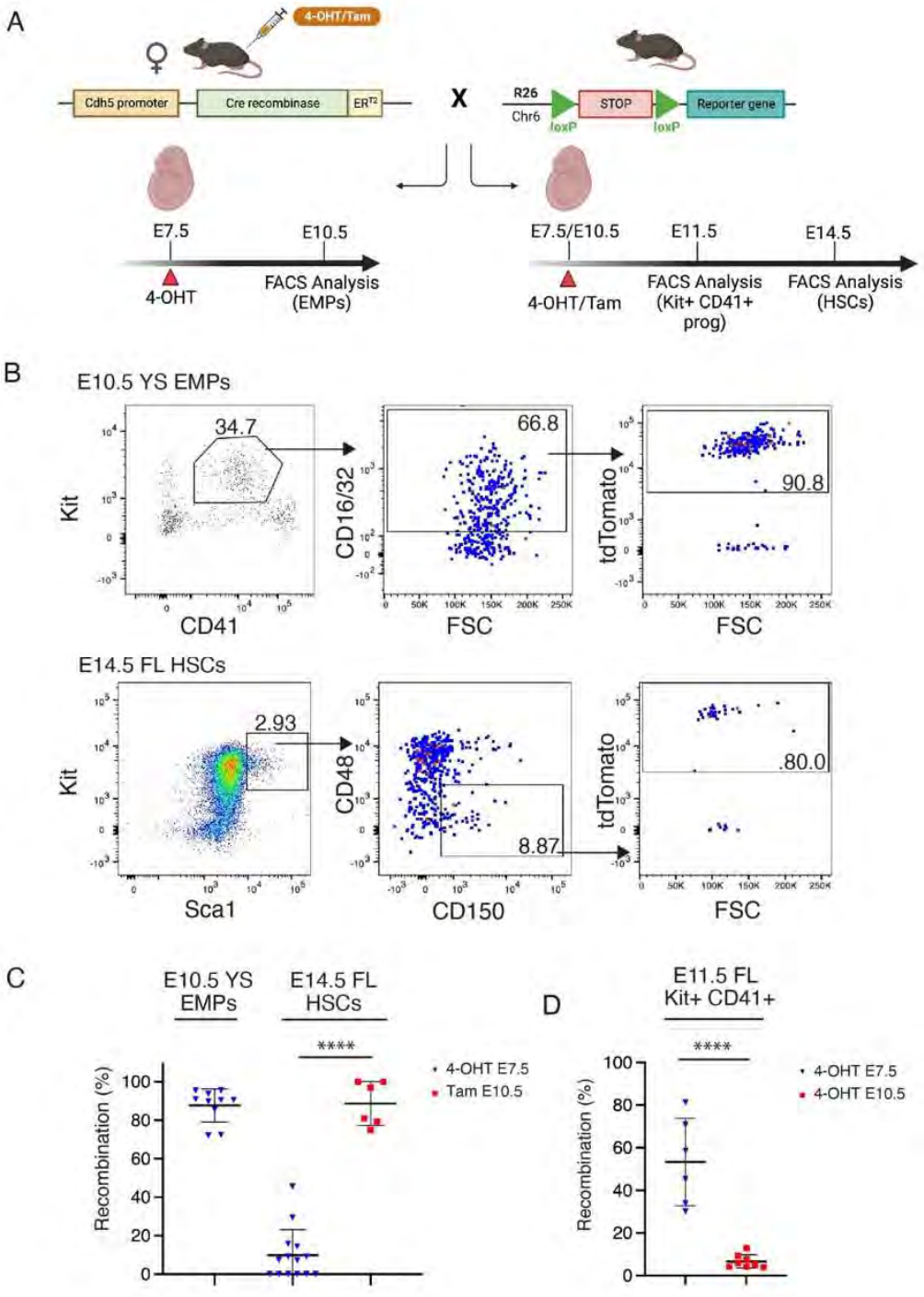


Figure 2. Validation of *Cdh5-CreER^{T2}* mice for selective labeling of EMPs and HSCs. (**A**) Schematic of crossing strategy and analyses of *Cdh5-CreER^{T2}::R26-tdTomato/zsGreen* mice for lineage tracing analysis. (**B**) Representative flow cytometric analysis of *CreER^{T2}::R26-tdTomato* embryos. E10.5 YS EMPs activated with 4-OHT at E7.5 (upper) and E14.5 FL HSCs activated with tamoxifen (Tam) at E10.5 (lower) are shown. (**C**) Quantification of flow cytometric analyses displayed in (**B**), plus analysis of E14.5 FL HSCs activated with 4-OHT at E7.5. Each dot represents a single embryo. (**D**) Flow cytometric analysis of Kit+ CD41+ hematopoietic progenitor cells in *Cdh5-CreER^{T2}::R26-tdTomato/zsGreen* E11.5 FL activated with 4-OHT at E7.5 (blue) or E10.5 (red). Each dot represents a single embryo. Data are shown as mean ± SD. **** $p < 0.0001$.

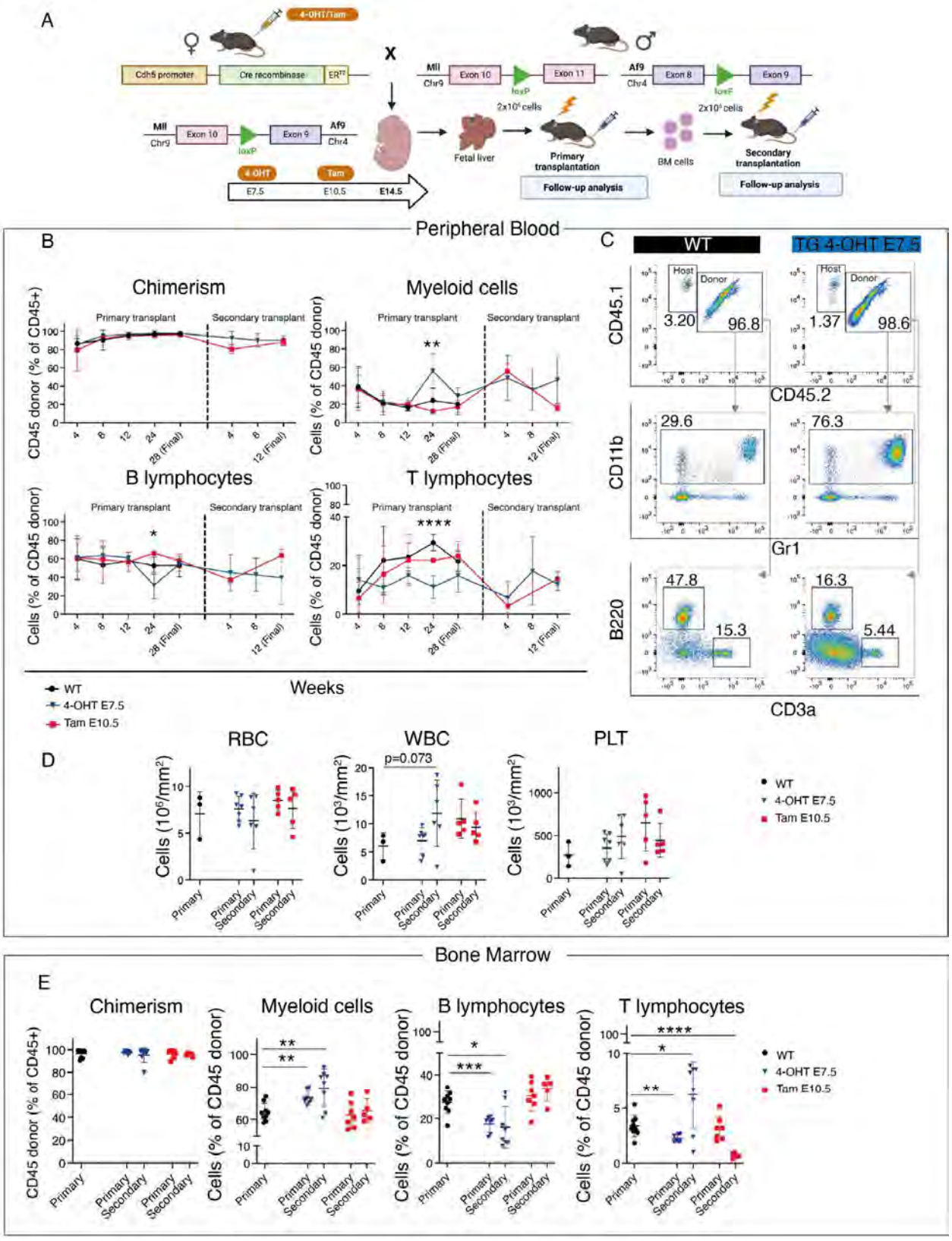


Figure 3. Analysis of *Cdh5-CreER*^{T2}*::Mll*^{loxP}*::Af9*^{loxP} primary and secondary transplants. (**A**) Schematic of crossing strategy, Cre activation (4-OHT at E7.5 or tamoxifen at E10.5) and analyses of primary fetal

liver (FL) and secondary bone marrow (BM) *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* transplants. (**B**) Quantification of PB flow cytometric analysis displayed in (**C**) of donor-derived chimerism (top left), donor myeloid (top right), (**B**) (bottom left) and T lymphocytes (bottom right) in primary and secondary transplant recipients of WT (black) and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* cells with the administration of 4-OHT at E7.5 (blue) or tamoxifen at E10.5 (red). Data are mean ± SD. ** $p < 0.01$, **** $p < 0.0001$. The statistical significance refers to the comparison of 4-OHT E7.5-activated *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* mice with WT mice. (**C**) Representative flow cytometric analysis of PB chimerism (top), donor myeloid (center), B and T lymphocytes (bottom) in primary transplant recipient mice of WT (left) or 4-OHT E7.5-activated *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* (TG, right) cells. Donor cells are CD45.1+/CD45.2+; host cells are CD45.1+. (**D**) PB, RBC (left), WBC (center), PLT (right) absolute counts of primary and secondary transplant recipient mice of WT (black) and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* cells with the administration of 4-OHT at E7.5 (blue) or tamoxifen at E10.5 (red). Analysis was performed in 28-week-old mice for primary transplants, and 12-week-old mice for secondary transplants. Data are mean ± SD. (**E**) Quantification of BM flow cytometric analysis of donor-derived chimerism (left), myeloid (center left), B (center right), and T lymphocytes (right) in primary and secondary transplant recipient mice of WT (black) and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* cells with the administration of 4-OHT at E7.5 (blue) or tamoxifen at E10.5 (red). Data are mean ± SD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

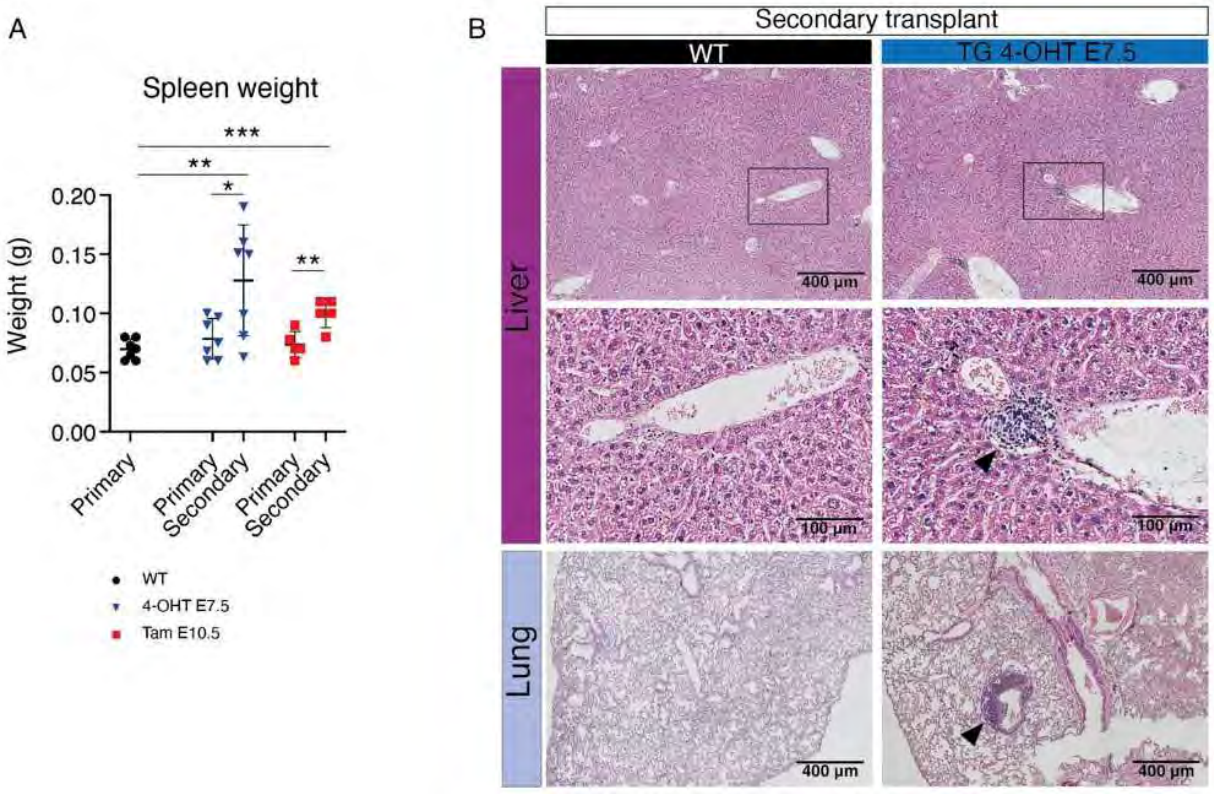


Figure 4. Morphological and histological analysis of extramedullary hematopoiesis in *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* transplanted mice. (**A**) Spleen weight of WT (black) and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* mice with the administration of 4-OHT at E7.5 (blue) or tamoxifen at E10.5 (red) in primary and secondary transplants. Data are mean ± SD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (**B**) Histological analysis showing the presence of extramedullary leukocyte infiltration (black arrow heads) in liver (top and middle) and lung (bottom) in 4-OHT E7.5-activated *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* (right) secondary transplants. WT controls (left) show absence of extramedullary infiltration. Middle panels are magnifications of boxed areas shown in top panels. Scale bar: 400 µm (top and bottom panels), 100 µm (magnification, middle panels).

Taken together, these results support the idea that HSC-independent progenitors, including EMPs, are susceptible to leukemic transformation upon acquisition of the Mll-Af9 inter-chromosomal translocation.

3.3. CRISPR/Cas9 Engineering of the Mll-Af9 Translocation in HSC-Independent Hematopoietic Progenitors Triggers an Aberrant Self-Renewal Program

An acute leukemic phenotype was only evident in secondary transplanted mice in which the Mll-Af9 rearrangement was introduced in HSC-independent progenitors, which suggests a long latency, possibly reflecting the low efficiency of the translocator mouse-based approach. It is not clear whether Mll-Af9 can directly affect the self-renewal capacity of HSC-independent progenitors. To answer this question, we decided to employ a CRISPR/Cas9-based approach. Thus, we combined CRISPR/Cas9 knockin mice [31] with an electroporation-based method for the engineering of chromosomal rearrangements [40,41], which we adapted and optimized here for mouse FL cells.

To selectively introduce the Mll-Af9 translocation in HSC-independent hematopoietic progenitors using CRISPR/Cas9, we crossed *Cdh5-CreER^{T2}* with *R26-LSL-Cas9-GFP* and we induced Cre recombination with 4-OHT at E8.5 (Figure 5A). This time, E8.5 was chosen as the selected time point of Cre induction instead of E7.5, as it results in higher recombination rates in FL c-Kit+ hematopoietic progenitors, including EMPs, but still largely avoiding the labeling of definitive-type adult HSCs, which emerge only after E10.5 ([17]; Barone et al., under revision). We next isolated CD45+ c-Kit+ GFP+ cells from E14.5 FL of *Cdh5-CreER^{T2}::R26-LSL-Cas9-GFP* embryos by FACS (Figure 5B), and nucleofected them with sgRNAs against Mll and/or Af9, which, for consistency, we designed to target intronic breakpoints (intron 10 for Mll, intron 8 for Af9) analogous to the ones present in *Mll^{loxP}::Af9^{loxP}* mice (Figure 5C). When both sgRNAs were used together, we obtained clear evidence of the Mll-Af9 inter-chromosomal rearrangement by PCR (Figure 5D), which was also confirmed by Sanger sequencing (Figure 5E). The variability in the sequence of the Mll-Af9 rearrangement at the breakpoints reflects the fact that this analysis was conducted in bulk (Figure 5E). Next, after a 7-day co-culture period with OP9 stromal cells, we plated translocated and control non-transfected FL hematopoietic progenitors in a methylcellulose-based media containing cytokines. Already upon the first plating, edited cells gave rise to significantly more colonies than control cells, including large granulocyte, erythroid, monocyte, and megakaryocyte (GEMM) colonies, which were absent in control cells (Figure 5F). Mll-Af9 edited cells produced more colonies than control cells until the third round of replating, indicating that they had acquired an enhanced self-renewal potential (Figure 5F). Accordingly, the transplantation of Mll-Af9 edited FL HSPCs in sub-lethally irradiated recipients yielded a strongly myeloid-biased engraftment (Figure 5G). These data support the idea that Mll-Af9 confers enhanced self-renewal to HSC-independent myeloid progenitors.

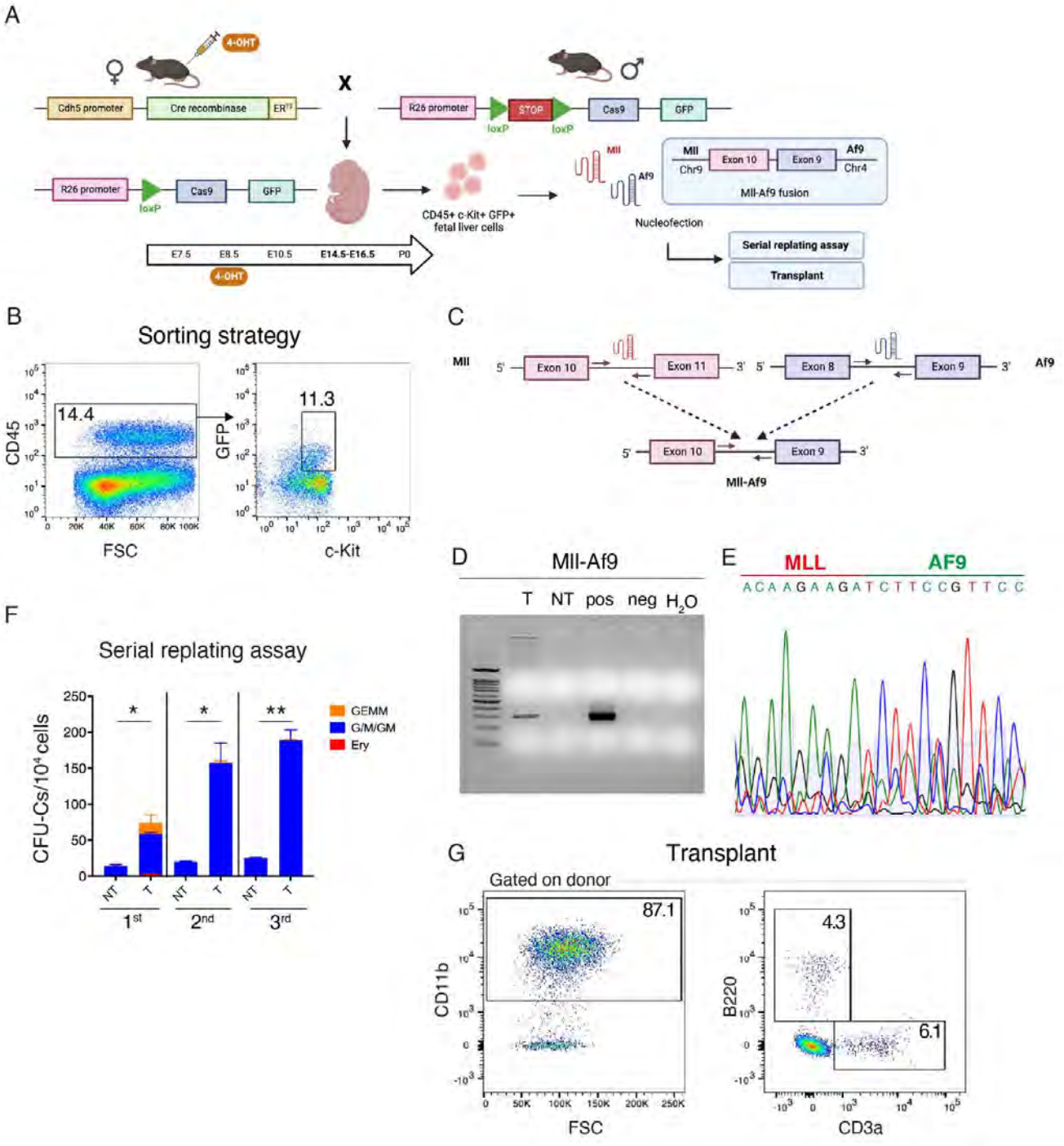


Figure 5. CRISPR/Cas9 engineering of the Mll-Af9 fusion gene in hematopoietic stem cell (HSC)-independent hematopoietic progenitors. (**A**) Experimental schematic of the strategy for the isolation, gene editing, and analysis of fetal HSC-independent progenitors. (**B**) FACS isolation of HSC-independent progenitors. Live CD45+ c-Kit+ GFP+ cells were sorted from the FL of *Cdh5-CreER^{T2}::R26-LSL-Cas9-GFP* E14.5 embryos (4-OHT at E8.5). (**C**) sgRNA were designed to target intron 10 of the Mll (red) and intron 8 of the Af9 (blue) genes. When electroporated in combination, the formation of the inter-chromosomal Mll-Af9 translocation is induced. Arrows represent PCR primers used to detect the translocation band. (**D**) Representative agarose gel showing the presence of the Mll-Af9 translocation with an amplicon of the expected size (300 bp) observed in transfected cells

(T), and absent in non-transfected cells (NT). Positive (pos) and negative (neg) controls are shown. (**E**) Sanger sequencing confirming the identity of the Mll-Af9 translocation. The variability in the peaks around the breakpoint is due to the sequencing reaction being performed on bulk-edited cells. A = green; C = blue; G = black; T = red. (**F**) Serial replating CFU assay with sorted hematopoietic stem and progenitor cells (HSPCs) transfected with the combination of the sgRNAs against Mll and Af9 genes (T) and non-transfected HSPCs (NT). Colonies were scored after 7 days following plating. Granulocyte/erythroid/monocyte/macrophage (GEMM, orange), granulocyte/macrophage (G/M/GM, blue), and erythroid (Ery, red) colonies were scored. Data are mean ± SD. * $p < 0.05$, ** $p < 0.01$. (**G**) Representative peripheral blood flow cytometric analysis of recipient mice engrafted with Mll-Af9 edited E16.5 FL HSPCs, sampled 4 weeks post-transplantation (n = 2 mice transplanted).

4. Discussion

The mutations driving AML change dramatically with age [6]. Adult AML is mostly driven by the progressive accumulation of focal mutations, whereas the large majority of pediatric AML are characterized by the presence of fusion proteins originating from different structural variants [6,9]. Such diverse genetics are also reflected by the overall mutation burden, which is much higher in adult AML [9]. However, the real biological meaning of this phenomenon is still unclear. One hypothesis that has been proposed is that translocations could simply occur more frequently during embryonic/fetal hematopoiesis than during postnatal life [13]. This was supported by the idea that fetal HSCs cycle more frequently than adult HSCs, which are instead normally present in a more quiescent state [42]. Thus, the age-specific mutation profile would be consistent with the distinct type of DNA damage that HSPCs are exposed to at different ages in life. Fetal hematopoiesis would present mutations especially induced by replication stress, whereas adult life would expose HSPCs to the physiological age-driven accumulation of single nucleotide variants or indel mutations. Although this can certainly be a plausible explanation, it is probably not the only one.

Recent work shows that the pool of HSCs fated to adult life undergoes limited cell divisions during gestation [23]; thus, in this scenario, the previously observed extensive proliferation of FL HSPCs would primarily be carried out by a pool of stem/progenitor cells ultimately fated to differentiation [21,23,43]. If this relatively new understanding of fetal hematopoiesis eventually holds to be true, a possible consequence would imply that adult-fated HSCs are likely not the cell of origin of most hematological malignancies originating prenatally. Moreover, the higher frequency of chromosomal rearrangements would not be explained by replication stress-induced DNA damage, at least not in adult-fated HSCs.

The hypothesis that the intrinsic properties of fetal hematopoiesis may act as vulnerabilities for pediatric hematological malignancies is not a new one. Thus, another possible explanation for the age-specific mutation profile of AML could be that fetal myeloid progenitors are more susceptible than adult ones to leukemic transformation upon the acquisition of AML-associated translocations. Indeed, some investigators provided experimental support for this hypothesis. For example, the ETO2-GLIS2 fusion was shown to transform fetal progenitors with more efficiency than adult ones in mice [44]. In this model, the leukemic phenotype and latency were also ontogeny-dependent [44]. Therefore, developmental hematopoiesis would provide a "window of opportunity" for leukemic transformation, which could be largely driven by the specific cellular environment and high proliferation capacity of HSPCs distinct from adult-fated HSCs.

Among HSC-independent progenitors, EMPs were recently shown to substantially contribute to fetal erythropoiesis [19] as well as to long-lasting immune cells that, when subject to various perturbations, have the potential to affect adult disease [45]. So far, EMPs have not been recognized as potential cells of origin of pediatric AML, or other childhood myeloid neoplasms driven by inter-chromosomal translocations, probably due to their transiency and lack of long-lasting engraftment potential [46]. However, being intrinsically very proliferative [46], EMPs would remarkably fit the criteria of the higher susceptibility to replication stress damage.

Our work here shows that EMPs are, in principle, susceptible to Mll-Af9-mediated leukemic transformation. Although we have not formally determined the progenitor cell type undergoing malignant transformation in our experimental model, previous reports which we and others were able to reproduce have demonstrated that 4-OHT activation at E7.5 in the *Cdh5-CreER*^{T2} model targets EMPs almost exclusively, avoiding recombination in the HSC-dependent wave [28,29]. Our lineage tracing experiments, presented in Figure 2, support the feasibility of using the *Cdh5-CreER*^{T2} model to separately target EMP- and HSC-derived hematopoietic waves.

To our knowledge, this is the first report demonstrating that the selective introduction of a chromosomal rearrangement in HSC-independent progenitors can result in leukemic disease. A recent study employed a similar Cre-Lox translocator mice-based strategy to develop embryonic and adult models of Mll-Enl driven leukemia; however, the induction strategy in this study could not discriminate between HSC-dependent and independent waves [15]. Other reports conclusively reported a fetal origin for the Mll-Af4 pro-B acute lymphoblastic leukemia, one of the most common hematological malignancies in infants [47,48]. The fetal lymphoid-primed multipotent progenitor (LMPP) [49] was suggested as a potential cell of origin [50], although the experimental models used were not able to discriminate between HSC-dependent or independent hematopoietic waves. Another study reported T-ALL development upon the overexpression of the Notch intracellular domain in early intraembryonic, but not yolk sac, precursors [51]. Once again, this study did not explain what the precise nature of leukemia propagating cells was in vivo, although the fact that only intraembryonic cells were susceptible to leukemic transformation in this setting was a strong indicator of an HSC origin, in agreement with EMPs being independent of Notch signaling [52,53]. Overall, these studies provide a strong experimental rationale supporting that several types of infant leukemias can have a fetal origin, but, at the same time, highlight the difficulty of identifying which progenitor cell types are susceptible to transformation.

One important characteristic of the experimental models we used in this study is that both the Cre-Lox model and the CRISPR/Cas9-based approach rely on inter-chromosomal translocation events; therefore, the expression of deriving fusion genes is driven by endogenous regulatory elements rather than by an extrinsic promoter, making the model more physiological than transgene overexpression. A consequential limitation is that the efficiency of leukemic transformation was probably low, and this was reflected by the long incubation times we observed. Additionally, in the Cre-Lox model, not all cells in which the Cre is expressed will generate an inter-chromosomal translocation. Our conditional Cre activation protocol induced recombination in a subset of cells in a short time window during embryonic development; therefore, it is highly probable that when using a conditional strategy, the clones in which the Mll-Af9 translocation is generated are of a small size. Indeed, except in rare cases, a clear leukemic phenotype was seen only in transplanted mice, possibly because lethal irradiation acted as a perturbation trigger that allowed transformed clones to expand.

Our data suggest that in infant AML, the acquisition of a myeloid-biased self-renewal program driven by Mll-Af9 in HSC-independent progenitors (EMPs) that are normally highly proliferating could lead to leukemic transformation. Interestingly, a myeloid phenotype was only observed when Mll-Af9 was introduced in EMPs, but not in embryonic HSCs, at least within the timeframe of our analysis. Our results are consistent with the delayed latency and mixed phenotype observed when Mll-Af9 was introduced in fetal HSCs using a knockin model [14]. Nevertheless, our results do not exclude that HSCs could also act as cells of origin for infant AML; however, it reinforces the possibility that differences in the cell of origin could influence the latency and the phenotype of the disease.

In our experimental setting, we could not determine whether embryonic-derived Mll-Af9 leukemia was more aggressive than its adult counterpart because all our experiments were performed in the context of embryonic hematopoiesis. Instead, we focused on determining which of the embryonic hematopoietic waves contained cells susceptible

to leukemic transformation upon the introduction of Mll-Af9 rearrangement. Besides the proliferation rate, what other specific characteristics of HSPCs belonging to distinct embryonic hematopoietic waves would make them more or less vulnerable to leukemic transformation remains an open question. Recent studies reported a window of opportunity for leukemogenesis in neonatal blood progenitors upon the introduction of the Mll-Enl translocation [16]. The same group also reported that the transition to fetal and adult hematopoiesis taking place in neonates was accompanied by a peak in IFN signaling [42]. Intriguingly, pro-inflammatory signaling, including IFNs, was also shown to be particularly important for embryonic HSC development [54]. It will be interesting to further explore the role of inflammation in pediatric leukemias, especially in the context of recent insight in developmental hematopoiesis.

5. Conclusions

In conclusion, here we provide an experimental "proof-of-principle" showing that HSC-independent hematopoietic progenitors, which include EMPs, are susceptible to leukemogenesis upon the acquisition of the Mll-Af9 inter-chromosomal rearrangement. Further studies should investigate the specific intrinsic and extrinsic characteristics of the cellular environment in embryonic and fetal hematopoiesis, and the molecular pathways that might be responsible for malignant transformation. Obtaining such insights will be important in order to identify the potentially targetable and fetal-specific vulnerabilities of infant leukemias.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/cancers15143624/s1, Figure S1: Additional analysis of *Tie2-Cre*::Mll^{loxP}::$Af9^{loxP}$ mice and transplants; Figure S2: Analysis of unperturbed adult *Cdh5-CreER*^{T2}::Mll^{loxP}::$Af9^{loxP}$ mice; Figure S3: Detection of Mll-Af9 translocation in *Cdh5-CreER*^{T2}::Mll^{loxP}::$Af9^{loxP}$ transplants; Table S1: Primer sequences used for Mll-Af9 detection by PCR.

Author Contributions: Conceptualization, S.B. (Silvia Brunelli) and E.A.; formal analysis, C.B. (Cristiana Barone), R.O., A.C., E.D., A.P. and E.A.; funding acquisition, E.A.; investigation, C.B. (Cristiana Barone), R.O., A.C., E.D., A.P., G.Q., B.V. and S.B. (Silvia Bombelli), S.D.M., C.D. and E.A.; methodology, T.H.R. and E.A.; project administration, C.B. (Cristiana Barone), A.C. and E.A.; resources, S.B. (Silvia Bombelli), S.D.M., C.B. (Cristina Bianchi), B.E.L., A.B., G.C. and T.H.R.; supervision, C.D., R.M., S.B. (Silvia Brunelli) and E.A.; writing—original draft, C.B. (Cristiana Barone), R.O., G.Q. and E.A.; writing—review and editing, E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by research grants from Fondazione Cariplo (Biomedical Research Conducted by Young Researchers, ID: 2018-0102) and the Leukemia Research Foundation (New Investigator Blood Cancer Research Grant Program, ID: 831382) to E.A.

Institutional Review Board Statement: Animal studies included in this manuscript were approved by the San Raffaele Scientific Institute Animal Welfare Body and by the Italian Ministry of Health (Authorization n°503/2019-PR of 22 July 2019).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We thank Dario Bonanomi for providing Tie2-Cre mice, and Chiara Buracchi for assistance with flow cytometric analysis. Schematics were created with a licensed version of BioRender.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Biondi, A.; Cimino, G.; Pieters, R.; Pui, C.H. Biological and therapeutic aspects of infant leukemia. *Blood* **2000**, *96*, 24–33. [CrossRef] [PubMed]
2. Duguid, A.; Mattiucci, D.; Ottersbach, K. Infant leukaemia—Faithful models, cell of origin and the niche. *Dis. Model. Mech.* **2021**, *14*, dmm049189. [CrossRef] [PubMed]
3. Ibrahimova, A.; Pommert, L.; Breese, E.H. Acute Leukemia in Infants. *Curr. Oncol. Rep.* **2021**, *23*, 27. [CrossRef] [PubMed]
4. Marcotte, E.L.; Spector, L.G.; Mendes-de-Almeida, D.P.; Nelson, H.H. The Prenatal Origin of Childhood Leukemia: Potential Applications for Epidemiology and Newborn Screening. *Front. Pediatr.* **2021**, *9*, 639479. [CrossRef] [PubMed]
5. Masetti, R.; Vendemini, F.; Zama, D.; Biagi, C.; Pession, A.; Locatelli, F. Acute myeloid leukemia in infants: Biology and treatment. *Front. Pediatr.* **2015**, *3*, 37. [CrossRef] [PubMed]
6. Bolouri, H.; Farrar, J.E.; Triche, T., Jr.; Ries, R.E.; Lim, E.L.; Alonzo, T.A.; Ma, Y.; Moore, R.; Mungall, A.J.; Marra, M.A.; et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nat. Med.* **2018**, *24*, 103–112. [CrossRef]
7. Cazzola, A.; Cazzaniga, G.; Biondi, A.; Meneveri, R.; Brunelli, S.; Azzoni, E. Prenatal Origin of Pediatric Leukemia: Lessons From Hematopoietic Development. *Front. Cell Dev. Biol.* **2020**, *8*, 618164. [CrossRef]
8. Rice, S.; Roy, A. MLL-rearranged infant leukaemia: A 'thorn in the side' of a remarkable success story. *Biochim. Biophys. Acta Gene Regul. Mech.* **2020**, *1863*, 194564. [CrossRef]
9. Mercher, T.; Schwaller, J. Pediatric Acute Myeloid Leukemia (AML): From Genes to Models Toward Targeted Therapeutic Intervention. *Front. Pediatr.* **2019**, *7*, 401. [CrossRef]
10. Chaudhury, S.S.; Morison, J.K.; Gibson, B.E.; Keeshan, K. Insights into cell ontogeny, age, and acute myeloid leukemia. *Exp. Hematol.* **2015**, *43*, 745–755. [CrossRef]
11. Horton, S.J.; Jaques, J.; Woolthuis, C.; van Dijk, J.; Mesuraca, M.; Huls, G.; Morrone, G.; Vellenga, E.; Schuringa, J.J. MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. *Leukemia* **2013**, *27*, 1116–1126. [CrossRef]
12. Rowe, R.G.; Lummertz da Rocha, E.; Sousa, P.; Missios, P.; Morse, M.; Marion, W.; Yermalovich, A.; Barragan, J.; Mathieu, R.; Jha, D.K.; et al. The developmental stage of the hematopoietic niche regulates lineage in MLL-rearranged leukemia. *J. Exp. Med.* **2019**, *216*, 527–538. [CrossRef]
13. Mendoza-Castrejon, J.; Magee, J.A. Layered immunity and layered leukemogenicity: Developmentally restricted mechanisms of pediatric leukemia initiation. *Immunol. Rev.* **2023**, *315*, 197–215. [CrossRef]
14. Chen, W.; O'Sullivan, M.G.; Hudson, W.; Kersey, J. Modeling human infant MLL leukemia in mice: Leukemia from fetal liver differs from that originating in postnatal marrow. *Blood* **2011**, *117*, 3474–3475. [CrossRef]
15. Sinha, R.; Porcheri, C.; d'Altri, T.; Gonzalez, J.; Ruiz-Herguido, C.; Rabbitts, T.; Espinosa, L.; Bigas, A. Development of embryonic and adult leukemia mouse models driven by MLL-ENL translocation. *Exp. Hematol.* **2020**, *85*, 13–19. [CrossRef]
16. Okeyo-Owuor, T.; Li, Y.; Patel, R.M.; Yang, W.; Casey, E.B.; Cluster, A.S.; Porter, S.N.; Bryder, D.; Magee, J.A. The efficiency of murine MLL-ENL-driven leukemia initiation changes with age and peaks during neonatal development. *Blood Adv.* **2019**, *3*, 2388–2399. [CrossRef]
17. Barone, C.; Orsenigo, R.; Meneveri, R.; Brunelli, S.; Azzoni, E. One Size Does Not Fit All: Heterogeneity in Developmental Hematopoiesis. *Cells* **2022**, *11*, 1061. [CrossRef]
18. Soares-da-Silva, F.; Elsaid, R.; Peixoto, M.M.; Nogueira, G.; Pereira, P.; Bandeira, A.; Cumano, A. Assembling the layers of the hematopoietic system: A window of opportunity for thymopoiesis in the embryo. *Immunol. Rev.* **2023**, *315*, 54–70. [CrossRef]
19. Soares-da-Silva, F.; Freyer, L.; Elsaid, R.; Burlen-Defranoux, O.; Iturri, L.; Sismeiro, O.; Pinto-do, O.P.; Gomez-Perdiguero, E.; Cumano, A. Yolk sac, but not hematopoietic stem cell-derived progenitors, sustain erythropoiesis throughout murine embryonic life. *J. Exp. Med.* **2021**, *218*, e20201729. [CrossRef]
20. Ulloa, B.A.; Habbsa, S.S.; Potts, K.S.; Lewis, A.; McKinstry, M.; Payne, S.G.; Flores, J.C.; Nizhnik, A.; Feliz Norberto, M.; Mosimann, C.; et al. Definitive hematopoietic stem cells minimally contribute to embryonic hematopoiesis. *Cell Rep.* **2021**, *36*, 109703. [CrossRef]
21. Yokomizo, T.; Ideue, T.; Morino-Koga, S.; Tham, C.Y.; Sato, T.; Takeda, N.; Kubota, Y.; Kurokawa, M.; Komatsu, N.; Ogawa, M.; et al. Independent origins of fetal liver haematopoietic stem and progenitor cells. *Nature* **2022**, *609*, 779–784. [CrossRef]
22. Ema, H.; Nakauchi, H. Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood* **2000**, *95*, 2284–2288. [CrossRef] [PubMed]
23. Ganuza, M.; Hall, T.; Myers, J.; Nevitt, C.; Sanchez-Lanzas, R.; Chabot, A.; Ding, J.; Kooienga, E.; Caprio, C.; Finkelstein, D.; et al. Murine foetal liver supports limited detectable expansion of life-long haematopoietic progenitors. *Nat. Cell Biol.* **2022**, *24*, 1475–1486. [CrossRef] [PubMed]
24. Azzoni, E.; Fantin, A. Fetal liver hematopoiesis revisited: A precast hierarchy. *Nat. Cardiovasc. Res.* **2022**, *1*, 872–873. [CrossRef]
25. Kisanuki, Y.Y.; Hammer, R.E.; Miyazaki, J.; Williams, S.C.; Richardson, J.A.; Yanagisawa, M. Tie2-Cre transgenic mice: A new model for endothelial cell-lineage analysis in vivo. *Dev. Biol.* **2001**, *230*, 230–242. [CrossRef]
26. Collins, E.C.; Pannell, R.; Simpson, E.M.; Forster, A.; Rabbitts, T.H. Inter-chromosomal recombination of Mll and Af9 genes mediated by cre-loxP in mouse development. *EMBO Rep.* **2000**, *1*, 127–132. [CrossRef] [PubMed]

27. Drynan, L.F.; Pannell, R.; Forster, A.; Chan, N.M.; Cano, F.; Daser, A.; Rabbitts, T.H. Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *EMBO J.* **2005**, *24*, 3136–3146. [CrossRef]
28. Gentek, R.; Ghigo, C.; Hoeffel, G.; Bulle, M.J.; Msallam, R.; Gautier, G.; Launay, P.; Chen, J.; Ginhoux, F.; Bajenoff, M. Hemogenic Endothelial Fate Mapping Reveals Dual Developmental Origin of Mast Cells. *Immunity* **2018**, *48*, 1160–1171.e1165. [CrossRef]
29. Gentek, R.; Ghigo, C.; Hoeffel, G.; Jorquera, A.; Msallam, R.; Wienert, S.; Klauschen, F.; Ginhoux, F.; Bajenoff, M. Epidermal gammadelta T cells originate from yolk sac hematopoiesis and clonally self-renew in the adult. *J. Exp. Med.* **2018**, *215*, 2994–3005. [CrossRef]
30. Wang, Y.; Nakayama, M.; Pitulescu, M.E.; Schmidt, T.S.; Bochenek, M.L.; Sakakibara, A.; Adams, S.; Davy, A.; Deutsch, U.; Luthi, U.; et al. Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **2010**, *465*, 483–486. [CrossRef]
31. Platt, R.J.; Chen, S.; Zhou, Y.; Yim, M.J.; Swiech, L.; Kempton, H.R.; Dahlman, J.E.; Parnas, O.; Eisenhaure, T.M.; Jovanovic, M.; et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **2014**, *159*, 440–455. [CrossRef]
32. Madisen, L.; Zwingman, T.A.; Sunkin, S.M.; Oh, S.W.; Zariwala, H.A.; Gu, H.; Ng, L.L.; Palmiter, R.D.; Hawrylycz, M.J.; Jones, A.R.; et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **2010**, *13*, 133–140. [CrossRef]
33. Azzoni, E.; Frontera, V.; McGrath, K.E.; Harman, J.; Carrelha, J.; Nerlov, C.; Palis, J.; Jacobsen, S.E.W.; de Bruijn, M.F. Kit ligand has a critical role in mouse yolk sac and aorta-gonad-mesonephros hematopoiesis. *EMBO Rep.* **2018**, *19*, e45477. [CrossRef] [PubMed]
34. O'Brien, A.; Bailey, T.L. GT-Scan: Identifying unique genomic targets. *Bioinformatics* **2014**, *30*, 2673–2675. [CrossRef]
35. Swiers, G.; Rode, C.; Azzoni, E.; de Bruijn, M.F. A short history of hemogenic endothelium. *Blood Cells Mol. Dis.* **2013**, *51*, 206–212. [CrossRef]
36. Tang, Y.; Harrington, A.; Yang, X.; Friesel, R.E.; Liaw, L. The contribution of the Tie2+ lineage to primitive and definitive hematopoietic cells. *Genesis* **2010**, *48*, 563–567. [CrossRef]
37. Barone, C.; School of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy; Azzoni, E.; School of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy. Personal communication, 2022.
38. Chen, M.J.; Li, Y.; De Obaldia, M.E.; Yang, Q.; Yzaguirre, A.D.; Yamada-Inagawa, T.; Vink, C.S.; Bhandoola, A.; Dzierzak, E.; Speck, N.A. Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* **2011**, *9*, 541–552. [CrossRef]
39. Martinez-Corral, I.; Makinen, T. Genetic Lineage Tracing of Lymphatic Endothelial Cells in Mice. *Methods Mol. Biol.* **2018**, *1846*, 37–53. [CrossRef]
40. Jeong, J.; Jager, A.; Domizi, P.; Pavel-Dinu, M.; Gojenola, L.; Iwasaki, M.; Wei, M.C.; Pan, F.; Zehnder, J.L.; Porteus, M.H.; et al. High-efficiency CRISPR induction of t(9;11) chromosomal translocations and acute leukemias in human blood stem cells. *Blood Adv.* **2019**, *3*, 2825–2835. [CrossRef]
41. Rice, S.; Jackson, T.; Crump, N.T.; Fordham, N.; Elliott, N.; O'Byrne, S.; Fanego, M.; Addy, D.; Crabb, T.; Dryden, C.; et al. A human fetal liver-derived infant MLL-AF4 acute lymphoblastic leukemia model reveals a distinct fetal gene expression program. *Nat. Commun.* **2021**, *12*, 6905. [CrossRef]
42. Li, Y.; Kong, W.; Yang, W.; Patel, R.M.; Casey, E.B.; Okeyo-Owuor, T.; White, J.M.; Porter, S.N.; Morris, S.A.; Magee, J.A. Single-Cell Analysis of Neonatal HSC Ontogeny Reveals Gradual and Uncoordinated Transcriptional Reprogramming that Begins before Birth. *Cell Stem Cell* **2020**, *27*, 732–747.e7. [CrossRef]
43. Bowie, M.B.; McKnight, K.D.; Kent, D.G.; McCaffrey, L.; Hoodless, P.A.; Eaves, C.J. Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *J. Clin. Investig.* **2006**, *116*, 2808–2816. [CrossRef]
44. Lopez, C.K.; Noguera, E.; Stavropoulou, V.; Robert, E.; Aid, Z.; Ballerini, P.; Bilhou-Nabera, C.; Lapillonne, H.; Boudia, F.; Thirant, C.; et al. Ontogenic Changes in Hematopoietic Hierarchy Determine Pediatric Specificity and Disease Phenotype in Fusion Oncogene-Driven Myeloid Leukemia. *Cancer Discov.* **2019**, *9*, 1736–1753. [CrossRef]
45. Mass, E.; Gentek, R. Fetal-Derived Immune Cells at the Roots of Lifelong Pathophysiology. *Front. Cell Dev. Biol.* **2021**, *9*, 648313. [CrossRef]
46. 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. Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Rep.* **2015**, *11*, 1892–1904. [CrossRef]
47. Barrett, N.A.; Malouf, C.; Kapeni, C.; Bacon, W.A.; Giotopoulos, G.; Jacobsen, S.E.W.; Huntly, B.J.; Ottersbach, K. Mll-AF4 Confers Enhanced Self-Renewal and Lymphoid Potential during a Restricted Window in Development. *Cell Rep.* **2016**, *16*, 1039–1054. [CrossRef]
48. Symeonidou, V.; Jakobczyk, H.; Bashanfer, S.; Malouf, C.; Fotopoulou, F.; Kotecha, R.S.; Anderson, R.A.; Finch, A.J.; Ottersbach, K. Defining the fetal origin of MLL-AF4 infant leukemia highlights specific fatty acid requirements. *Cell Rep.* **2021**, *37*, 109900. [CrossRef]
49. Boiers, C.; Carrelha, J.; Lutteropp, M.; Luc, S.; Green, J.C.; Azzoni, E.; Woll, P.S.; Mead, A.J.; Hultquist, A.; Swiers, G.; et al. Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells. *Cell Stem Cell* **2013**, *13*, 535–548. [CrossRef]
50. Malouf, C.; Ottersbach, K. The fetal liver lymphoid-primed multipotent progenitor provides the prerequisites for the initiation of t(4;11) MLL-AF4 infant leukemia. *Haematologica* **2018**, *103*, e571–e574. [CrossRef]

51. Ding, J.; Cardoso, A.A.; Yoshimoto, M.; Kobayashi, M. The Earliest T-Precursors in the Mouse Embryo Are Susceptible to Leukemic Transformation. *Front. Cell Dev. Biol.* **2021**, *9*, 634151. [CrossRef]
52. Azzoni, E.; Frontera, V.; Anselmi, G.; Rode, C.; James, C.; Deltcheva, E.M.; Demian, A.S.; Brown, J.; Barone, C.; Patelli, A.; et al. The onset of circulation triggers a metabolic switch required for endothelial to hematopoietic transition. *Cell Rep.* **2021**, *37*, 110103. [CrossRef]
53. Hadland, B.K.; Huppert, S.S.; Kanungo, J.; Xue, Y.; Jiang, R.; Gridley, T.; Conlon, R.A.; Cheng, A.M.; Kopan, R.; Longmore, G.D. A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood* **2004**, *104*, 3097–3105. [CrossRef]
54. Espin-Palazon, R.; Weijts, B.; Mulero, V.; Traver, D. Proinflammatory Signals as Fuel for the Fire of Hematopoietic Stem Cell Emergence. *Trends Cell Biol.* **2018**, *28*, 58–66. [CrossRef]

Supplementary Figures and Legends

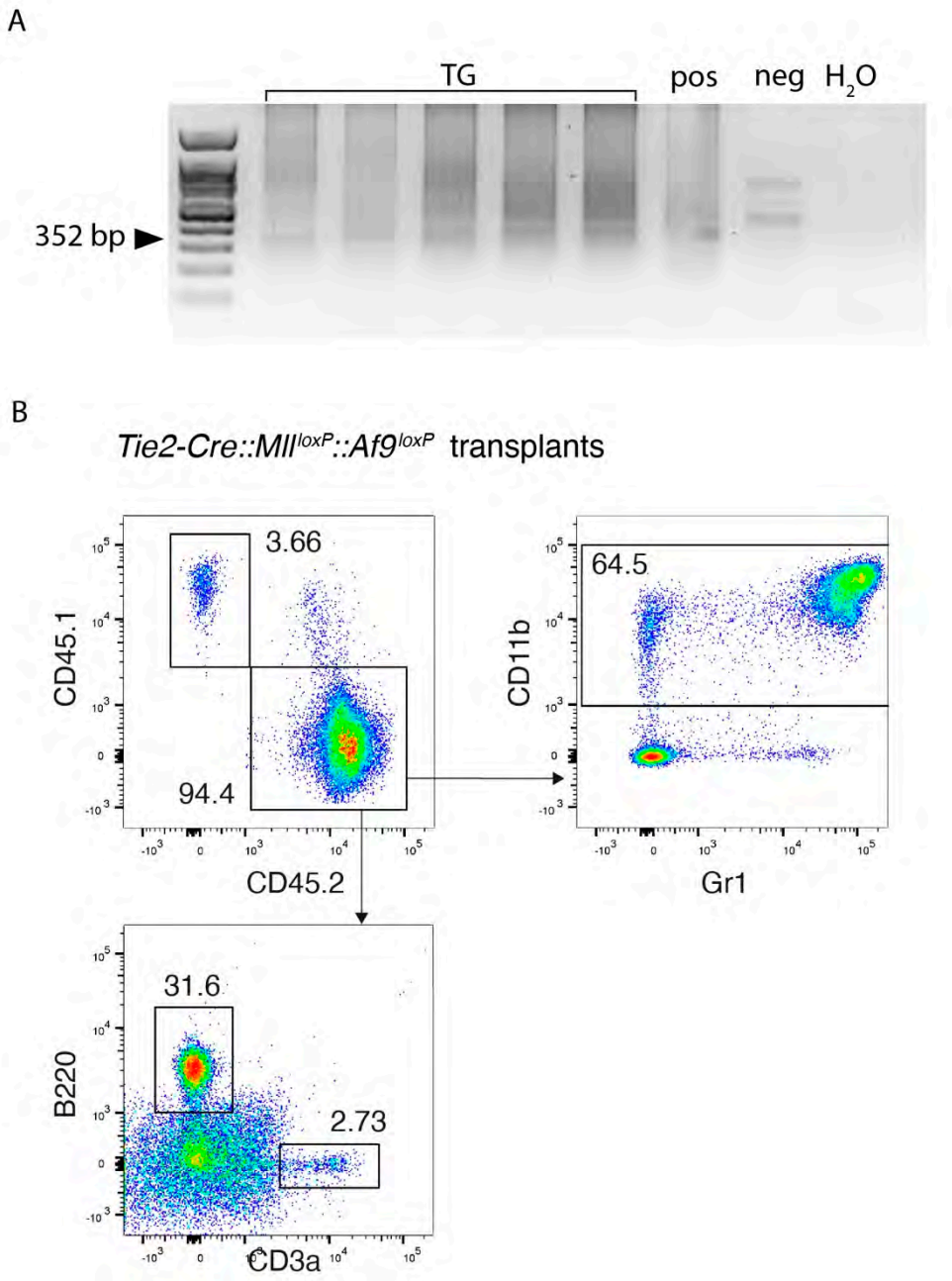


Figure S1. Additional analysis of *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* mice and transplants.

(A) Agarose gel showing the presence of the expected size amplicon (black arrow head) corresponding to the Mll-Af9 translocation in the bone marrow (BM) of *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* mice (TG). Positive (pos) and negative (neg) controls are shown.
(B) Representative flow cytometric peripheral blood (PB) analysis of lethally irradiated mice transplanted with BM cells derived from *Tie2-Cre::Mll^{loxP}::Af9^{loxP}* adult mice. Donor-derived chimerism (CD45.1 vs CD45.2, top left), myeloid cells (CD11b vs Gr1, top right), B and T lymphocytes (B220 vs CD3a, bottom) were assessed.

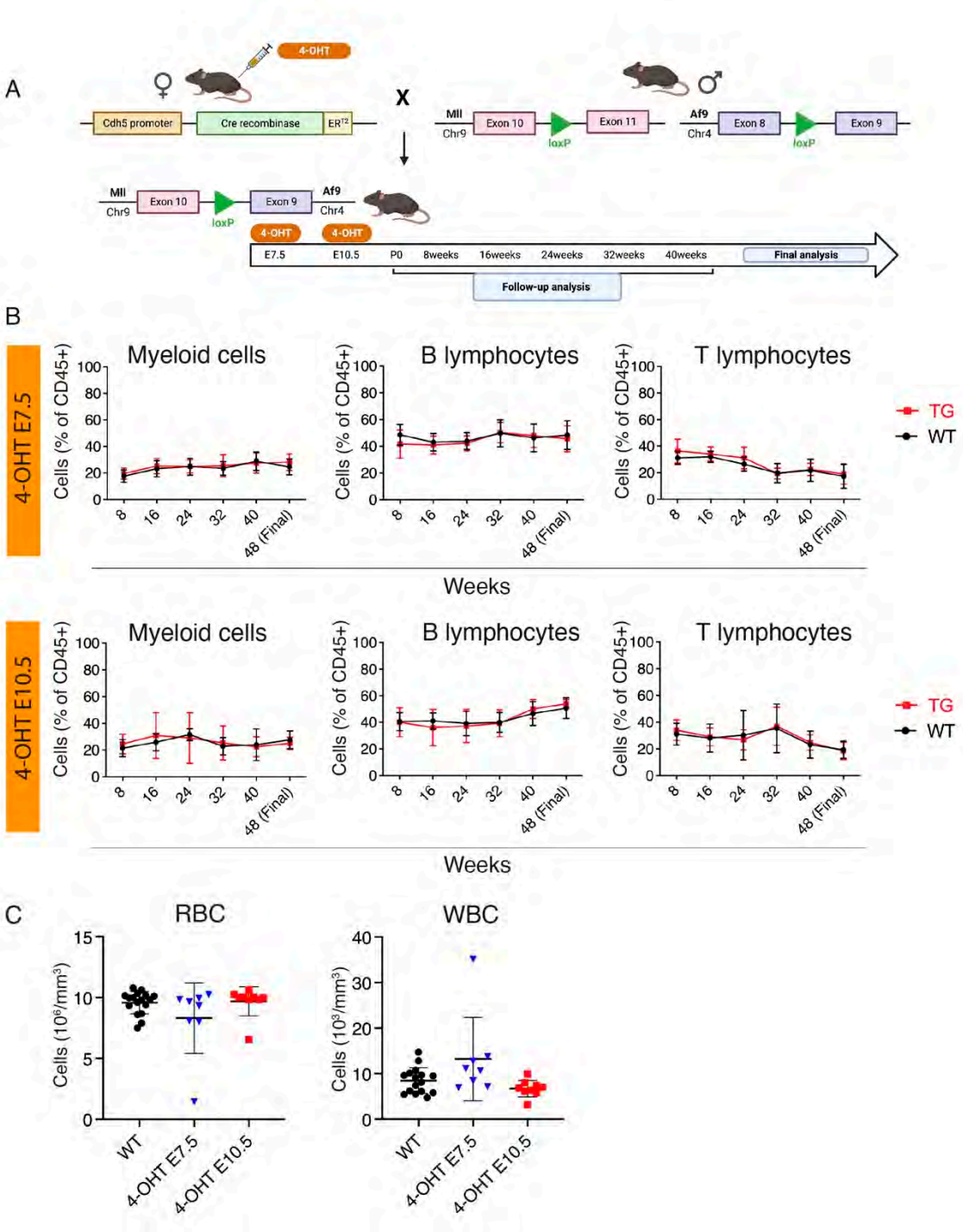


Figure S2. Analysis of unperturbed adult *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* mice.

(A) Schematic of crossing strategy, Cre activation (4-OHT at E7.5 or E10.5) and analyses of adult *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* mice.
(B) Quantification of flow cytometric analysis of myeloid (left), B (middle) and T lymphocytes (right) in WT (black) and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* (TG, red) unperturbed adult mice PB, with the administration of 4-OHT at E7.5 (top) or E10.5 (bottom). Data are mean ± SD.
(C) Red blood cells (RBC, left) and white blood cells (WBC, right) absolute counts of 48 weeks-old WT (black) and *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* adult mice with the administration of 4-OHT at E7.5 (blue) or E10.5 (red). Data are mean ± SD.

A

	FL		BM primary tx		BM secondary tx		Fraction of Mll-AF9 pos mice
	Mll-AF9 positive	Total	Mll-AF9 positive	Total	Mll-AF9 positive	Total	
Tx 4-OHT E7.5	6	6	6	6	5	8	17/20 (85%)
Tx Tam E10.5	4	8	2	8	4	5	10/21 (47.6%)

B

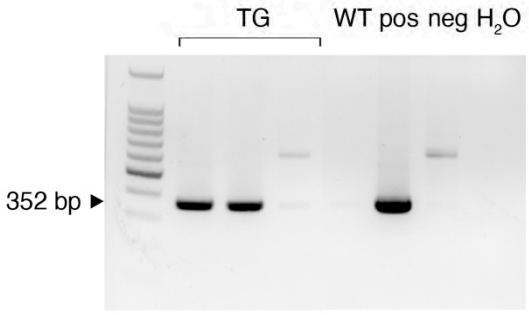


Figure S3. Detection of Mll-Af9 translocation in *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* transplants.

(A) Table representing the number of samples, per condition and total, used for primary and secondary transplantation analysis of *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* (4-OHT at E7.5 or Tamoxifen at E10.5) FL and BM. This table summarizes the number of samples in which the Mll-Af9 translocation was detected by PCR. The "BM primary tx" and "BM secondary tx" columns refer to the analysis done on the BM of transplant recipients.
(B) Representative agarose gel showing the presence of the expected 352-bp amplicon corresponding to the Mll-Af9 inter-chromosomal translocation in 2/3 *Cdh5-CreER^{T2}::Mll^{loxP}::Af9^{loxP}* transgenic (TG) FL samples (Tamoxifen at E10.5). Positive control (pos), wild type (WT) and negative (neg) controls are shown.

Primer name	Sequence (5'-3')	Cre/CRISPR
1A_Mll-intron_FW	GTC CCC ATA ACA CCC AGA GTA GTG	Cre
1B_AF9-intron_Rev	CCT CAT TCT GAC AGA CCA GAG CCA	Cre
2A_Mll-intron_FW	GGG CAT GTA GAG GTA AGA CGC CTG	Cre
2B_AF9-intron_Rev	ATC TCC AGG GAC TGA ATC TAG GGC	Cre
Mll_int10-CRISPR_FW	AAA CCA ACA GCA ACC CTT TTT	CRISPR
Mll_int10-CRISPR_Rev	GCA GTG GGC ATG TAG AGG TAA	CRISPR
AF9_int8-CRISPR_FW	TTT GTT CCC ATC ACA TCT GCC	CRISPR
AF9_int8-CRISPR_Rev	GGC AGC CAC TCA GTA ACT TG	CRISPR

Table S1. Primer sequences used for Mll-Af9 detection by PCR.
The third column indicates if primers were used to detect Cre/LoxP-induced or CRISPR/Cas9-induced translocation.

Conclusion

The studies presented in this thesis provide new insight into the complex landscape of developmental hematopoiesis in health and disease, from the origin and the specification of different hematopoietic stem and progenitor cells (HSPCs) to their leukemic transformation.

The first part of this thesis (Neo, Fadlullah et al., 2025; *Nature Cardiovascular Research*) focused on the endothelial to hematopoietic transition (EHT), and the mechanisms regulating the HSPCs fate choice. By using single-cell transcriptomics (including full-length transcriptome sequencing), whole-mount confocal imaging and functional assays, we identified and anatomically localized three distinct trajectories through which HE gives rise to specialized progenitor populations, each associated with a unique molecular and transcriptional signature. Importantly, we identified significant differences in the expression of chromatin modifiers and spliceosome components in different HE progenitor subsets, which correlated with distinct isoform expression patterns, including multiple stemness associated factors. We functionally validated the role of one of these isoforms (Runx1 Δexon6), which was found to be required for HSC emergence. Our data suggest that hematopoietic fate decisions that govern embryonic HSC fate could be, at least in part, be regulated through the expression of specific isoforms.

Through a combination of genetic lineage tracing, whole-mount imaging, and single-cell transcriptomic analyses, we characterized the contribution of different subsets of HSPC precursors to fetal and postnatal hematopoiesis (Barone et al., *Biorxiv*, 2024; under revision). Notably, we showed that a wave of fetal-restricted HSPCs arises from the HE in the vitelline and umbilical arteries, between E8.5 and E9.5, an intermediate time window between the emergence of YS-derived EMPs and adult-type HSCs. Although transient, this wave exerts a significant contribution to fetal lympho-myelopoiesis, and it is characterized by transcriptional and functional programs distinct from those of adult-type HSCs. These findings emphasize the complexity and heterogeneity of embryonic hematopoiesis, demonstrating that multiple waves of progenitors coexist, emerge in a spatially and temporally segregated manner, and are intrinsically programmed toward specific fates.

Understanding the identity and specific roles of embryonic HSPCs is not only important for developmental biology but can also have significant implications for the study of pediatric leukemias. Indeed, many hematological malignancies of the infancy were shown to originate *in utero*, therefore underlying the need of gaining a deeper knowledge of normal prenatal hematopoiesis, which can serve as a landmark to investigate its pathological counterpart. To this purpose, we focused on the

study of juvenile myelomonocytic leukemia (JMML). In order to investigate the prenatal origin of the disease, we combined the lineage tracing approach that we previously validated with the targeting of JMML-associated mutation $Kras^{G12D}$. We generated a new murine model (Cdh5-$CreER^{T2}$::LSL-$Kras^{G12D}$::$R26^{TdTomato}$) in which the $Kras^{G12D}$ mutation was specifically targeted to distinct subsets of fetal HSPCs at defined developmental time points, corresponding to their emergence from hemogenic endothelium.

Our data show that the cellular origin of $Kras^{G12D}$-driven JMML is not limited to the fetal HSCs compartment, and interestingly, the fetal-restricted HSPCs that we identified, could be responsible for JMML initiation. However, the resulting phenotype showed marked differences, suggesting that clinical heterogeneity could be related to differences in the cell of origin in which the RAS driving mutation first occurs. Moreover, we were able to identify a pre-leukemic state of the disease, further confirming its prenatal origin. Our transcriptomic and epigenetic analysis showed that fetal HSPCs undergo a Kras-driven transcriptional and epigenetic reprogramming already apparent at the fetal liver stage, which includes a metabolic rewiring with hypoxia-related pathway expression and the presence of Nfkb1-driven inflammation. Interestingly, a broad inflammatory signature was also detected in adult BM, thus highlighting a new pathway potentially amenable to therapeutic targeting. Based on these findings, we are currently testing anti-inflammatory treatments in this model, which have shown promising results in ameliorating disease features. We also plan to extend these studies to xenograft models of JMML, which we have successfully established in our laboratory. If successful, this strategy could complement existing treatments, potentially improving their efficacy. A manuscript reporting these results is currently in preparation (Quattrini et al.).

With a similar approach, we demonstrated that HSC-independent progenitors can also act as cells of origin of AML (Barone et al., 2023; *Cancers*). By using a conditional Cre-Lox transgenic mouse, we engineered the Mll-Af9 translocation, commonly found in pediatric AML, in defined subsets of fetal HSPCs. The selective targeting of Mll-Af9 in HSCs-independent progenitors, but not in HSCs, resulted in a transplantable myeloid leukemia. By using a CRISPR-Cas9-based approach, we also showed that Mll-Af9 rearrangement ex vivo in HSC-independent progenitors activate an abnormal myeloid-biased self-renewal program.

Overall, our work refines the current knowledge of embryonic hematopoiesis, as a complex and multi-layered process and highlights how transcriptional programs including specific isoforms of key genes converge to define fetal HSPCs identity and their contribution to embryonic and postnatal hematopoiesis. In addition, it also provides a critical framework and a new preclinical model that can be used to investigate the pathogenesis of JMML. By integrating developmental and pathological

perspectives, our findings open new avenues for understanding the early pathological events, leading different HSPCs populations to leukemic transformation and for identifying potential targets for early intervention and improving patients outcome.