Title

MLLT3 Isoforms Regulate Hematopoietic Stem Cell Maturation and Fate Decisions

Author

Vavilina-Halstead, Anastasia

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

MLLT3 Isoforms Regulate Hematopoietic Stem Cell Maturation and Fate Decisions

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Anastasia Dmitrievna Vavilina-Halstead

2024

ABSTRACT OF THE DISSERTATION

MLLT3 Isoforms Regulate Hematopoietic Stem Cell Maturation and Fate Decisions

by

Anastasia Dmitrievna Vavilina-Halstead

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2024

Professor Hanna K.A. Mikkola, Chair

Hematopoietic stem cell (HSC) transplantation can cure life-threatening blood disorders such as leukemias and inherited blood diseases. However, shortage of HLA-matched donors, which disproportionally affects minorities and patients of mixed ethnic backgrounds, limits the number of patients that can be treated. *Ex vivo* expansion or *de novo* generation of transplantable human HSCs has not been successful due to poor understanding of the basic biology underpinning key HSC traits - self-renewal, engraftment and multi-lineage differentiation ability (together referred to as "stemness"). Previous studies identified MLLT3 as a key regulator of stemness in human HSCs whose expansion declines in culture and differentiation, and demonstrated that maintaining MLLT3 expression in culture expands transplantable HSCs.

The focus of this is the characterization of a truncated isoform of MLLT3 . Analysis of RNA-seq data and epigenetic marks associated with the MLLT3 gene in human HSCs revealed a second TSS linked to a novel MLLT3 isoform (MLLT3-S), which encodes a truncated protein that can interact with known MLLT3 protein partners such as the Superelongation Complex (SEC) and Dot1L, but is unable to bind chromatin. MLLT3-L and MLLT3-S expression has opposing effects on gene expression and expansion of human HSCs in culture, suggesting distinct but complementary roles for the two isoforms. However, both isoforms of MLLT3 are necessary for proper HSC function. The MLLT3-S enhancer is accessible prior to its induction in fetal liver (FL) HSCs during their maturation, and coincides with downregulation of IGFBP2 expression. IGFBP2 is typically downregulated during HSC maturation, but its renewed expression is required for HSC proliferation in culture and MLLT3-L driven HSC expansion. The interplay between long and short isoforms of MLLT3 in human HSCs may provide a mechanism by which mature HSCs balance between expansion and maintenance modes.

The dissertation of Anastasia Vavilina-Halstead is approved.

Donald Barry Kohn

Tracy L. Johnson

John P. Chute

Hanna KA Mikkola, Committee Chair

University of California ,Los Angeles

2024

DEDICATION

I'd like to dedicate this thesis to Mom and Dad, for your unwavering support; to Tom, for being my rock; to Dr. Vincenzo Calvanese and all my labmates for showing me how it's done; and to Dr. Joe Pogliano and Dr. Alan Derman, for starting me on this journey.

A very special thank you to Dr. Hanna Mikkola, for being my inspiration and my very patient mentor throughout this work.

I couldn't have done it without you!

TABLE OF CONTENTS

Preliminary Pages:

Dissertation Research:

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

LIST OF TABLES

LIST OF ABBREVIATIONS

AGM	**Aorta-Gonad-Mesonephrous**
AHD	ANC1 Homology Domain
BM	Bone marrow
CB	(Umbilical) cord blood
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CS	Carnegie stage
DotCom	DOT1L complex
FL	Fetal liver
HE	Hemogenic endothelium
HLA	Human leukocyte antigen
HPC	Hematopoietic progenitor cell
HSC	Hematopoietic stem cell
HSPC	Hematopoietic stem/progenitor cell
KD	Knockdown
MHC	Major compatibility complex
NK	Natural killer cell
OE	Overexpression
PSC	Pluripotent stem cell
SEC	SuperElongation Complex
TSS	Transcription start site
UTR	Untranslated region

ACKNOWLEDGEMENTS

Chapters 2 and 3 are an adaptation of **Vavilina A**, Calvanese V, Fares I, Colombo G, Ma F, Capellera-Garcia S, Aguadé-Gorgorió J, Perrod C, Liebscher S, Wang Y, Shin J, Goodridge H, Crooks G, Wohlschlegel J, Schenke-Layland K, Mikkola HKA, "Developmentally regulated long and short isoforms of MLLT3 balance human hematopoietic stem cell fate decisions". *Manuscript in preparation*

Author contributions: A.V, V.C., J.A.G, G.C and H.K.A.M. designed experiments and interpreted data. Bioinformatic analysis was performed by V.C., A.V. and F.M. A.V, V.C. G.C and I.F. performed and/or supervised wet lab experiments and related data analysis. S.L coordinated fetal tissue collection. G.C and J.W assisted with data interpretation. A.V, V.C. and H.K.A.M. wrote the manuscript, which all authors edited and approved.

Correspondence and requests for materials should be addressed to H.K.A.M

Appendix 1 is a version of Calvanese V, Nguyen AT, Bolan TJ, **Vavilina A**, Su T, Lee LK, Wang Y, Lay FD, Magnusson M, Crooks GM, Kurdistani SK. "MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment". *Nature 2019* Dec;576(7786):281-6.

Appendix 2 is a version of Calvanese V, Capellera-Garcia S, Ma F, Fares I, Liebscher S, Ng ES, Ekstrand S, Aguadé-Gorgorió J, Vavilina A, Lefaudeux D, Nadel B, Li JY, Wang Y, Lee LK, Ardehali R, Iruela-Arispe MR, Pellegrini M, Stanley EG, Elefanty AG, Schenke-Layland K, Mikkola HKA. "Mapping human hematopoietic stem cells from hemogenic endothelium to birth." *Nature 2022* Apr;604(7906):534-40

I would like to thank the BSCRC FACS and sequencing cores, and TPCL, TCGB and CFAR cores (NIH AI028697-21) at UCLA. I thank Yanling Wang, Julia Shin, for assistance with harvesting cord blood, and Valerie Rezek for assistance with the breeding and transplantation of HSCs into NBSGW mice.

This work was supported by Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Jonsson Cancer Center Foundation and UCLA David Geffen School of Medicine Regenerative Medicine Theme Award for H.K.A.M.; NIH 1RO1DK125097 for HKAM, Swiss National Science Foundation P2ZHP3_178113 and EMBO ALTF 433-2019 for J.A.G, BSCRC post-doctoral fellowships for J.A.G and I.F, and T32 HL-086345-13 Developmental Hematology fellowship for I.F and B.N., BSCRC Rose Hills Foundation Graduate Training Program and Ruth L. Kirschstein National Research Service Award T32HL069766 for A.V., and Deutsche Forschungsgemeinschaft Cluster of Excellence iFIT (EXC 2180-390900677) for K.S-L.

Biographical Sketch:

Anastasia Vavilina-Halstead

EDUCATION

2012-2016 BS in Biochemistry and Cell Biology, Cum Laude

PUBLICATIONS:

Calvanese V, Capellera-Garcia S, Ma F, Fares I, Liebscher S, Ng ES, Ekstrand S, Aguadé-Gorgorió J, **Vavilina A**, Lefaudeux D, Nadel B, Li JY, Wang Y, Lee LK, Ardehali R, Iruela-Arispe MR, Pellegrini M, Stanley EG, Elefanty AG, Schenke-Layland K, Mikkola HKA. "Mapping human hematopoietic stem cells from hemogenic endothelium to birth." Nature 2022 Apr;604(7906):534-40

Calvanese V, Nguyen AT, Bolan TJ, **Vavilina A**, Su T, Lee LK, Wang Y, Lay FD, Magnusson M, Crooks GM, Kurdistani SK. "MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment". Nature 2019 Dec;576(7786):281-6.

Fung TC, Vuong HE, Luna CD, Pronovost GN, Aleksandrova AA, Riley NG, **Vavilina A**, McGinn J, Rendon T, Forrest LR, Hsiao EY. "Intestinal serotonin and fluoxetine exposure modulate bacterial colonization in the gut". Nature microbiology 2019 Dec;4(12):2064-73.

Chaikeeratisak V, Khanna K, Nguyen K, Sugie J, MacKennon ME, Erb ML, **Vavilina A**, Nonejuie P, Pogliano J "Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria" Cell 2019 Jun 13;177(7):1771-80

Chaikeeratisak V, Nguyen K, Khanna K, Brilot AF, Erb ML, Coker JK, **Vavilina A**, Newton GL, Buschauer R, Pogliano K, Villa E."Assembly of a nucleus-like structure during viral replication in bacteria" Science 2017 Jan 13;355(6321):194-7.

Chaikeeratisak V, Nguyen K, Egan ME, Erb ML, **Vavilina A**, Pogliano J. "The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages". Cell reports 2017 Aug 15;20(7):1563-71.

Pope WH, Bowman CA, Russell DA, Jacobs-Sera D, Asai DJ, Cresawn SG, Jacobs Jr WR, Hendrix RW, Lawrence JG, Hatfull GF, Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science, Phage Hunters Integrating Research and Education, and Mycobacterial Genetics Course. "Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity". Elife 2015;4.

- I was part of the Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science [SEA-PHAGES] group that isolated bacteriophages for the study and carried out preliminary characterization of isolated phage

FELLOWSHIPS AND HONORS:

2012-2016 UCSD Provost Honors

2017 Honorable Mention – NSF Graduate Research Fellowship

2016-2018 UCLA Dean's Scholar Award

2018-2021 Vascular Biology Training Grant

2018-2019 BSCRC Training Grant (awarded but declined due to overlap)

2018-2019 CMB Training Grant (awarded but declined due to overlap)

2020-2021 Whitcome Training Grant

2020-2021 BSCRC Training Grant (awarded but declined due to overlap)

2021-2023 BSCRC Training Grant

WORK & EXTRACURRICULARS

Volunteer, Skype A Scientist a April 2020 - present
Teaching Assistant, Stem Cell Biology Winter 2019
Teaching Assistant, Stem Cell Biology Winter 2018
BMSIS Young Scientist Program, NASA Ames Research Center March 2015 – May 2016
Undergraduate Instructional Apprentice, Cell Biology Winter 2016
Undergraduate Instructional Apprentice, Structural Biochemistry Fall 2015
Space Life Sciences Training Program, NASA Ames Research Center Summer 2014
Class Representative for the SEA-PHAGES Research Conference June 2013
Volunteer, Ostional, Costa Rica sea turtle research station June-July 2012

Chapter 1

Introduction

Preface

Since their discovery, stem cells have shown great promise, offering a multifaceted approach to revolutionize our understanding and treatment of many diseases. Stem cells have been explored for use in cell therapies for an ever-growing number of diseases and injuries, such as Parkinson's Disease, ALS, heart failure and spinal cord injuries[1]. They also offer a promising platform for disease modeling and drug development[2,3]. The stem cell type most frequently used for clinical treatments are hematopoietic stem cells (HSCs)[4]. HSCs are specified early in embryonic development and are the life-long source of all cell types in blood. Hematopoietic stem cell transplantation has long been used to cure blood disorders such as leukemias and inherited immunodeficiencies[5]. However, shortage of human leukocyte antigen (HLA)-matched donors, limits the number of patients who can be treated, especially among ethnic minority groups[6]. The development of protocols for expansion of transplantable human HSCs in *ex vivo* culture without sacrificing their long-term health and engraftability could ameliorate this shortfall, but has so far met with only limited success due to an incomplete understanding of the factors that maintain HSC self-renewal and regulate HSC fate decisions, as well as approaches for sustaining the expression of these factors *ex vivo*[7]. Adult HSCs can switch between quiescence and expansion as needed in response to intrinsic and extrinsic signals without losing the capacity for multilineage differentiation or exhausting the stem cell pool. Attempts to recapitulate these changes of cell state in culture currently result in loss of self-renewal and transplantability[8]. For this reason, it is essential to define the mechanisms underpinning HSC fate decisions and identify how these mechanisms are maintained. In addition to providing new mechanistic insight into the fundamental biology of HSCs, our findings could potentially be used to help establish or maintain HSC function while expanding or manipulating human HSCs in culture prior to transplantation.

Another potential source of transplantable HSCs could be de novo generation of HSCs from human pluripotent stem cells (PSCs). However, current methods for producing PSC-derived HSCs result in cells that retain many traits of fetal HSCs and are unable to recapitulate the maturation HSCs undergo during development[9-13]. Directed differentiation is unable to establish the mechanisms needed for long-term reconstituting human HSCs. Our work seeks to not only identify factors regulating HSC fate decisions, but also to uncover when these mechanisms are established and how they contribute to HSC maturation.

1.1 – The Hematopoietic Hierarchy

Hematopoietic stem cells are tissue-specific multipotent stem cells that reside at the top of the hematopoietic hierarchy[14]. The tissues in which HSCs are found, and which provide the microenvironment to maintain HSCs in their undifferentiated state, are known as the niche. Bone marrow is the main HSC niche in the adult[4]. Adult HSCs are primarily quiescent but can re-enter the cell cycle and divide either symmetrically, whereupon both daughter cells remain HSCs and facilitate the self-renewal and/or expansion of the stem cell pool, or asymmetrically, whereupon one daughter cell remains an HSC while the other daughter cell begins to differentiate and becomes a hematopoietic progenitor cell (HPC) to maintain active production of blood cells[15,16]. Crucially, dividing HSCs can also return to quiescence to prevent exhaustion of the HSC pool.

Although HPCs retain the multipotency of HSCs, they are more proliferative, and are unable to self-renew or provide long-term engraftment upon transplantation[14]. HPCs give rise to lineage-restricted progenitors, such as common lymphoid progenitors (CLP) and common myeloid progenitors (CMP), downstream of which are increasingly committed progenitor cells that

ultimately produce terminally differentiated cells with highly specialized functions. The lymphoid lineage gives rise to natural killer (NK) cells, T-cells, and B-cells, whereas the myeloid lineage gives rise to erythrocytes, platelets, macrophages, and granulocytes.

A common method for the identification and isolation of hematopoietic cell types is through surface marker expression. The currently accepted definition of a human HSC is positive for the markers CD34, CD90/THY1, and CD49f/ITGA6, has low expression of CD38, and is negative for CD45RA [17]. This population can further be enriched for long-term repopulating HSCs in cord blood and fetal liver samples through the expression of EPCR/CD201. Additionally, GPI80/VNN2 can be used to enrich undifferentiated HSCs in fetal liver samples[18]. However, GPI80 has been confirmed to enrich for transplantable HSCs only in uncultured samples, and have not yet been confirmed to do so in *ex vivo* expanded cells, and not all EPCR+ cells are engraftable HSCs. The population of cells characterized by this immunophenotype is still heterogenous, and only a fraction of them have proven to be true long-term repopulating HSCs upon transplantation. Furthermore, while it is possible to generate hPSC-derived HSCs that are CD34+ CD38-CD90+, these cells remain functionally immature and do not contain self-renewing, robustly transplantable HSCs[9,18,19].

Figure 1.1

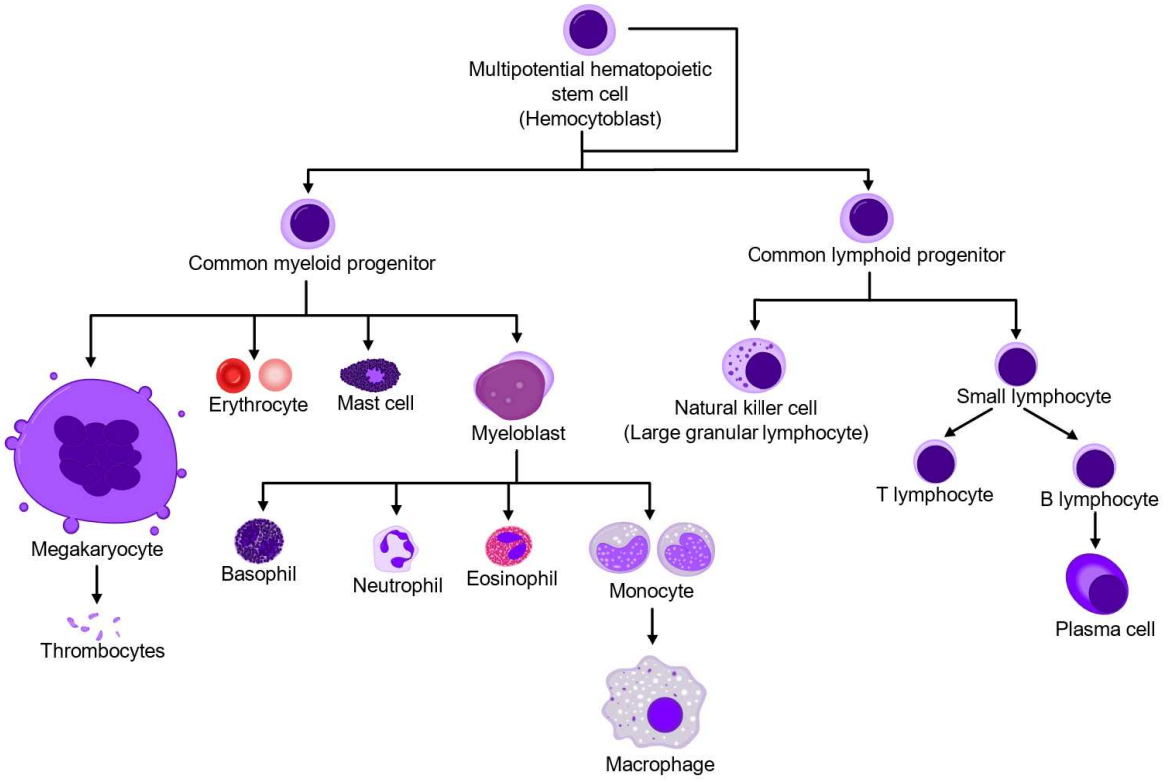

Figure 1.1 The Hematopoietic Hierarchy

Diagram of the hematopoietic hierarchy, showing self-renewing, multipotent HSCs and their progressively more differentiated progeny in the lymphoid and myeloid lineages. By A. Rad and M. Häggström. *released under the Attribution-Share Alike 3.0 Unported license*

1.2 – Developmental Hematopoiesis

The development of the hematopoietic system can be broken up into three waves, of which only the last leads to the emergence of definitive HSCs capable of maintaining lifelong hematopoiesis. Even after the definitive HSCs emerge, they are immature and must undergo maturation before they resemble adult HSCs[9]. This process is broadly preserved in vertebrates, although the exact sites of HSC generation and maturation in some model organisms, such as

zebrafish) differ from those in mammals[10]. In mammalian development, hematopoiesis occurs sequentially in the yolk sack, the aorta-gonad-mesonephrous (AGM) region, the placenta, the fetal liver, and finally the bone marrow.

The first wave is generated by primitive hematopoietic progenitors originating from the yolk sac before the start of circulation, at Carnegie stage (CS) 7-8 of human development. Their potential is restricted to the myeloid lineage, and their primary purpose is to rapidly generate primitive erythroblasts, although they also give rise to megakaryocytes and macrophages [10]. The primitive erythroblasts, which are initially nucleated, enter circulation with the initiation of the heart beat (CS 10), whereupon they make their way to the placental vasculature and exit into the villous stroma. Unlike definitive erythroblasts, which mature and enucleate in the site where they are generated (e.g. fetal liver or bone marrow), primitive erythroblasts first enter circulation and enucleate after arriving in the villous stroma with the aid of placental macrophages (Hofbauer cells). These macrophages are generated in the placenta, suggesting that pre-circulation macrophages may arise from several sources[20].

The second wave of hematopoiesis also consists of HSC-independent transient-definitive progenitors emerging in the yolk sac. Transient definitive progenitors are still unable to self-renew, but are partially multipotent, although most of them have limited lymphoid potential[21-23].

The third wave of hematopoiesis consists of the emergence and maturation of definitive HSCs that replace previous progenitors. Definitive HSCs are self-renewing, able to produce all blood cell types, and engraft upon transplantation (the gold standard for defining a "true" functional HSC). The process of definitive HSC specification and emergence takes place in the AGM, on the ventral side of the dorsal aorta. Following HOXA patterning, aortic endothelial cells in this region begin to lose their aortic identity and transition first to pre-hemogenic endothelial

cells, identifiable by their expression of IL33, ALDH1A1, and DKK1, and then to hemogenic endothelium (HE) cells, marked by the additional upregulation of RUNX1 and KCNK17[9]. Definitive HSCs emerge from HE at CS14-16, via intra-aortic hemogenic cluster[24,25], and contribute significantly to human hematopoiesis from 6-8 developmental weeks onwards. Transplantable cells are initially very rare in both mouse and human embryos, although single-cell studies in embryos of both species identified an HSC population much larger than transplantation assays suggest, implying that the functional immaturity of AGM HSCs impairs their engraftment ability[24]. Following their emergence, nascent HSCs passage through the placenta and subsequently colonize the fetal liver, which is the main site of active hematopoiesis throughout the first and second trimester. After undergoing dramatic expansion and functional maturation that involves transition to a more quiescent, homeostatic state, HSCs embark on their final migration, this time into the bone marrow, where the HSC pool will remain through birth and later life.

scRNAseq transcriptome mapping of human hematopoietic tissues from the first trimester to birth revealed that definitive HSCs are distinguished from earlier hematopoietic progenitors by the gene signature RUNX1+ HOXA9+ MLLT3+ MECOM+ HLF+ SPINK2+. This effort also revealed that HSC maturation involves temporal shifts in gene expression, including the suppression of genes characteristic of immature HSCs, such as CDH5, MEIS2, IGFBP2, and HOXB9, and the upregulation of genes characteristic of mature HSCs, such as PROM1, HLA-DRA, HEMGN, and MSI2 [24]. The changing biological processes within maturing HSCs can be seen in the suppression of megakaryocytic and endothelial surface features, Fetal specific genes and proliferation genes, upregulation of HSC stemness regulators, such as MLLT3 and HLF, and acquisition of maturity surface markers such as PROM1/CD133 and MHC class II[24].

Figure 1.2

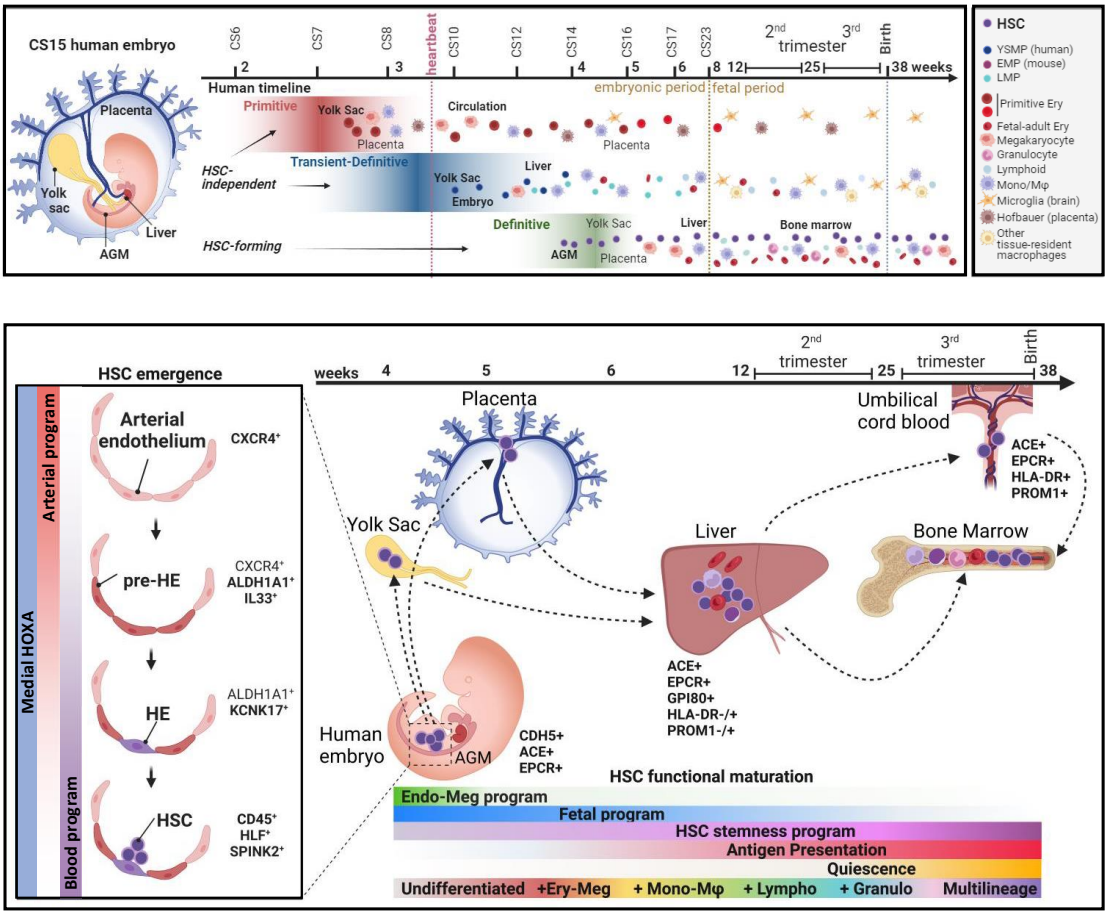

Figure 1.2 Waves of human hematopoiesis and the emergence of definitive HSCs

A. Diagram showing the three waves of hematopoiesis, their timing in respect to human development, the tissues in which the respective hematopoietic HSPCs emerge, and the differentiated cell types generated by each wave. An illustration depicting the anatomy of a human embryo at CS15 is included as a reference for the localization of hematopoietic tissues.

B. Schematic illustrating the establishment of HE in the AGM, followed by HSC specification and emergence. Key genes characteristic of each cell type and stage of development are indicated next to the cell types that express them. Colored bars indicate temporal changes in gene expression programs through this process.

\

1.3 – HSC Transcription Regulators

There are many transcription factors required sequentially for definitive HSC specification and emergence, as well as the maintenance and expansion of HSCs throughout development and later life. Most studies examining these factors have been done in animal (eg. mouse and zebrafish) models or in the context of directed differentiation of hPSCs. Early in HSC specification, SCL/tal1 and its protein partner, LMO2, as well as SOX17 are necessary for the specification of blood lineage from the mesoderm. Following the establishment of hemogenic endothelium, RUNX1, MLL (also known as MLL1, KMT2A, or ALL1), and Cdx4 have been demonstrated to be essential for HSC emergence[10].

In definitive HSCs, a plethora of other factors are needed to maintain HSC identity. The transcription factors HOXA9, HOXB4, HLF, and MECOM/EVI1 all contribute to HSC self-renewal and proliferation[24]. However, continued expression of some HSC specification factors, such as SCL and RUNX1, is not required for adult HSC survival and self-renewal, although it can affect the differentiation of subsequent progenitor cells. Furthermore, single-cell studies of HSCs and early progenitors revealed that these populations express low levels of factors characteristic of multiple differentiated lineages, a phenomenon called lineage priming. This highlights the heterogeneity of HSC populations, and suggests that expression of these factors is a way for stem cells to maintain the possibility of many alternative fates until further differentiation occurs.

Many HSC transcription factors were first discovered as their dysregulation or mutation was observed in hematologic malignancies. Deregulation of the SCL/tal1 locus leads to T cell acute leukemias, while GATA1 mutations are commonly found in cases of transient abnormal hematopoiesis and myeloid leukemia of Down syndrome[26]. Acute myeloid and lymphoid

leukemias can be driven by chromosomal translocations creating fusion proteins partnering MLL with other transcription regulators, including MLLT3/AF9, AFF4/AF4, ELL, and ENL[27-29].

Our previous research identified MLT3 as a novel upstream regulator of human HSC transcriptional program[30]. MLLT3 expression is downregulated both during HSC differentiation and culture, when HSCs lose self-renewal ability. MLLT3 (previously known as AF9) was first identified as a component of the superelongation complex (SEC), which regulates the rate of transcription elongation[31,32], and the DOT1L complex (DotCom), which catalyzes H3K79 di- and trimethylation to promote active transcription[29,33-35]. The MLLT3 protein has two domains critical to its function: an N-terminal YEATS domain which allows it to recognize activating histone marks (e.g.H3K9ac)[36], and a C-terminal ANC1 homology domain (AHD) that enables interaction with protein partners in MLLT3-containing complexes[29,37]. The intrinsically disordered structure of the AHD allows MLLT3 to interact with many different proteins, such as DOT1L, AFF1, AFF4, and MLLT10[29,31,32], and as shown more recently PRC1 Polycomb group proteins CBX8 and BCOR, at least in leukemias[38]. Notably, truncated MLL/MLLT3 fusion proteins, which combine the AHD of MLLT3 with the N-terminal region of MLL, are associated with aggressive leukemias, providing additional evidence that MLLT3 can be a powerful regulator of transcription in hematopoietic cells[28,29,38].

We demonstrated that sustaining MLLT3 expression at physiological levels using lentiviral overexpression (OE) vectors in human fetal liver (FL) and cord blood (CB) hematopoietic stem and progenitor cells (HSPCs) promoted symmetric self-renewal and allowed the expansion of transplantable HSCs without blocking their differentiation or inducing transformation to leukemia[30]. Knockdown (KD) of MLLT3 led to rapid loss of HSPCs in culture and rendered them unable to engraft. We found that MLLT3 binds the transcription start site (TSS) of key HSC

regulatory genes such as ERG, HLF, and MECOM, and showed that MLLT3 OE increased H3K79me2 in HSC regulatory genes bound by MLLT3 in cultured HSPCs. Importantly, MLLT3 binds genes in a cell-type specific manner, binding different sets of genes in HSPCs compared to erythroid cells. These data suggest that MLLT3 is not a pioneer factor that establishes new transcriptional programs, but rather maintains the transcription of genes activated by other factors, thus protecting the expression of HSC stemness programs when HSCs divide[30]. Examination of MLLT3 expression in scRNAseq of human hematopoietic tissues throughout development also showed that MLLT3 expression is present in hemogenic endothelium prior to HSC emergence and increased throughout HSC maturation, suggesting that not only is MLLT3 expression key to maintaining HSC cell identity upon ex vivo expansion, but also that HSC maturation involves amplification of MLLT3-driven transcription programs[24].

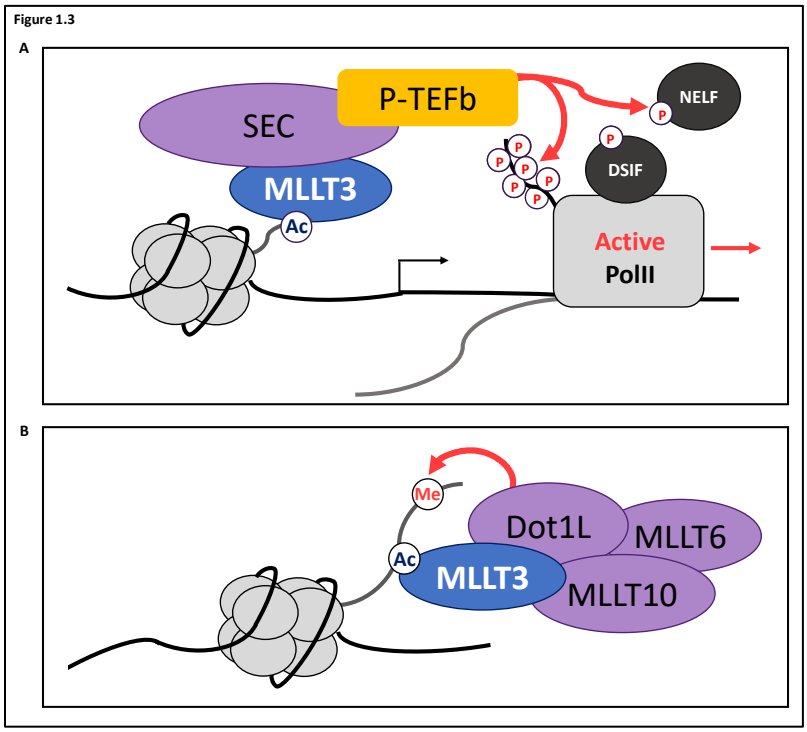

Figure 1.3 MLLT3 is a component of SEC and DotCom

Diagrams of the Superelongation Complex (SEC) (**A**) and Dot1L Complex (DotCom) (**B**), illustrating their major protein components and molecular function in regulating transcription

1.5 – Significance and Aims of This Dissertation

Key Question: ***How does MLLT3 promote HSC expansion without rendering them unable to respond to environmental cues prompting quiescence or differentiation?***

At the beginning of this study, our laboratory had made the initial discovery that human HSCs may express an alternate isoform of MLLT3. The goal of this thesis is to test the hypothesis that the newly identified second, truncated isoform of MLLT3 contributes to fine tuning the regulation of HSC fate decisions.

Aim 1 – Validate the existence of MLLT3-S

Firstly, we aim to examine which isoforms of MLLT3 are expressed in human HSCs and validate the expression of MLLT3-S transcript. We will also verify that MLLT3-S expression results in a stable protein that is necessary for human HSC function.

Aim 2 – Investigate the role of a truncated MLLT3 isoform (MLLT3-S) in human HSC fate decisions

We also aim to investigate the expression of a hitherto-uncharacterized truncated isoform of MLLT3 (MLLT3-S) in HSCs and their downstream progeny, and evaluate the role it plays in HSC fate decisions and transplantability. We will also investigate how MLLT3-S affects HSC transcriptional programs, and how the effects of its expression compare to those of full-length MLLT3 (MLLT3-L). This will expand our understanding of the mechanisms by which human HSCs retain self-renewal, engraftment, and differentiation ability while alternating between active and dormant states. Ultimately, greater understanding of these processes may lead to improved protocols for expanding HSCs in culture while maintain their function and transplantability.

Aim 3 – Investigate the role of MLLT3 isoforms in human HSC developmental maturation

Secondly, we intend to evaluate the expression of MLLT3-S and MLLT3-L in fetal hematopoietic tissues to determine at which stages of development each isoform is induced. We will also define how changes in MLLT3 isoform expression affect the expression of genes involved in HSC maturation and investigate the molecular mechanisms by which MLLT3 isoforms regulate expression of these genes. Through this, we will enhance our understanding of the regulation of HSC developmental maturation, which could lead to improved approaches for ensuring the functional maturation of PSC-derived HSCs.

Bibliography

1. Yamanaka, S. (2020). Pluripotent Stem Cell-Based Cell Therapy—Promise and Challenges. Cell Stem Cell *27*, 523-531. 10.1016/j.stem.2020.09.014.
2. Argentati, C., Tortorella, I., Bazzucchi, M., Morena, F., and Martino, S. (2020). Harnessing the Potential of Stem Cells for Disease Modeling: Progress and Promises. Journal of Personalized Medicine *10*, 8. 10.3390/jpm10010008.
3. Musunuru, K. (2013). Genome editing of human pluripotent stem cells to generate human cellular disease models. Disease Models & Mechanisms *6*, 896-904. 10.1242/dmm.012054.
4. Crane, G.M., Jeffery, E., and Morrison, S.J. (2017). Adult haematopoietic stem cell niches. Nature Reviews Immunology *17*, 573-590. 10.1038/nri.2017.53.
5. Bordignon, C. (2006). Stem-cell therapies for blood diseases. Nature *441*, 1100-1102. 10.1038/nature04962.

6. Mosaad, Y.M. (2014). Immunology of hematopoietic stem cell transplant. Immunological investigations *43*, 858-887.

7. Haltalli, M.L., Wilkinson, A.C., Rodriguez-Fraticelli, A., and Porteus, M. (2022). Hematopoietic stem cell gene editing and expansion: State-of-the-art technologies and recent applications. Experimental Hematology *107*, 9-13.

8. Magnusson, M., Sierra, M.I., Sasidharan, R., Prashad, S.L., Romero, M., Saarikoski, P., Van Handel, B., Huang, A., Li, X., and Mikkola, H.K.A. (2013). Expansion on Stromal Cells Preserves the Undifferentiated State of Human Hematopoietic Stem Cells Despite Compromised Reconstitution Ability. PLoS ONE *8*, e53912. 10.1371/journal.pone.0053912.

9. Calvanese, V., and Mikkola, H. (2023). The genesis of human hematopoietic stem cells. Blood Journal, blood. 2022017934.

10. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631-644. 10.1016/j.cell.2008.01.025.

11. Ding, J., Li, Y., and Larochelle, A. (2023). De Novo Generation of Human Hematopoietic Stem Cells from Pluripotent Stem Cells for Cellular Therapy. Cells *12*, 321. 10.3390/cells12020321.

12. Ivanovs, A., Rybtsov, S., Ng, E.S., Stanley, E.G., Elefanty, A.G., and Medvinsky, A. (2017). Human haematopoietic stem cell development: from the embryo to the dish. Development *144*, 2323-2337. 10.1242/dev.134866.

13. Ditadi, A., Sturgeon, C.M., and Keller, G. (2017). A view of human haematopoietic development from the Petri dish. Nature Reviews Molecular Cell Biology *18*, 56-67. 10.1038/nrm.2016.127.

14. Zhang, Y., Gao, S., Xia, J., and Liu, F. (2018). Hematopoietic hierarchy–an updated roadmap. Trends in cell biology *28*, 976-986.

15. Loeffler, D., and Schroeder, T. (2021). Symmetric and asymmetric activation of hematopoietic stem cells. Current Opinion in Hematology *28*, 262-268.

16. Ho, A.D., and Wagner, W. (2007). The beauty of asymmetry: asymmetric divisions and self-renewal in the haematopoietic system. Current opinion in hematology *14*, 330-336.

17. Fares, I., Calvanese, V., and Mikkola, H.K. (2022). Decoding Human Hematopoietic Stem Cell Self-Renewal. Current Stem Cell Reports *8*, 93-106.

18. Prashad, S.L., Calvanese, V., Yao, C.Y., Kaiser, J., Wang, Y., Sasidharan, R., Crooks, G., Magnusson, M., and Mikkola, H.K.A. (2015). GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. Cell stem cell *16*, 80-87.

19. Dou, D.R., Calvanese, V., Sierra, M.I., Nguyen, A.T., Minasian, A., Saarikoski, P., Sasidharan, R., Ramirez, C.M., Zack, J.A., Crooks, G.M., et al. (2016). Medial HOXA genes demarcate haematopoietic stem cell fate during human development. Nature Cell Biology *18*, 595-606. 10.1038/ncb3354.

20. Van Handel, B., Prashad, S.L., Hassanzadeh-Kiabi, N., Huang, A., Magnusson, M., Atanassova, B., Chen, A., Hamalainen, E.I., and Mikkola, H.K. (2010). The first trimester human placenta is a site for terminal maturation of primitive erythroid cells. Blood, The Journal of the American Society of Hematology *116*, 3321-3330.

21. Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., and Bian, L. (2019). Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. Cell research *29*, 881-894.

22. Ghosn, E., Yoshimoto, M., Nakauchi, H., Weissman, I.L., and Herzenberg, L.A. (2019). Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. Development *146*, dev170571.

23. Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., and Li, Z. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. Nature *533*, 487-492.

24. Calvanese, V., Capellera-Garcia, S., Ma, F., Fares, I., Liebscher, S., Ng, E.S., Ekstrand, S., Aguadé-Gorgorió, J., Vavilina, A., and Lefaudeux, D. (2022). Mapping human haematopoietic stem cells from haemogenic endothelium to birth. Nature *604*, 534-540.

25. Tavian, M., Hallais, M.-F., and Péault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. Development *126*, 793-803.

26. Kanezaki, R., Toki, T., Terui, K., Xu, G., Wang, R., Shimada, A., Hama, A., Kanegane, H., Kawakami, K., and Endo, M. (2010). Down syndrome and GATA1 mutations in transient abnormal myeloproliferative disorder: mutation classes correlate with progression to myeloid leukemia. Blood, The Journal of the American Society of Hematology *116*, 4631-4638.

27. Britten, Ragusa, Tosi, and Kamel (2019). MLL-Rearranged Acute Leukemia with t(4;11)(q21;q23)—Current Treatment Options. Is There a Role for CAR-T Cell Therapy? Cells *8*, 1341. 10.3390/cells8111341.

28. Mohan, M., Lin, C., Guest, E., and Shilatifard, A. (2010). Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nature reviews Cancer *10*, 721-728.

29. Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S.A., and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered transcriptional

regulator that recruits multiple partners via coupled folding and binding. Structure *21*, 176-183.

30. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., and Crooks, G.M. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281-286.

31. Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nature reviews Molecular cell biology *13*, 543-547.

32. He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., Alber, T., Benkirane, M., and Zhou, Q. (2011). Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. Proceedings of the National Academy of Sciences *108*, E636-E645.

33. Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes & development *24*, 574-589.

34. Kuntimaddi, A., Achille, N.J., Thorpe, J., Lokken, A.A., Singh, R., Hemenway, C.S., Adli, M., Zeleznik-Le, N.J., and Bushweller, J.H. (2015). Degree of recruitment of DOT1L to MLL-AF9 defines level of H3K79 Di-and tri-methylation on target genes and transformation potential. Cell reports *11*, 808-820.

35. Barry, E.R., Krueger, W., Jakuba, C.M., Veilleux, E., Ambrosi, D.J., Nelson, C.E., and Rasmussen, T.P. (2009). ES cell cycle progression and differentiation require the action of the histone methyltransferase Dot1L. Stem Cells *27*, 1538-1547.

36. Wu, H., Chen, L., Zhou, Q., and Zhang, W. (2011). AF17 facilitates Dot1a nuclear export and upregulates ENaC-mediated Na+ transport in renal collecting duct cells. PLoS One *6*, e27429.

37. Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y., and Li, W. (2014). AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. Cell *159*, 558-571.

38. Hu, H., and Muntean, A.G. (2023). The YEATS domain epigenetic reader proteins ENL and AF9 and their therapeutic value in leukemia. Experimental Hematology *124*, 15-21. https://doi.org/10.1016/j.exphem.2023.06.001.

Figure 1.1 By A. Rad and M. Häggström. *released under the Attribution-Share Alike 3.0 Unported license*

Chapter 2

Role of MLLT3 isoforms in human HSC function

Abstract:

MLLT3 is a key regulator of human hematopoietic stem cell (HSC) self-renewal but it's expression declines in culture[30]. Maintaining MLLT3 levels in cultured cord blood (CB) HSCs results in expansion of transplantable HSCs without transformation or differentiation bias. Analysis of RNA-sequencing data and epigenetic marks associated with the MLLT3 gene in human HSCs revealed a second TSS linked to a novel MLLT3 isoform (MLLT3-S). MLLT3-S encodes a truncated protein that retains the AHD domain responsible for protein-protein interactions[29,36], but lacks the chromatin binding YEATS domain[37,39]. MLLT3-S overexpression (OE) in hematopoietic cell lines confirmed the generation of a stable protein that interacts with known MLLT3 protein partners DOT1L[29,33-35] and superelongation complex (SEC)[31,32,40], but cannot bind chromatin. While MLLT3-L OE in CB HSCs promoted their expansion, MLLT3-S OE suppressed it, implying distinct functions for the two isoforms in HSCs. Conversely, while knockdown (KD) of MLLT3-L in CB HSCs triggered premature differentiation, MLLT3-S KD resulted in relative expansion of immunophenotypic HSCs. However, although MLLT3-S deficient HSPCs appeared phenotypically normal in culture, they engrafted poorly in immunodeficient mice, showing that both isoforms of MLLT3 are necessary for proper HSC function. scRNA and bulk RNA seq of HSPCs with MLLT3-L and/or MLLT3-S OE and KD revealed that the two isoforms have opposing effects on expression of genes regulating key HSC processes, such as mitochondrial biosynthesis and oxidative phosphorylation, translation, and splicing, as well as several HSC transcriptional regulators. The interplay between long and short isoforms of MLLT3 in human HSCs may provide a mechanism by which mature HSCs balance between expansion and maintenance modes.

Introduction:

It has been a long-standing goal to develop culture protocols that would enable ex vivo expansion or de novo generation of transplantable human HSCs[5,7]. To ensure life-long blood cell production while preserving a healthy HSC pool, HSCs must be able to alternate between different fates: quiescence, self-renewal (symmetric or asymmetric), mobilization/homing, differentiation, and death[15,16]. Adult HSCs can also enter cell cycle and expand in response to external stimuli, but returning to quiescence is crucial for preventing HSC exhaustion[8]. These HSC fate decisions are dictated by a myriad of intrinsic and extrinsic factors that need to be precisely regulated to maintain stemness and healthy HSC pool. However, the molecular machinery that enables human HSCs to alter their fate in response to signals in a reversible fashion and maintain stemness without transformation or differentiation block, remains poorly defined. As such, it has proven difficult to dictate these changes in HSC states in culture without compromising their self-renewal and engraftment ability[17,41].

By searching for the genes in human HSCs that decline when HSCs lose self-renewal ability during differentiation and culture, we previously identified MLLT3/AF9 as a master regulator of human HSC "stemness"[8,18,19,30]. Restoring MLLT3 expression in human cord blood (CB) HSCs during culture expanded transplantable HSCs without impairing their ability to differentiate. MLLT3's ability to sustain the expression of regulatory factors involved in HSC expansion (e.g. HSC transcription factors HLF and MECOM, protein synthesis machinery) was linked to interactions with the DOT1L complex[30], which maintains the active chromatin mark H3K79me2[29,33-36]. MLLT3 also interacts with the Superelongation complex (SEC), which regulates transcription elongation[31,32,37,39,40]. Here we show that, in addition to expressing the full-length MLLT3, human HSCs express a hitherto-unreported short form of MLLT3, MLLT3-S, that

is transcribed from its own TSS. Together with MLLT3-L, MLLT3-S regulates HSC gene expression and balances HSC expansion and maintenance modes.

Results:

MLLT3 gene harbors a second active TSS in human HSCs that promotes the expression of a short MLLT3 isoform (MLLT3-S)

To understand MLLT3 regulation in human HSCs, we sought to identify regulatory elements in the MLLT3 gene and performed ChIP-seq for histone marks in 2nd trimester human fetal liver (FL) CD34+CD38loCD90+ HSPCs. Analysis of epigenetic marks characteristic of active transcription start sites (TSS) (H3K4me3, H3K27ac and low H3K4me1) identified two potential promoters within the MLLT3 gene: the expected TSS at the 5' end of the transcript, and a putative alternative TSS between exons 5 and 6 **(Figure 2.1A**). The second TSS matched with an annotated short MLLT3 transcript predicted to contain six 3' exons (exons 6-11) of full length MLLT3 and a unique 5' untranslated region (alternative 5' exon 1). The isoform associated with the second TSS, hereby named MLLT3-S, is predicted to encode a truncated protein comprised of the 162 C-terminal amino acids of the full-length MLLT3 (hereby referred to as MLLT3-L) (**Figure 2.1B**). This truncated isoform lacks the sequences encoding the YEATS domain, which is critical for chromatin binding by MLLT3, but retains the ANC1 homology domain responsible for MLLT3 protein-protein interactions.

Examination of MLLT3 active regulatory regions in FL HSPCs by ATAC-seq revealed that the putative MLLT3-S TSS is accessible in the most undifferentiated CD34+CD38-CD90+GPI80+ FL HSPCs and becomes progressively less accessible in CD34+CD38-

CD90+GPI80- and CD34+CD38-CD90- downstream progeny, while the canonical MLLT3 TSS still remains accessible (**Figure 2.1A**). ATAC-seq also revealed an adjacent HSC-specific regulatory site categorized as an enhancer based on histone marks (low H3K4me3, high H3K4me1 and H3K27ac), that closed with similar kinetics as MLLT3-S TSS upon FL HSPC differentiation.

RNA-seq in human 2nd trimester FL and cord blood (CB) HSPCs demonstrated peaks corresponding to the unique 5' region of MLLT3-S (**Figure 2.1A**). The predicted MLLT3-S transcript in human FL HSPCs was also verified by RT-PCR, and further validated by cloning and sequencing using primers spanning the transcript from the alternative 5' exon 1 to exon 11 (**Figure 2.1C, Figure S2.1A**). Detailed analyses suggested that human FL HSPCs may express two different subvariant transcripts of MLLT3-S (7 or 8 exons) that have identical coding sequences (spanning exons 6-11) but differ by the length of their 5'UTR. While the 5' UTR of the 7 exon MLLT3-S (named MLLT3-S1) contains one unique exon, the 5' UTR of the 8 exon variant (named MLLT3-S2) contains two unique exons (Figure S2.1A, B). 5' RACE in human CB HSPCS confirmed the presence of both subvariants of MLLT3-S with the 7 exon variant being detected predominantly (**Figure S2.1C**). PacBio long read RNA sequencing analysis of 20 week fetal liver and cord blood CD34+CD38lo HSPCs confirmed that isoforms matching the sequences of MLLT3-L, MLLT3-S1, and MLLT3-S2 were the predominant MLLT3 isoforms expressed in these tissues (data not shown).

Lentiviral overexpression of either the predicted coding region or the two MLLT3-S subvariants with different length 5' UTR in hematopoietic cell lines resulted in a stable protein, as verified by Western blot of the V5-tagged truncated protein (**Figure 2.1E, F, Figure S2.1D, E**). Evaluation of MLLT3 isoform localization by subcellular fractionation showed that, in contrast to MLLT3-L, which was observed in both nuclear soluble and chromatin bound fractions, MLLT3-

S could only be detected in the nuclear soluble fraction, consistent with lack of a YEATS domain (**Figure 2.1F**).

RNA-seq and Q-RT-PCR of different hematopoietic subsets in 2nd trimester FL showed that MLLT3-S transcript was most abundant in undifferentiated FL HSPCs and became downregulated in more differentiated progenitor cells (**Figure 2.1D, Figure S2.1F**). Q-RT-PCR and RNA-seq data from freshly isolated and cultured CD34+CD38loCD90+ FL HSPCs showed that MLLT3-S is downregulated during HSC culture to even greater degree than MLLT3-L, even when different HSC supportive conditions were used (**Figure 2.1D, Figure S2.1G**). Together, these data evidence highly regulated, HSC-enriched expression of a short isoform of MLLT3, suggesting a specialized function in human HSCs.

Figure 2.1

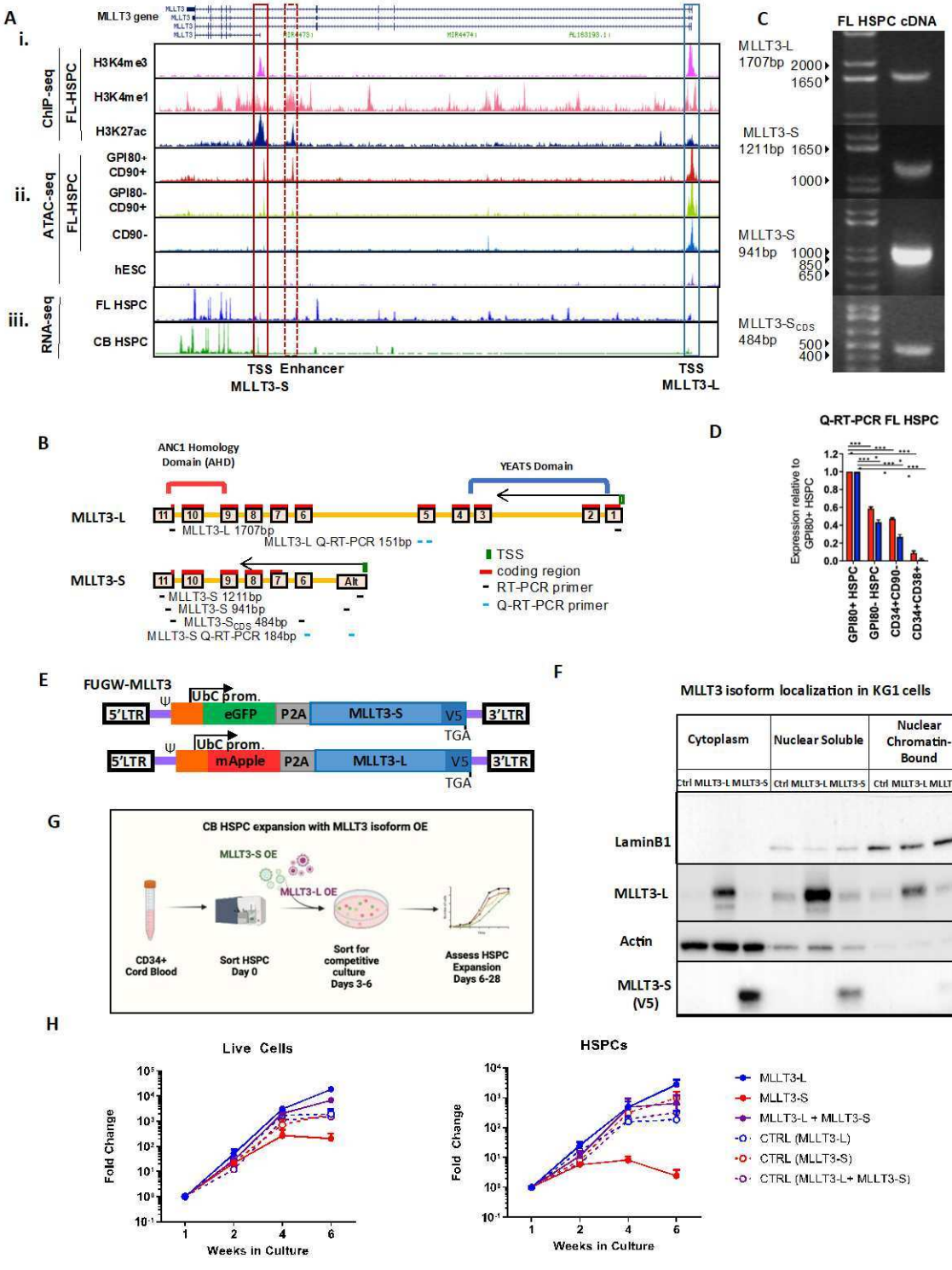
A
MLLT3 gene
i.
ChIP-seq
FL-HSPC
H3K4me3
H3K4me1
H3K27ac
ii.
ATAC-seq
FL-HSPC
GPI80+ CD90+
GPI80- CD90+
CD90-
hESC
iii.
RNA-seq
FL HSPC
CB HSPC
TSS Enhancer
MLLT3-S
TSS
MLLT3-L
C
FL HSPC cDNA
MLLT3-L
1707bp
2000
1650
MLLT3-S
1211bp
1650
1000
MLLT3-S
941bp
1000
850
650
MLLT3-S$_{CDS}$
484bp
500
400
B
ANC1 Homology Domain (AHD)
YEATS Domain
MLLT3-L
MLLT3-L 1707bp
MLLT3-L Q-RT-PCR 151bp
MLLT3-S
MLLT3-S 1211bp
MLLT3-S 941bp
MLLT3-S$_{CDS}$ 484bp
MLLT3-S Q-RT-PCR 184bp
TSS
coding region
RT-PCR primer
Q-RT-PCR primer
D
Q-RT-PCR FL HSPC
Expression relative to GPI80+ HSPC
GPI80+ HSPC
GPI80- HSPC
CD34+CD90-
CD34+CD38+
E
FUGW-MLLT3
5'LTR
Ψ
UbC prom.
eGFP
P2A
MLLT3-S
V5
TGA
3'LTR
mApple
MLLT3-L
F
MLLT3 isoform localization in KG1 cells
Cytoplasm
Nuclear Soluble
Nuclear Chromatin-Bound
Ctrl MLLT3-L MLLT3-S
LaminB1
MLLT3-L
Actin
MLLT3-S (V5)
G
CB HSPC expansion with MLLT3 isoform OE
MLLT3-S OE
MLLT3-L OE
CD34+ Cord Blood
Sort HSPC Day 0
Sort for competitive culture Days 3-6
Assess HSPC Expansion Days 6-28
H
Live Cells
HSPCs
Fold Change
Weeks in Culture
MLLT3-L
MLLT3-S
MLLT3-L + MLLT3-S
CTRL (MLLT3-L)
CTRL (MLLT3-S)
CTRL (MLLT3-L+ MLLT3-S)

Figure 2.1. Human HSCs express a truncated MLLT3 isoform that is downregulated during HSC differentiation.

A. ChIP-seq for histone modifications and ATAC-seq (iii) in human fetal liver HSPCs document an alternative TSS (red box) and candidate enhancer (red dashed box) corresponding to a short isoform of MLLT3. iii. RNA-seq of human fetal liver (FL), cord blood (CB) HSPCs identify a peak corresponding to the alternative 5' exon of MLLT3-S (red box) demonstrating MLLT3-S expression in human HSPCs. **B.** Diagram of MLLT3-L and MLLT3-S exons with corresponding protein domains and location of RT-PCR and Q-RT-PCR primers. **C.** RT-PCR in FL-HSPCs with isoform specific and primers shows bands of predicted length for MLLT3-S transcript **D.** Q-RT-PCR quantification of MLLT3-L and MLLT3-S shows that MLLT3-S expression declines during FL-HSPC differentiation. **E**. Diagram of FUGW lentiviral vector used for MLLT3 isoform overexpression in HSCs. **F.** Subcellular fractionation of KG1 cells overexpressing MLLT3-L and MLLT3-S demonstrating MLLT3-L enrichment in the nucleus and ability to bind chromatin. Overexpression of MLLT3-S generates a stable protein that is not associated with chromatin (n=3). **G.** Strategy for MLLT3 isoform lentiviral overexpression in CB HSPCs **H.** Expansion of fetal liver HSPCs upon overexpression of MLLT3-L, MLLT3-S, and MLLT3-L + MLLT3-S over 6 weeks of competitive culture, documenting expansion of CD34+CD38^{lo}CD90+ HSPC upon MLLT3-L overexpression, but not MLLT3-S overexpression. Data from n=2 independent experiments, n=3 samples total

Overexpression of MLLT3-L, but not MLLT3-S, promotes human HSPC expansion in culture

To compare the function of MLLT3-S to that of MLLT3-L in human HSPCs, we created CD34+CD38−/loCD90+GPI80+ FL HSPCs that overexpressed MLLT3-S coding sequence (MLLT3-SCDS), alone or together with MLLT3-L, in culture (**Figure 2.1E-H**). Although HSPC culture results in gradual downregulation of both isoforms, FUGW lentiviral vectors can maintain MLLT3 transcripts at physiological levels (Calvanese et al., 2019). Overexpression of MLLT3-L promoted the expansion of undifferentiated HSPCs over time, as previously shown (Calvanese et al., 2019), whereas overexpression of MLLT3-S did not expand the undifferentiated HSPC population (**Figure 2.1G**). The cells that overexpressed MLLT3-S together with MLLT3-L also sustained HSPC maintenance, but to a lesser extent than the cells expressing MLLT3-L alone. These data suggest that the long and short MLLT3 isoforms have distinct effects on human HSPCs, and confirm the previous results that MLLT3-L is the key isoform promoting HSPC ex vivo expansion.

Loss of MLLT3-L and MLLT3-S exerts opposing functions in cultured human HSPCs

To understand the unique functions of MLLT3 isoforms, we used lentiviral vectors to knock down long and short MLLT3 isoforms alone, or together **(Figure 2.2A**). We designed and validated two MLLT3-S specific short hairpins (sh3, sh4) that target the 1st alternative exon in MLLT3-S 5'UTR (thereby targeting both MLLT3-S1 and MLLT3-S2), and used commercially available and previously validated sh92 to target MLLT3-L (targets exon 2), and sh93 to knockdown both MLLT3-L and MLLT3-S (targets exon 10 in the ANC homology domain)

(**Figure S2.2B**). The ability of lentiviral shRNA vectors to knock down each isoform at the protein level was first confirmed in K562 cells transduced with pLVX lentiviral vectors that co-express V5-tagged MLLT3-S (MLLT3-SCDS, MLLT3-S1 or MLLT3-S2) or FLAG-tagged MLLT3-L, and fluorescent proteins mCherry and GFP, respectively, via IRES (internal ribosomal entry site). FACS analysis confirmed suppression of the fluorescent protein expression upon knockdown of the overexpressed MLLT3-L or the different subvariants of MLLT3-S transcript (MLLT3-SCDS, MLLT3-S1 and MLLT3-S2) (**Figure S2.2C-E**). Specifically, FACS analysis confirmed that shRNAs against MLLT3-S (sh3 and 4) and both isoforms (sh93) knocked down both 5'UTR subvariants of MLLT3-S, whereas MLLT3-SCDS could only be knocked down by sh93. Both sh92and sh93 resulted in knockdown MLLT3-L. Downregulation of the protein and transcript from two different MLLT3-S subvariants by sh3, sh4 and sh93 was also confirmed by Western blot and Q-RT-PCR in K562 cells engineered to overexpress these variants **(Figure S2, 2F,G).**

Functional effects of loss of MLLT3 isoforms in sorted CB CD34+CD38−/loCD90+ HSPC were evaluated ex vivo by assessing colony formation and HSPC culture expansion over time (**Figure 2.2A**). Methylcellulose assays for transduced, puromycin selected HSPCs plated at day 6 showed that loss of MLLT3-L (either alone or together with MLLT3-S) results significant loss of total colonies, with distinct loss of mixed erythroid/myeloid and erythroid colonies (**Figure 2.2B**). Loss of MLLT3-S preserved all colony types but resulted in slight decline of total colonies that was not statistically significant.

HSPC expansion culture revealed that knockdown of MLLT3-L (sh92) in sorted HSPCs leads to precocious emergence of CD34+CD38hi cells by day 6 in culture, followed by a gradual decline of undifferentiated HSPCs and total CD34+ cells (**Figure 2.2C-G**). In contrast, by day 6, knockdown of MLLT3-S resulted in increased frequency of the most undifferentiated

CD34+CD38−/loCD90+EPCR+ subset of CB HSPC compared to the control vector. Simultaneous knockdown of both MLLT3 isoforms (sh93) created a complex phenotype – an immediate relative expansion of CD34+CD38−/loCD90+EPCR+ cells by day 6, similar to MLLT3-S KD, followed by a gradual loss of undifferentiated CD34+CD90+ cells, similar to MLLT3-L KD, but without an increase in CD34+CD38+ cells (**Figure 2.2C-G**). These distinct responses to the knockdown of MLLT3-S and MLLT3-L further confirm that the two isoforms have different roles in regulating human HSC behavior, further shows that MLLT3-L is the key isoform required for the expansion of undifferentiated, multipotent HSPC.

Figure 2.2

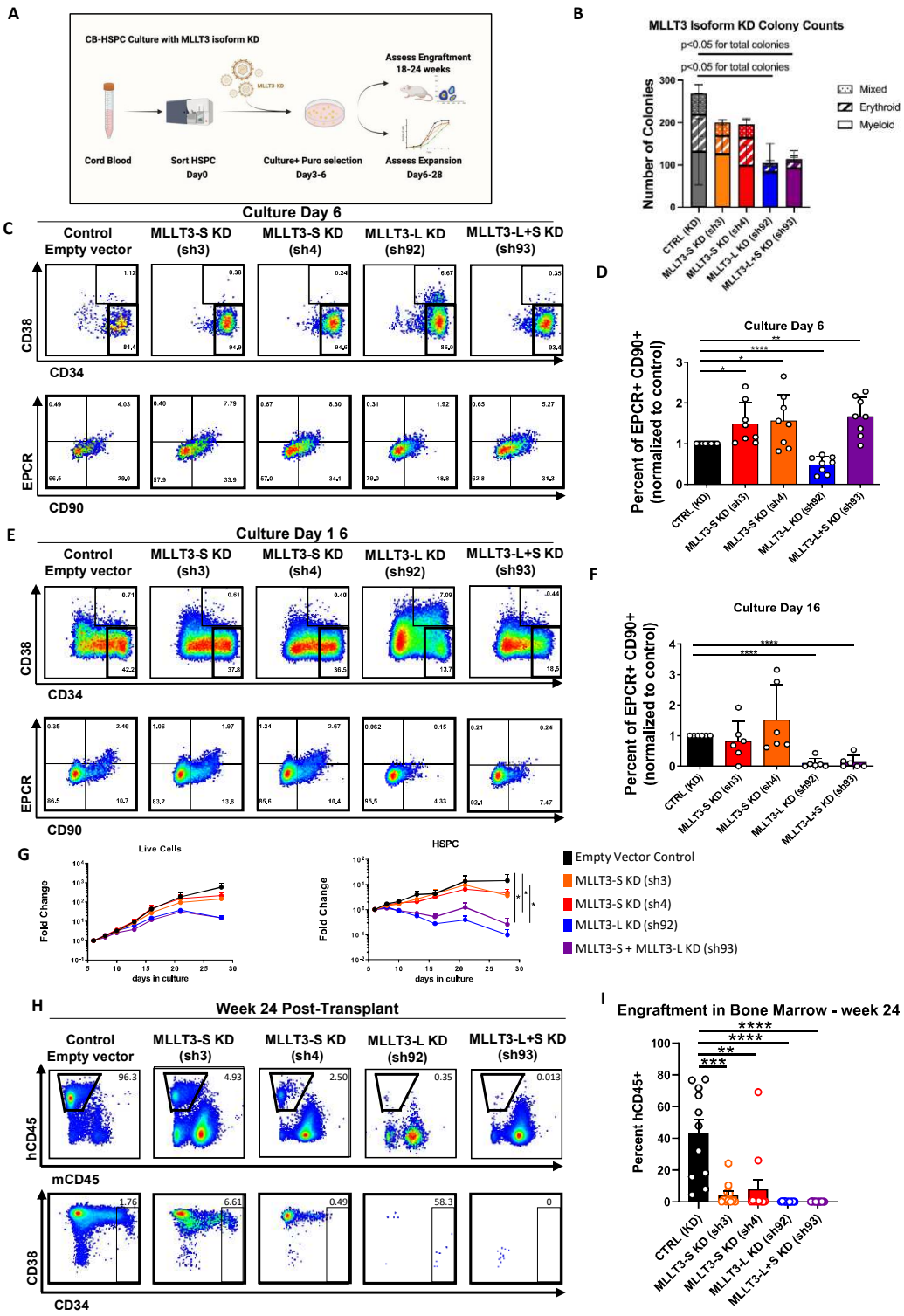
A
CB-HSPC Culture with MLLT3 isoform KD
Assess Engraftment 18-24 weeks
Cord Blood
Sort HSPC Day0
Culture+ Puro selection Day3-6
Assess Expansion Day6-28
B
MLLT3 Isoform KD Colony Counts
p<0.05 for total colonies
p<0.05 for total colonies
Number of Colonies
Mixed
Erythroid
Myeloid
CTRL (KD)
MLLT3-S KD (sh3)
MLLT3-S KD (sh4)
MLLT3-L KD (sh92)
MLLT3-L+S KD (sh93)
C
Culture Day 6
Control Empty vector
MLLT3-S KD (sh3)
MLLT3-S KD (sh4)
MLLT3-L KD (sh92)
MLLT3-L+S KD (sh93)
CD38
CD34
EPCR
CD90
D
Culture Day 6
Percent of EPCR+ CD90+ (normalized to control)
E
Culture Day 1 6
F
Culture Day 16
G
Live Cells
HSPC
Fold Change
days in culture
Empty Vector Control
MLLT3-S KD (sh3)
MLLT3-S KD (sh4)
MLLT3-L KD (sh92)
MLLT3-S + MLLT3-L KD (sh93)
H
Week 24 Post-Transplant
hCD45
mCD45
I
Engraftment in Bone Marrow - week 24
Percent hCD45+

Figure 2.2. Both long and short isoforms of MLLT3 are required for human HSC function.

A. Strategy for MLLT3 isoform lentiviral shRNA knockdown in CB HSPCs **B**. Quantification of CFU-C in methylcellulose colony assay (mean and sd, data from n=2 independent replicates, n=4 CB samples total). **C-F.** Flow cytometry analysis (**C,E**) and quantification **(D,F**) of human HSPC marker expression in the progeny of $CD34^{+}CD38^{lo}CD90^{+}$ CB HSPCs transduced with empty vector control or knockdown vectors targeting MLLT3 isoforms on day 6 (**C, D**) or day 16 (**E, F**) of culture. Data from n=3 independent experimental replicates, n=4 CB samples total. **G**. Quantification of CB HSPC expansion in vitro following MLLT3 isoform KD (shown as mean and SEM). **H.** Flow cytometry analysis of bone marrow from NBSGW mice transplanted with the progeny of $CD34^{+}CD38^{lo}CD90^{+}$ CB HSPCs transduced with MLLT3 KD vectors or empty vector control. Representative FACS plots showing total human engraftment (hCD45) and HSPCs ($hCD45^{+}CD34^{+}CD38^{lo}$) 24 weeks after transplantation. **I.** Quantification of engraftment (hCD45) in mouse bone marrow 24 weeks after transplantation (n=3 independent experimental replicates, n=11-12 mice per condition).

Both MLLT3 isoforms are critical for human CB HSPC engraftment and repopulation in NBSGW mice

We next evaluated the effects of knockdown of the long and short MLLT3 isoforms in vivo by transplantation to immunodeficient NBSGW mice (**Figure 2.2A**). CB HSPCs transduced with lentiviral vectors against the MLLT3 isoforms were transplanted to female NBSGW mice at day 6 in culture, when all KD samples still contain undifferentiated HSPCs. The mice were analyzed

by bone marrow biopsy at 18 weeks after transplantation and by harvesting hematopoietic tissues at 24 weeks (**Figure 2.2H,I**). The empty vector control transduced HSPCs showed consistent engraftment of human CD45 cells, many of which contained CD34+CD38+/- HSPCs and both myeloid and lymphoid cells, while mice transplanted with HSPCs deficient for MLLT3-L or both isoforms showed no engraftment. Interestingly, mice transplanted with HSPCs deficient for MLLT3-S (sh3 or sh4) also showed minimal repopulation of hCD45 cells (**Figure 2.2H,I**).These data confirmed that loss of full length MLLT3 results in complete failure of human hematopoietic engraftment, and also revealed that although knockdown of MLLT3-S preserves immunophenotypic human HSPCs in culture, these cells are unable to robustly reconstitute the hematopoietic system upon transplantation. These data imply that expression of both long and short isoforms of MLLT3 in human HSPCs is essential for proper HSC function, but through unique functions.

scRNA sequencing reveals that MLLT3-L and MLLT3-S regulate HSC stemness programs in opposing directions

To understand at the molecular level how MLLT3 isoforms control HSPC function, we performed scRNA seq for the progeny of CD34+CD38loCD90 HSPCs from two cord bloods (male and female) that were transduced with lentiviral knockdown vectors and sorted for CD34+ cells at day 6 in culture (**Figure 2.3A**). Evaluation of HSPCs based on HSC marker genes HLF+ (HSC transcription factor) or EPCR+ (marker of transplantable HSCs in culture), PROM1/CD133 (a more broadly expressed HSPC marker in mature HSCs), and MYCT1 (novel HSC regulator that declines in cultured HSCs) implied that all modified cells still possessed undifferentiated HSPCs, although these populations were reduced in MLLT3-L KD (sh92) and MLLT3-L+MLLT3-S

(sh93) KD knockdown cells (**Figure 2.3B, C, Figure S2.3A**). Distinct from all other HLF+ HSPC populations that overlapped in UMAP, MLLT3-L KD HLF+ cells formed a distinct population. Analysis of progenitor marker CD38+ expression in total CD34+ cells showed a drastic increase in CD38+ cells in MLLT3-L KD cells, consistent with FACS analysis **(Figure S2.2A).**

To determine the molecular programs triggering the unique behaviors of MLLT3 isoform KD HSPCs, we compared genes differentially expressed in HLF HSCs between control and MLLT3-S KD (sh3, sh4), MLLT3-L KD (sh92) or knockdown of both isoforms (sh93). Pathway analysis of the differentially expressed genes comparing MLLT3-L or MLLT3-S knockdown to control cells using Enrichr revealed that many programs were dysregulated in opposing directions by knockdown of the two isoforms (**Figure S2.3B-D**). Programs such as transcription (containing classical MLLT3 targets such as HSPC transcription factors ERG[30,42,43], GATA2[30,44-47]) and protein synthesis machinery that were upregulated by MLLT3-OE in previous studies[30] were downregulated in MLLT3-L KD, but often upregulated in MLLT3-S KD HSPCs (**Figure 2.3D-E, Figure S2.3B-D**). Conversely, programs that became upregulated upon loss of MLLT3-L such as aerobic electron transport, mitochondrial translation, and cholesterol biosynthesis were downregulated in MLLT3-S KD HSPCs (**Figure 2.3D, F, Figure S2.3B-D**), suggesting that the two isoforms of MLLT3 can regulate the metabolic state of human HSPCs. Splicing was another novel category upregulated upon loss of MLLT3-L, and downregulated upon loss of MLLT3-S.

To assure that the results were not skewed by method of selecting HSPCs based on expression of one gene, HLF, we also selected HSPCs using PROM1 expression. The dysregulated programs identified were very similar whether HSPC were defined by the expression of HLF or PROM1 (Data not shown).

To assess whether the programs differentially regulated by MLLT3-L and MLLT3-S in culture are direct target genes of MLLT3-L, the isoform that can bind chromatin, ChIP-seq was performed on cultured CB CD34+ cells with and without MLLT3-L overexpression.MLLT3-L binding was detected in all categories of genes, including the previously identified transcription factor and protein synthesis genes that are activated by MLLT3-L, as well as the newly identified MLLT3-L repressed categories such as splicing and aerobic electron transport (**Figure 2.3G**). Together, these data show that MLLT3 long and short isoforms regulate many key HSC stemness programs in opposing directions, suggesting MLLT3 isoforms may be used to control HSC behavior and metabolic states.

Figure 2.3

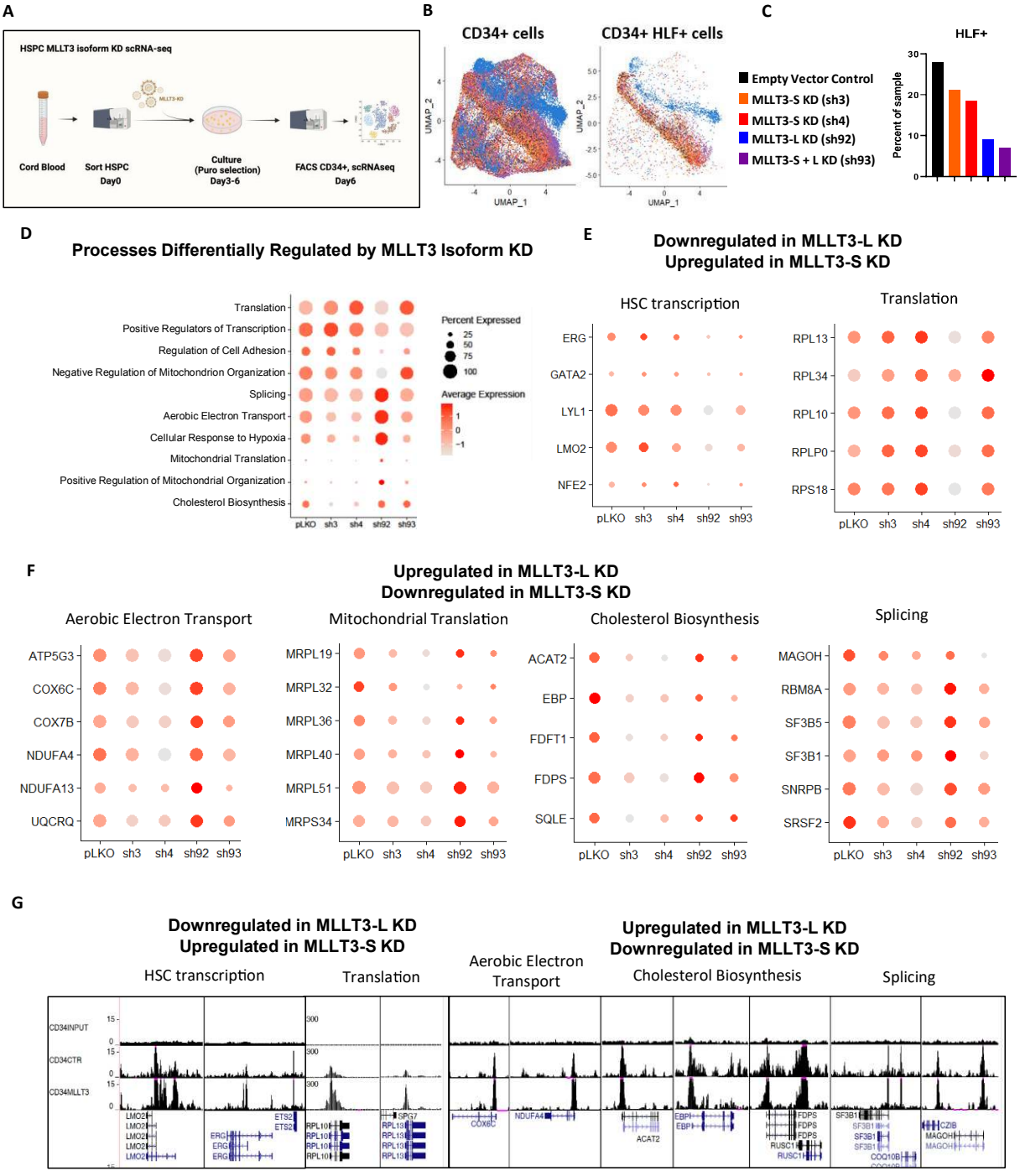
A
HSPC MLLT3 isoform KD scRNA-seq
Cord Blood
Sort HSPC Day0
Culture (Puro selection) Day3-6
FACS CD34+, scRNAseq Day6
B
CD34+ cells
CD34+ HLF+ cells
UMAP_1
UMAP_2
C
Empty Vector Control
MLLT3-S KD (sh3)
MLLT3-S KD (sh4)
MLLT3-L KD (sh92)
MLLT3-S + L KD (sh93)
HLF+
Percent of sample
D
Processes Differentially Regulated by MLLT3 Isoform KD
Translation
Positive Regulators of Transcription
Regulation of Cell Adhesion
Negative Regulation of Mitochondrion Organization
Splicing
Aerobic Electron Transport
Cellular Response to Hypoxia
Mitochondrial Translation
Positive Regulation of Mitochondrial Organization
Cholesterol Biosynthesis
Percent Expressed
Average Expression
pLKO sh3 sh4 sh92 sh93
E
Downregulated in MLLT3-L KD
Upregulated in MLLT3-S KD
HSC transcription
ERG
GATA2
LYL1
LMO2
NFE2
Translation
RPL13
RPL34
RPL10
RPLP0
RPS18
F
Upregulated in MLLT3-L KD
Downregulated in MLLT3-S KD
Aerobic Electron Transport
ATP5G3
COX6C
COX7B
NDUFA4
NDUFA13
UQCRQ
Mitochondrial Translation
MRPL19
MRPL32
MRPL36
MRPL40
MRPL51
MRPS34
Cholesterol Biosynthesis
ACAT2
EBP
FDFT1
FDPS
SQLE
Splicing
MAGOH
RBM8A
SF3B5
SF3B1
SNRPB
SRSF2
G
Downregulated in MLLT3-L KD
Upregulated in MLLT3-S KD
HSC transcription
Translation
Upregulated in MLLT3-L KD
Downregulated in MLLT3-S KD
Aerobic Electron Transport
Cholesterol Biosynthesis
Splicing
CD34INPUT
CD34CTR
CD34MLLT3

Figure 2.3: Loss of MLLT3 long and short isoforms has opposing effects on HSC programs

A. Strategy for scRNAseq following MLLT3 isoform-specific lentiviral shRNA knockdown of MLLT3 in CB CD34+CD38loCD90+ HSPCs **B.** UMAP clustering of CD34+ and CD34+ HLF+ cells following MLLT3 isoform-specific KD. **C.** Quantification of HLF+ cells in each condition. **D.** Module scores showing selected gene ontology biological processes differentially expressed in MLLT3-S and MLLT3-L knockdown HLF+ HSCs **E.** Selected genes in categories that are upregulated by MLLT3-S KD and downregulated by MLLT3-L knockdown HLF+ HSCs. **F.** Selected genes in categories that are downregulated by MLLT3-S KD and upregulated by MLLT3-L knockdown HLF+ HSCs. **G.** MLLT3-L binding in selected genes from (**E, F**) with and without MLLT3 overexpression in cultured human CB HSPC

Conclusion

MLLT3 is a critical human HSCs self-renewal regulator that couples the presence of activating histone marks with the activity of protein complexes that regulate transcription elongation, such as the SEC and DOT1L complex[30,37,39]. Here, we demonstrate that human HSC "stemness" is governed not only by the full-length MLLT3 but also by its truncated isoform, MLLT3-S, that can counteract the function of MLLT3-L. Rescuing the declining expression of MLLT3 transcripts in cultured human HSCs by sustained expression of MLLT3-L drives the expansion of cultured CB HSCs, whereas MLLT3-S has a suppressive effect. MLLT3-L and MLLT3-S work together to offer two-dimensional regulation to key HSC programs such as expression of HSC stemness transcription factors[10,42-48], protein synthesis machinery[49,50], mitochondrial metabolism[51-56], and splicing[57-59], all of which are critical regulators of HSC fate,

and their dysregulation can lead to dramatic consequences. The long and short isoforms of MLLT3 provide an additional level of control over MLLT3-mediated gene expression, allowing fine-tuning of stemness programs so that HSCs can respond to different developmental and physiological demands. This need for a dynamic response makes the expression of both MLLT3 isoforms critical for HSC function; although loss of MLLT3-S appears to have only a moderate effect on HSC behavior in the ex vivo culture environment, it profoundly disrupts their repopulation ability in the more dynamic context of in vivo transplantation.

Figure S2.1

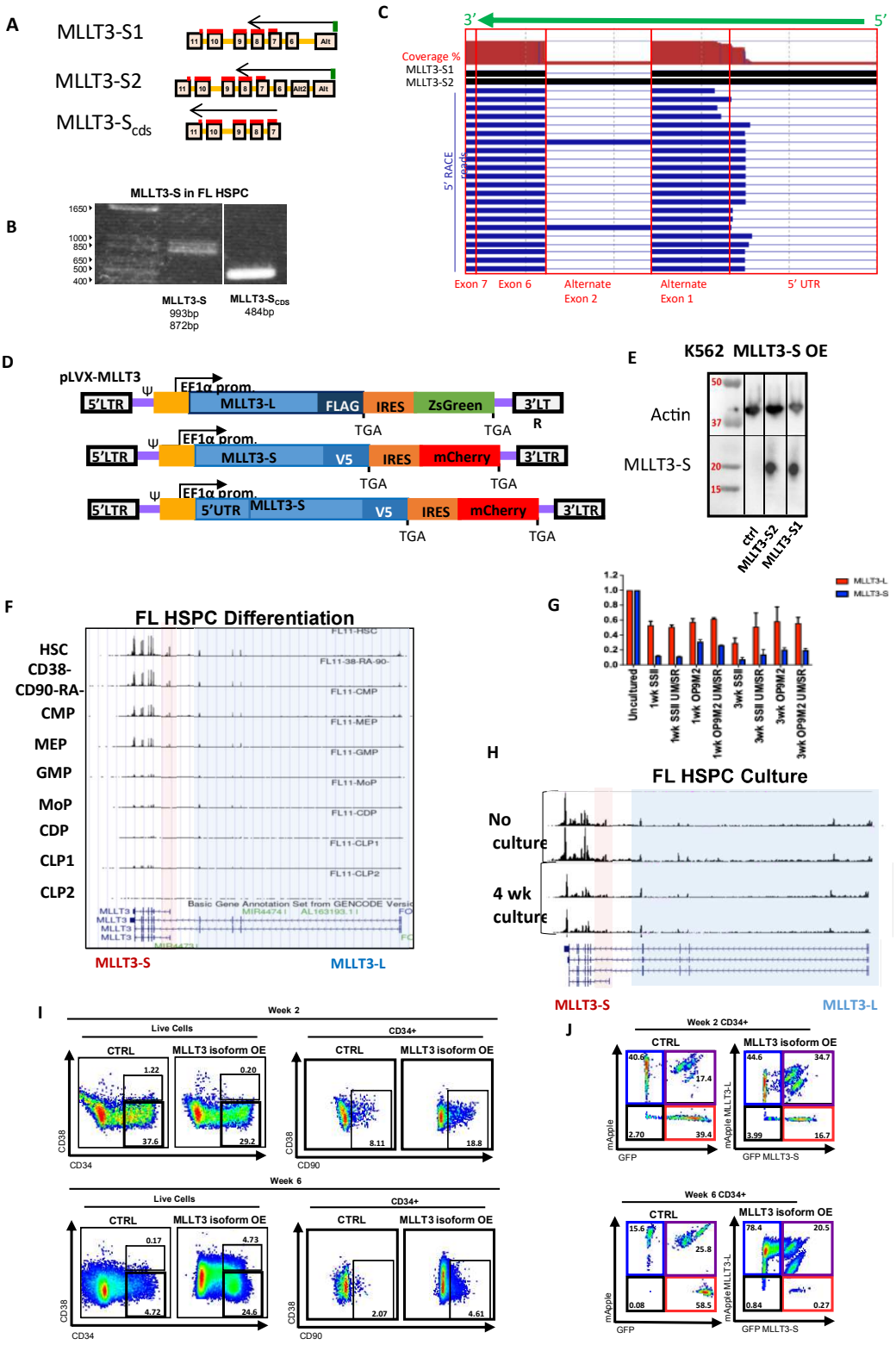
A
MLLT3-S1
MLLT3-S2
MLLT3-S$_{cds}$
B
MLLT3-S in FL HSPC
1650
1000
850
650
500
400
MLLT3-S
993bp
872bp
MLLT3-S$_{CDS}$
484bp
C
3′
5′
Coverage %
MLLT3-S1
MLLT3-S2
5′ RACE reads
Exon 7
Exon 6
Alternate Exon 2
Alternate Exon 1
5′ UTR
D
pLVX-MLLT3
5′LTR
Ψ
EF1α prom.
MLLT3-L
FLAG
IRES
ZsGreen
3′LTR
TGA
MLLT3-S
V5
mCherry
5′UTR
E
K562 MLLT3-S OE
Actin
MLLT3-S
50
37
20
15
ctrl
MLLT3-S2
MLLT3-S1
F
FL HSPC Differentiation
HSC
CD38-
CD90-RA-
CMP
MEP
GMP
MoP
CDP
CLP1
CLP2
MLLT3-S
MLLT3-L
G
MLLT3-L
MLLT3-S
Uncultured
1wk SSII
1wk SSII UM/SR
1wk OP9M2
1wk OP9M2 UM/SR
3wk SSII
3wk SSII UM/SR
3wk OP9M2
3wk OP9M2 UM/SR
H
FL HSPC Culture
No culture
4 wk culture
MLLT3-S
MLLT3-L
I
Week 2
Live Cells
CD34+
CTRL
MLLT3 isoform OE
CD38
CD34
CD90
Week 6
J
Week 2 CD34+
Week 6 CD34+
mApple
GFP
mApple MLLT3-L
GFP MLLT3-S

Figure S2.1. **Validation of MLLT3-S transcript and protein in human hematopoietic cells**

Related to Figure 2.1

A. Diagram of MLLT3-S transcript variants showing one or two alternative exons in the 5' UTR, as well as their shared coding sequence. **B.** RT-PCR from FL $CD34^{+}CD38^{lo}CD90^{+}GPI80^{+}$ HSPC demonstrates bands of expected size based on predicted sequences of MLLT3-S 5'UTR variants **C.** Alignment of 5' RACE reads from $CD34^{+}CD38^{lo}CD90^{+}$ CB HSPCs demonstrates presence of predicted exons at 5' UTR **D.** Diagram of pLVX MLLT3 lentiviral overexpression vectors, including a version allowing expression from native 5'UTR, used for KG1 and K562 cell line overexpression. **E**. Stable overexpression of MLLT3-S1 or MLLT3-S2 transcript variants in K562 cell line produces stable protein of expected size. **F, G.** Bulk RNAseq demonstrating expression of unique MLLT3-S exons in FL hematopoietic cells during differentiation (F) and culture (G). **H**. Q-RT-PCR quantification of MLLT3 isoform expression in different culture conditions. **I.** Representative flow cytometry analysis showing CD34+ CD38lo population and CD34+ CD38lo CD90+ populations in competitive culture of human FL HSPCs following lentiviral overexpression of MLLT3 isoforms. **J.** Representative flow cytometry analysis of fluorescent marker expression in same assay as I. Blue shows MLLT3-L OE or empty vector control, red shows MLLT3-S OE or empty vector control, purple shows MLLT3-L and MLLT3-S or empty vector control. (n=2 independent replicates, n=3 samples total)

Figure S2.2

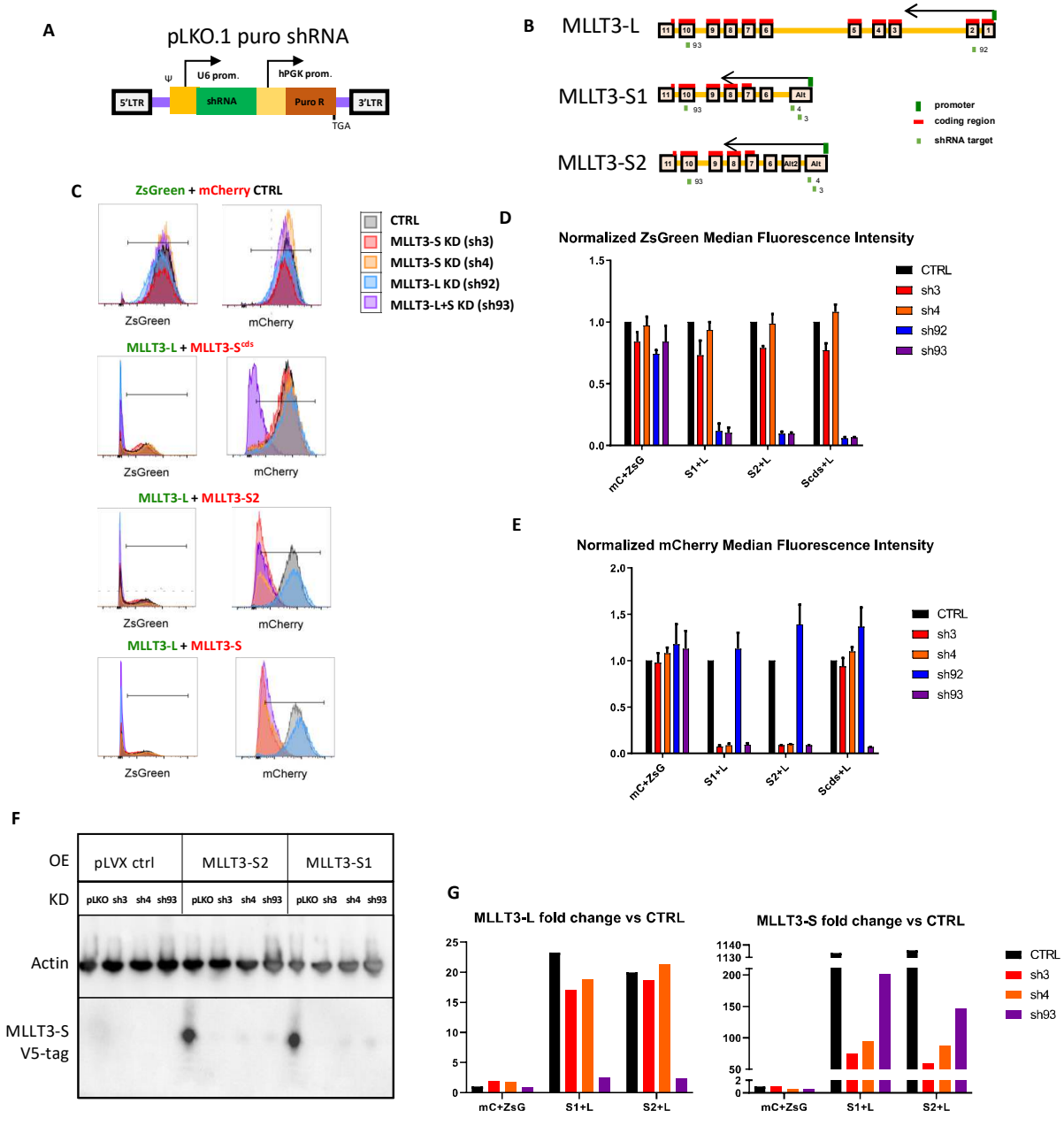
A
pLKO.1 puro shRNA
Ψ
U6 prom.
hPGK prom.
5'LTR
shRNA
Puro R
3'LTR
TGA
B
MLLT3-L
MLLT3-S1
MLLT3-S2
promoter
coding region
shRNA target
C
ZsGreen + mCherry CTRL
ZsGreen
mCherry
MLLT3-L + MLLT3-S^cds
MLLT3-L + MLLT3-S2
MLLT3-L + MLLT3-S
CTRL
MLLT3-S KD (sh3)
MLLT3-S KD (sh4)
MLLT3-L KD (sh92)
MLLT3-L+S KD (sh93)
D
Normalized ZsGreen Median Fluorescence Intensity
CTRL
sh3
sh4
sh92
sh93
mC+ZsG
S1+L
S2+L
Scds+L
E
Normalized mCherry Median Fluorescence Intensity
F
OE
KD
pLVX ctrl
MLLT3-S2
MLLT3-S1
pLKO sh3 sh4 sh93
Actin
MLLT3-S
V5-tag
G
MLLT3-L fold change vs CTRL
MLLT3-S fold change vs CTRL

Figure S2.2. Strategy of knockdown of MLLT3 isoforms. Related to Figure 2.2.

A. Diagram of pLKO vector for MLLT3 isoform-specific lentiviral knockdown in cell lines and primary human HSPCs. **B.** Diagram showing locations of shRNA targets for isoform-specific MLLT3 shRNA knockdown. **C.** Flow cytometry analysis showing fluorescent selection marker expression in K562 cell lines stably overexpressing MLLT3-L and MLLT3-S1, MLLT3-S2, or MLLT3-S coding sequence (MLLT3-Scds) following MLLT3-S isoform-specific shRNA knockdown. Reduction of selection marker fluorescence indicates decreased expression of corresponding overexpressed protein. Representative plots shown n=3. **D, E.** Quantification of median fluorescence intensity (MFI) of MLLT3-L selection marker fluorescence (ZsGreen) D. and MLLT3-S transcript variant selection marker fluorescence (mCherry) E., following isoform-specific MLLT3 knockdown (n=3). **F.** Western Blot showing expression of MLLT3-S (detected with anti-V5 antibody) in K562 cell lines with stable overexpression of V5-tagged MLLT3-S1 or MLLT3-S2 variants (also shown in Figure S1) and their loss following lentiviral shRNA knockdown of MLLT3-S (n=1) **G.** Quantification of Q-RT-PCR showing MLLT3 isoform expression in K562 cell lines stably overexpressing MLLT3-L and MLLT3-S1 or MLLT3-S2 following MLLT3-S isoform-specific shRNA knockdown (n=1)

Figure S2.3

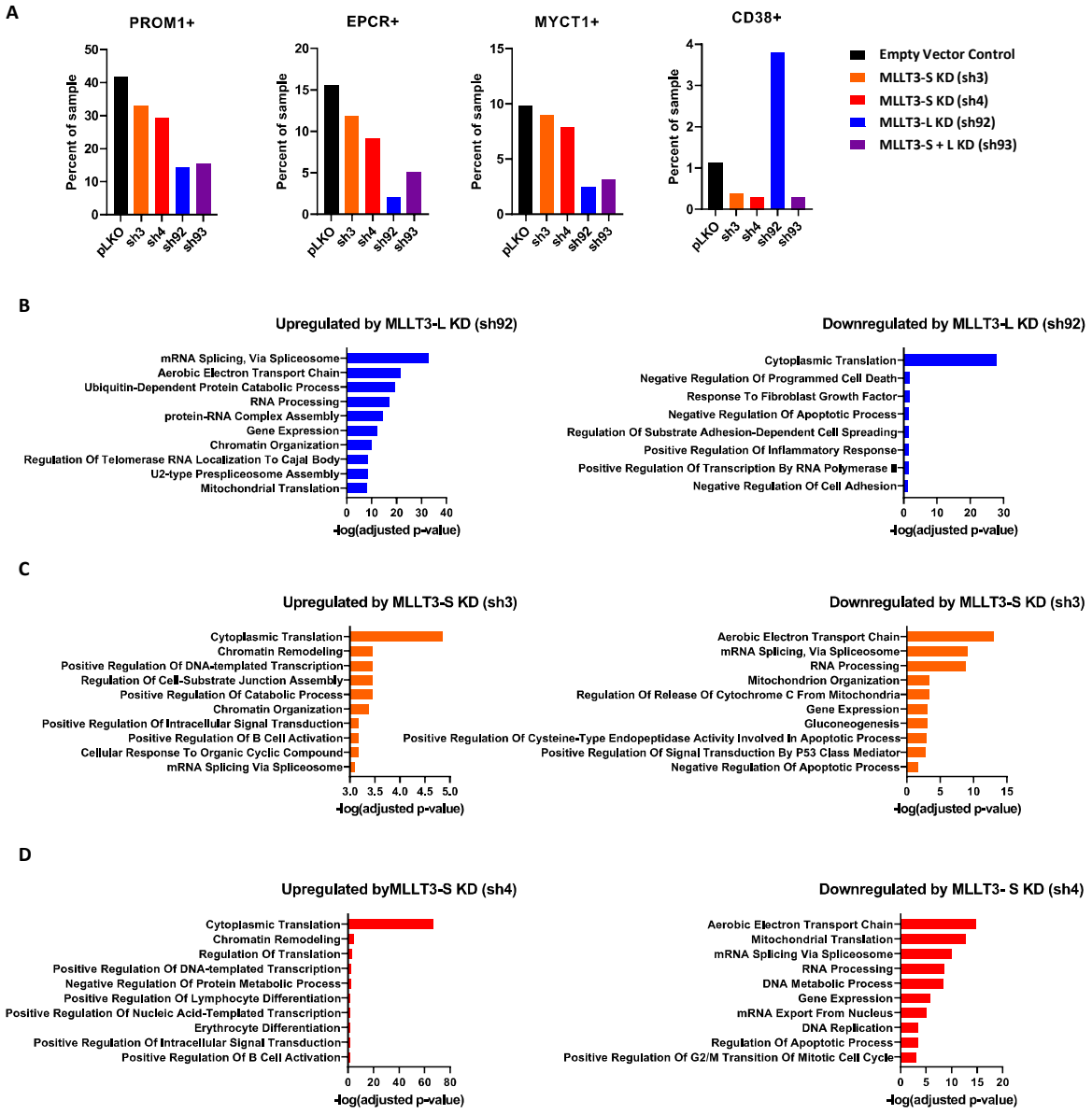

Figure S2.3. MLLT3 isoform knockdown effects on HSPC numbers and differentially regulated programs Related to Figure 2.3

A. Quantification of key human HSC markers in scRNAseq following MLLT3 isoform-specific lentiviral shRNA knockdown of MLLT3 in CB CD34+ CD38lo CD90+ HSPCs **B-D.** Top GO categories that are either up- or downregulated following knockdown of MLLT3-L (B), or MLLT3-S (C, D).

Bibliography

1. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., and Crooks, G.M. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281-286.
2. Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S.A., and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure *21*, 176-183.
3. Wu, H., Chen, L., Zhou, Q., and Zhang, W. (2011). AF17 facilitates Dot1a nuclear export and upregulates ENaC-mediated Na+ transport in renal collecting duct cells. PLoS One *6*, e27429.
4. Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y., and Li, W. (2014). AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. Cell *159*, 558-571.
5. Li, Y., Sabari, B.R., Panchenko, T., Wen, H., Zhao, D., Guan, H., Wan, L., Huang, H., Tang, Z., and Zhao, Y. (2016). Molecular coupling of histone crotonylation and active transcription by AF9 YEATS domain. Molecular cell *62*, 181-193.
6. Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes & development *24*, 574-589.
7. Kuntimaddi, A., Achille, N.J., Thorpe, J., Lokken, A.A., Singh, R., Hemenway, C.S., Adli, M., Zeleznik-Le, N.J., and Bushweller, J.H. (2015). Degree of recruitment of DOT1L to

MLL-AF9 defines level of H3K79 Di-and tri-methylation on target genes and transformation potential. Cell reports *11*, 808-820.

8. Barry, E.R., Krueger, W., Jakuba, C.M., Veilleux, E., Ambrosi, D.J., Nelson, C.E., and Rasmussen, T.P. (2009). ES cell cycle progression and differentiation require the action of the histone methyltransferase Dot1L. Stem Cells *27*, 1538-1547.

9. Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nature reviews Molecular cell biology *13*, 543-547.

10. Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output. Molecular and Cellular Biology *32*, 2608-2617. 10.1128/mcb.00182-12.

11. He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., Alber, T., Benkirane, M., and Zhou, Q. (2011). Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. Proceedings of the National Academy of Sciences *108*, E636-E645.

12. Bordignon, C. (2006). Stem-cell therapies for blood diseases. Nature *441*, 1100-1102. 10.1038/nature04962.

13. Haltalli, M.L., Wilkinson, A.C., Rodriguez-Fraticelli, A., and Porteus, M. (2022). Hematopoietic stem cell gene editing and expansion: State-of-the-art technologies and recent applications. Experimental Hematology *107*, 9-13.

14. Ho, A.D., and Wagner, W. (2007). The beauty of asymmetry: asymmetric divisions and self-renewal in the haematopoietic system. Current opinion in hematology *14*, 330-336.

15. Loeffler, D., and Schroeder, T. (2021). Symmetric and asymmetric activation of hematopoietic stem cells. Current Opinion in Hematology *28*, 262-268.

16. Magnusson, M., Sierra, M.I., Sasidharan, R., Prashad, S.L., Romero, M., Saarikoski, P., Van Handel, B., Huang, A., Li, X., and Mikkola, H.K.A. (2013). Expansion on Stromal Cells Preserves the Undifferentiated State of Human Hematopoietic Stem Cells Despite Compromised Reconstitution Ability. PLoS ONE *8*, e53912. 10.1371/journal.pone.0053912.

17. Fares, I., Calvanese, V., and Mikkola, H.K. (2022). Decoding Human Hematopoietic Stem Cell Self-Renewal. Current Stem Cell Reports *8*, 93-106.

18. Sakurai, M., Ishitsuka, K., Ito, R., Wilkinson, A.C., Kimura, T., Mizutani, E., Nishikii, H., Sudo, K., Becker, H.J., and Takemoto, H. (2023). Chemically defined cytokine-free expansion of human haematopoietic stem cells. Nature *615*, 127-133.

19. Dou, D.R., Calvanese, V., Sierra, M.I., Nguyen, A.T., Minasian, A., Saarikoski, P., Sasidharan, R., Ramirez, C.M., Zack, J.A., Crooks, G.M., et al. (2016). Medial HOXA genes demarcate haematopoietic stem cell fate during human development. Nature Cell Biology *18*, 595-606. 10.1038/ncb3354.

20. Prashad, S.L., Calvanese, V., Yao, C.Y., Kaiser, J., Wang, Y., Sasidharan, R., Crooks, G., Magnusson, M., and Mikkola, H.K.A. (2015). GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. Cell stem cell *16*, 80-87.

21. Loughran, S.J., Kruse, E.A., Hacking, D.F., De Graaf, C.A., Hyland, C.D., Willson, T.A., Henley, K.J., Ellis, S., Voss, A.K., and Metcalf, D. (2008). The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nature immunology *9*, 810-819.

22. Taoudi, S., Bee, T., Hilton, A., Knezevic, K., Scott, J., Willson, T.A., Collin, C., Thomas, T., Voss, A.K., and Kile, B.T. (2011). ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. Genes & development *25*, 251-262.

23. Tipping, A.J., Pina, C., Castor, A., Hong, D., Rodrigues, N.P., Lazzari, L., May, G.E., Jacobsen, S.E.W., and Enver, T. (2009). High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. Blood, The Journal of the American Society of Hematology *113*, 2661-2672.

24. De Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper, S.A., Speck, N., and Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. Journal of Experimental Medicine *210*, 2843-2850.

25. Lim, K.-C., Hosoya, T., Brandt, W., Ku, C.-J., Hosoya-Ohmura, S., Camper, S.A., Yamamoto, M., and Engel, J.D. (2012). Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning. The Journal of clinical investigation *122*, 3705-3717.

26. Vicente, C., Conchillo, A., García-Sánchez, M.A., and Odero, M.D. (2012). The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Critical reviews in oncology/hematology *82*, 1-17.

27. El Omari, K., Hoosdally, S.J., Tuladhar, K., Karia, D., Vyas, P., Patient, R., Porcher, C., and Mancini, E.J. (2011). Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex. Blood, The Journal of the American Society of Hematology *117*, 2146-2156.

28. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631-644. 10.1016/j.cell.2008.01.025.

29. Signer, R.A., Magee, J.A., Salic, A., and Morrison, S.J. (2014). Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature *509*, 49-54.

30. Buszczak, M., Signer, R.A., and Morrison, S.J. (2014). Cellular differences in protein synthesis regulate tissue homeostasis. Cell *159*, 242-251.

31. Rimmelé, P., Liang, R., Bigarella, C.L., Kocabas, F., Xie, J., Serasinghe, M.N., Chipuk, J., Sadek, H., Zhang, C.C., and Ghaffari, S. (2015). Mitochondrial metabolism in hematopoietic stem cells requires functional FOXO 3. EMBO reports *16*, 1164-1176.

32. Bonora, M., Ito, K., Morganti, C., Pinton, P., and Ito, K. (2018). Membrane-potential compensation reveals mitochondrial volume expansion during HSC commitment. Experimental hematology *68*, 30-37. e31.

33. Moon, J., Kim, H.R., and Shin, M.G. (2018). Rejuvenating aged hematopoietic stem cells through improvement of mitochondrial function. Annals of Laboratory Medicine *38*, 395.

34. Papa, L., Djedaini, M., and Hoffman, R. (2019). Mitochondrial role in stemness and differentiation of hematopoietic stem cells. Stem cells international *2019*.

35. Qiu, J., and Ghaffari, S. (2022). Mitochondrial deep dive into hematopoietic stem cell dormancy: not much glycolysis but plenty of sluggish lysosomes. Experimental hematology.

36. Takubo, K., Nagamatsu, G., Kobayashi, C.I., Nakamura-Ishizu, A., Kobayashi, H., Ikeda, E., Goda, N., Rahimi, Y., Johnson, R.S., and Soga, T. (2013). Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. Cell stem cell *12*, 49-61.

37. Goldstein, O., Meyer, K., Greenshpan, Y., Bujanover, N., Feigin, M., Ner-Gaon, H., Shay, T., and Gazit, R. (2017). Mapping whole-transcriptome splicing in mouse hematopoietic stem cells. Stem Cell Reports *8*, 163-176.

38. Tan, D.Q., Li, Y., Yang, C., Li, J., Tan, S.H., Chin, D.W.L., Nakamura-Ishizu, A., Yang, H., and Suda, T. (2019). PRMT5 modulates splicing for genome integrity and preserves proteostasis of hematopoietic stem cells. Cell reports *26*, 2316-2328. e2316.

39. Wang, F., Tan, P., Zhang, P., Ren, Y., Zhou, J., Li, Y., Hou, S., Li, S., Zhang, L., and Ma, Y. (2022). Single-cell architecture and functional requirement of alternative splicing during hematopoietic stem cell formation. Science advances *8*, eabg5369.

Chapter 3

Role of MLLT3 isoforms in human HSC Maturation

Abstract

Understanding the mechanisms that regulate human HSC maturation during development is key for the improvement of protocols for directed differentiation of transplantable PSC-derived HSCs. So far, it has been documented that human PSCs can differentiate to HSPCs that resemble closely nascent HSCs found in human AGM region, but they do not mature to fully functional, robustly transplantable HSCs that are found in fetal liver or bone marrow[9-13,24]. The transcriptional regulator MLLT3 has previously been shown to be expressed early during HSC specification, and increase in expression throughout HSC maturation, suggesting it may play a role in this process[24]. However, this scRNA seq dataset does not by default differentiate between MLLT3 isoforms. Bulk RNA-seq and scRNAseq analysis of human developmental tissues revealed that, while MLLT3-L is expressed already in hemogenic endothelium in the embryo, MLLT3-S is induced in maturing fetal liver HSCs. The induction of MLLT3-S expression is closely correlated with the downregulation of IGFBP2, a regulator associated with HSC expansion and highly proliferative immature HSCs. scRNA and bulk RNA-seq of HSPCs with MLLT3-L and/or MLLT3-S OE and KD revealed that MLLT3-S suppresses IGFBP2 expression *in vitro*, while MLLT3-L increases it. This suggests that by suppressing MLLT3-L-driven expression of IGFBP2 and HSC expansion, MLLT3-S may promote HSC maturation in the fetal liver and their transition to a more homeostatic state. Co-immunoprecipitation (Co-IP) experiments in hematopoietic cell lines demonstrated that MLLT3-S and MLLT3-L are able to bind the same protein complexes. As MLLT3-S is unable to bind chromatin, we propose that MLLT3-S sequesters these protein complexes, preventing them from binding to MLLT3-L and being directed to genes whose expression is regulated by MLLT3-L

Introduction While significant progress has been made in the design of protocols for the reprogramming of hPSCs into HSPCs, the HSPCs produced using these methods remain functionally immature and cannot be used for transplantation in the clinic[9-13,24]. Our poor understanding of the factors involved in HSC maturation during human development is a continued roadblock to recapitulating this process in PSC-derived HSCs. Human development includes multiple waves of hematopoiesis arising from multiple anatomical sites[9,14,20-23,25]. Only the third, final wave generates definitive HSCs[10,24,25]. These cells emerge from hemogenic endothelium in the AGM and travel through the bloodstream and extraembryonic hematopoietic tissues to colonize the fetal liver, where they undergo rapid expansion and maturation[9,24]. When they first emerge, only a small fraction of AGM HSPCs engraft upon transplantation[12,60-62], suggesting they are functionally immature. FL HSPCs in the second trimester, however, are robustly transplantable, although they are still more proliferative and differ from transplantable HSPCs found in CB at birth and later in adult bone marrow[9,24], indicating that HSC maturation continues throughout human development and early life.

To identify the temporal changes in transcription programs that occur during HSC developmental maturation, our lab carried out scRNA-seq on fetal hematopoietic tissues from the first and second trimester as well as CB. This project identified a "maturation scorecard" of genes whose expression serves as a hallmark of HSC maturation[24]. scRNA-seq comparisons of PSC-derived HSPCs to first and second trimester human HSCs have revealed that they most closely resemble AGM HSCs from CS14/15, not maturing FL HSCs or CB HSCs, underscoring that their inability to engraft upon transplantation could be due to their functional immaturity. Tracking MLLT3 expression using this scRNA-seq map of human HSC development revealed induction of MLLT3 already during endothelial to hematopoietic transition in the AGM and increased

expression during HSC maturation, suggesting an earlier role in human HSC ontogeny[24]. However, scRNA analysis did not distinguish between MLLT3 isoforms. Here we show that MLLT3 isoform expression is developmentally regulated, and are expressed at different stages of HSC emergence and maturation. We show that expression of MLLT3-S is induced robustly in fetal liver HSCs and may contribute to HSC maturation by suppressing the expression of the onco-fetal gene IGFBP2. We further propose a mechanism by which MLLT3-S counteracts the effect of MLLT3-L on gene expression through the interaction with same protein partners.

Results

MLLT3-S expression is developmentally regulated during human HSC maturation

Our previous studies have shown that MLLT3 expression is enriched in human fetal liver and cord blood HSCs and maintains their stemness[63], and our new analysis shows that both MLLT3 isoforms are expressed in FL and CB HSPCs, but the timing when MLLT3 isoforms are induced is unknown. Analysis of single cell RNA seq data that captures human HSCs throughout ontogeny revealed that MLLT3 expression can already be observed during human HSC emergence. Specifically, scRNAseq analysis of human 5 week (CS14-15) AGM revealed MLLT3 expression during EHT (endothelial to hematopoietic transition) in arterial pre-HE (pre-hemogenic endothelium), HE (hemogenic endothelium) and nascent HSCs **(Figure 3.1A)**, and further upregulation during HSC maturation[24,63] **.** However, since the standard 10X Genomics scRNAseq analysis does not distinguish between MLLT3 isoforms. we analyzed bulk RNA seq data from hematopoietic tissues throughout human fetal development and defined the expression patterns for the long and short isoforms. This analysis verified induction of MLLT3-L in human 5 week (CS15)

AGM endothelium (CD34+CD90+CD43-) and HSPCs (CD34+CD90+CD43+), and minimal expression in CD34+CD90-CD43+ HPCs. In contrast, MLLT3-S expression was not detected in AGM endothelium and HPCs and was very low or undetectable in AGM HSPCs **(Figure 3.1B).** MLLT3-S expression was greatly increased in FL HSPCs during the second trimester (15-17 weeks). These data suggest that MLLT3-S is induced after HSPC emergence and is developmentally regulated in maturing human HSPCs. The induction MLLT3-S during HSC maturation was subsequently validated at single cell level by selecting HLF+ HSCs from the scRNAseq data and visualizing MLLT3-S specific peak in maturing FL HSCs **(Figure 3.1C).** These data imply that MLLT3 isoforms are developmentally regulated, with MLLT3-L being induced already during human HSC specification and emergence in hemogenic endothelium, and MLLT3-S being induced later during human HSC maturation.

Figure 3.1

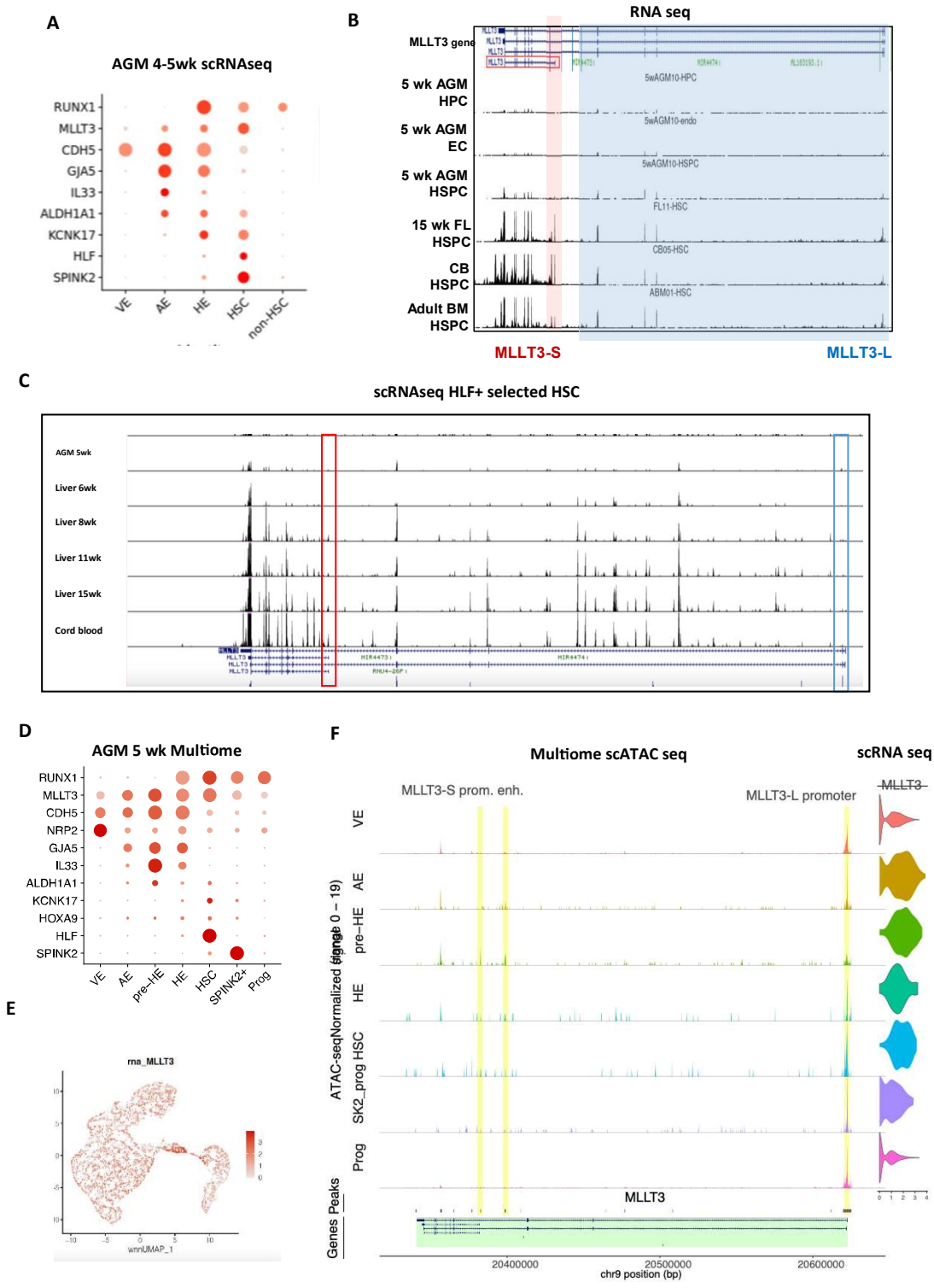
A
AGM 4-5wk scRNAseq
RUNX1
MLLT3
CDH5
GJA5
IL33
ALDH1A1
KCNK17
HLF
SPINK2
VE
AE
HE
HSC
non-HSC
B
RNA seq
MLLT3 gene
5 wk AGM HPC
5 wk AGM EC
5 wk AGM HSPC
15 wk FL HSPC
CB HSPC
Adult BM HSPC
5wAGM10-HPC
5wAGM10-endo
5wAGM10-HSPC
FL11-HSC
CB05-HSC
ABM01-HSC
MLLT3-S
MLLT3-L
C
scRNAseq HLF+ selected HSC
AGM 5wk
Liver 6wk
Liver 8wk
Liver 11wk
Liver 15wk
Cord blood
D
AGM 5 wk Multiome
RUNX1
MLLT3
CDH5
NRP2
GJA5
IL33
ALDH1A1
KCNK17
HOXA9
HLF
SPINK2
VE
AE
pre–HE
HE
HSC
SPINK2+
Prog
E
rna_MLLT3
wnnUMAP_1
F
Multiome scATAC seq
scRNA seq
MLLT3-S prom. enh.
MLLT3-L promoter
MLLT3
ATAC-seqNormalized (range 0 – 19)
VE
AE
pre–HE
HE
SK2_prog HSC
Prog
Genes
Peaks
MLLT3
20400000
20500000
20600000
chr9 position (bp)

Figure 3.1. MLLT3-S expression is developmentally regulated during human HSC maturation

A. Analysis of single cell RNA seq data from CS14-15 (4-5wk) human AGM shows induction of MLLT3 expression during EHT and in emerging HSCs. **B.** Bulk RNA-seq shows induction of MLLT3-L already in human 5 week AGM HSPCs, whereas MLLT3-S is only robustly expressed in the 2nd trimester fetal liver, cord blood and adult BM. **C**. Conversion of scRNA seq data of HLF+ expressing HSCs from 5 week AGM and 6, 8, 11 and 15 week liver and cord blood to UCSC browser session documents induction of MLLT3-S specific 5' exon in HSCs maturing in the liver. **D.** Single cell 10X Multiome analysis visualizing EHT genes in specified populations documents similar progression of EHT in nuclear RNA as in total RNA (see 4A). **E**. UMAP clustering showing that MLLT3 is expressed highly throughout EHT. **F.** Evaluation of accessibility of MLLT3 gene in Multiome shows opening of MLLT3-S regulatory elements in pre-HE, HE, and HSCs, and minimal accessibility in venous endothelium or differentiating progenitors.

MLLT3-S regulatory elements become accessible during endothelial-to-hematopoietic transition in the AGM

To track the induction of MLLT3-S during human HSC ontogeny, we carried out 10X Genomics Multiome analysis (combined scRNAseq and scATACseq) on human developmental hematopoietic tissues. Multiome analysis of 5 week AGM confirmed expected developmental transitions from arterial hemogenic endothelium to HSCs **(Figure 3.1D),** as observed with scRNA seq data **(Figure 3.1A).** Analysis of MLLT3 peaks during endothelium to hematopoietic transition revealed that MLLT3-S regulatory elements (promoter and associated HSC-specific enhancer) are

already accessible during HSC specification, most notably in pre-HE (IL33+ arterial endothelium), HE (hemogenic endothelium, CDH5+RUNX1+KCNK17+PTPRC-) and HSCs (HLF+) in 5wk AGM region, before MLLT3-S transcript can be detected **(Figure 3.1E, F, Figure S3.1A).** Venous endothelium and non-HSC hematopoietic cells did not show accessibility in MLLT3-S regulatory elements, although MLLT3-L TSS was open.

Comparison of Multiome data between 5 wk AGM, liver and 13 week FL HSC (HLF+) showed robust accessibility of both elements in 13 week FL, consistent with abundant expression of MLLT3-S specific peak in FL HSPCs in bulk RNAseq and HLF selected cells from Multiome data **(Figure S3.1 B,C).** In contrast, FL erythroid cells showed expression of MLLT3-S transcript and accessibility of MLLT3-S, but did not show accessibility of MLLT3-S HSC enhancer, but rather a different erythroid specific regulatory element distal to MLLT3-S TSS **(Figure S3.1C).**

Together these data document that MLLT3 isoforms are sequentially induced during HSC development, with MLLT3-L being induced already during EHT and MLLT3-S later during fetal development, and mature HSCs express both isoforms. Nevertheless, the earlier opening of the MLLT3-S TSS and HSC enhancer in AGM pre-HE, HE and nascent HSCs may be used as an indicator of specification to the HSC lineage. Moreover, these data raised the question of whether the later induction of MLLT3-S may be functionally linked to maturation of HSCs during development.

MLLT3 isoforms differentially regulate genes that reflect FL HSC maturation stage

To understand the significance of MLLT3-S on HSPC function during fetal stage, we performed RNAseq analysis on sorted 2nd trimester FL CD34^{+}CD38$^{-/lo}$CD90^{+} HSPCs after 4 week

culture with control vector, or MLLT3-L or MLLT3-S overexpression vector. Although cultured cells downregulate both isoforms and overexpressing the different isoforms have different effects on HSPC expansion rate, even the control cells with HSPC immunophenotype retain baseline expression of MLLT3 isoforms. As such, this experimental set up does not create HSPCs in which only one MLLT3 isoform is expressed, but a model where relative ratios of the isoforms are altered. Moreover, we observed that MLLT3-L also binds to both its own TSS and MLLT3-S TSS, and when overexpressed in cultured HSPCs, also improves MLLT3-S expression, indicating auto- and cross regulation between the two isoforms **(Figure S3.2A).**

Analyzing the differentially regulated genes in cultured FL HSPCs documented that altering levels of the two MLLT3 isoforms influences many key HSC programs, and suggested that also in FL-HSCs, MLLT3-S helps balance many important processes involved in HSC fate decisions that are regulated by MLLT3-L **(Figure S3.2B).** MLLT3-L overexpression in FL-HSPCs enhanced the expression of many HSC transcription factors (HLF, MECOM, ERG, LMO2, GATA2)[10,42-48,64-68] and protein synthesis machinery, as shown previously, whereas MLLT3-S overexpression suppressed many of these genes **(Figure 3.2A),** concordant with CB HSPC MLLT3 isoform KD studies **(Figure 2.3)**. MLLT3-L also promoted the expression of other regulatory factors expressed in developing HSCs (KIT, ADGRG6, IGFBP2),[9,24,69-71] whereas MLLT3-S suppressed them. In contrast, MLLT3-S expression within cultured FL-HSPCs led to upregulation of B lymphoid genes (IL7R, EBF1, JCHAIN), which were suppressed by MLLT3-L, implying MLLT3 isoforms can also differentially regulate HSC lineage potency or priming. Other MLLT3-S induced genes include genes involved in cell cycle checkpoints and cell adhesion and migration (ROCK2, ARHGAP5) **(Figure 3.2B**).

One of the genes most strongly differentially regulated in cultured FL HSPCs by the long and short MLLT3 isoforms was IGFBP2 (Insulin-like growth factor binding protein 2) (1.97-fold upregulated by MLLT3-L, and 4.41-fold downregulated by MLLT3-S, compared to control) **(Figure 3.2A, D, E).** IGFBP2 is an onco-fetal gene expressed in highly proliferative developmental tissues[72,73], and is reactivated in some cancers such as treatment resistant leukemias[74,75]. ScRNA seq analysis of human HSCs at different stages of ontogeny identified IGFBP2 expression as being characteristic of immature human HSPCs, and that decrease in IGFBP2 expression in liver HLF+ HSPCs between 6-8 weeks of development coincides with other molecular hallmarks of HSC maturation and acquisition of adult type functional properties[24], including gradual transition to quiescence and establishment of robust BM engraftment ability **(Figure 3.2E).** The induction of MLLT3-S expression in liver HSPCs between 5 and 6 weeks coincided with the downregulation of IGFBP2 in these cells **(Figure 3.2F).**

Although IGFBP2 expression becomes silenced in fetal liver HSCs by the second trimester, our culture data revealed that prolonged culture can reactivate the expression of IGFBP2 in FL and CB HSPCs **(Figure 3.2C, D, E).** MLLT3-L overexpression further upregulated IGFBP2 expression while MLLT3-S suppresses it in cultured FL- and CB HSPCs. Since supplementation of soluble IGFBP2 in culture has been shown to stimulate HSC expansion, these data raised the hypothesis that the differential regulation of IGFBP2 by MLLT3 isoforms in cultured HSPCs may contribute to the observed isoform-specific functional effect that MLLT3-L expression stimulates HSC expansion in culture, whereas MLLT3-S suppresses it. Similar to FL HSPCs, culture-induced upregulation of IGFBP2 was also observed in cultured CB HSPCs **(Figure 3.2D, E).**

Figure 3.2

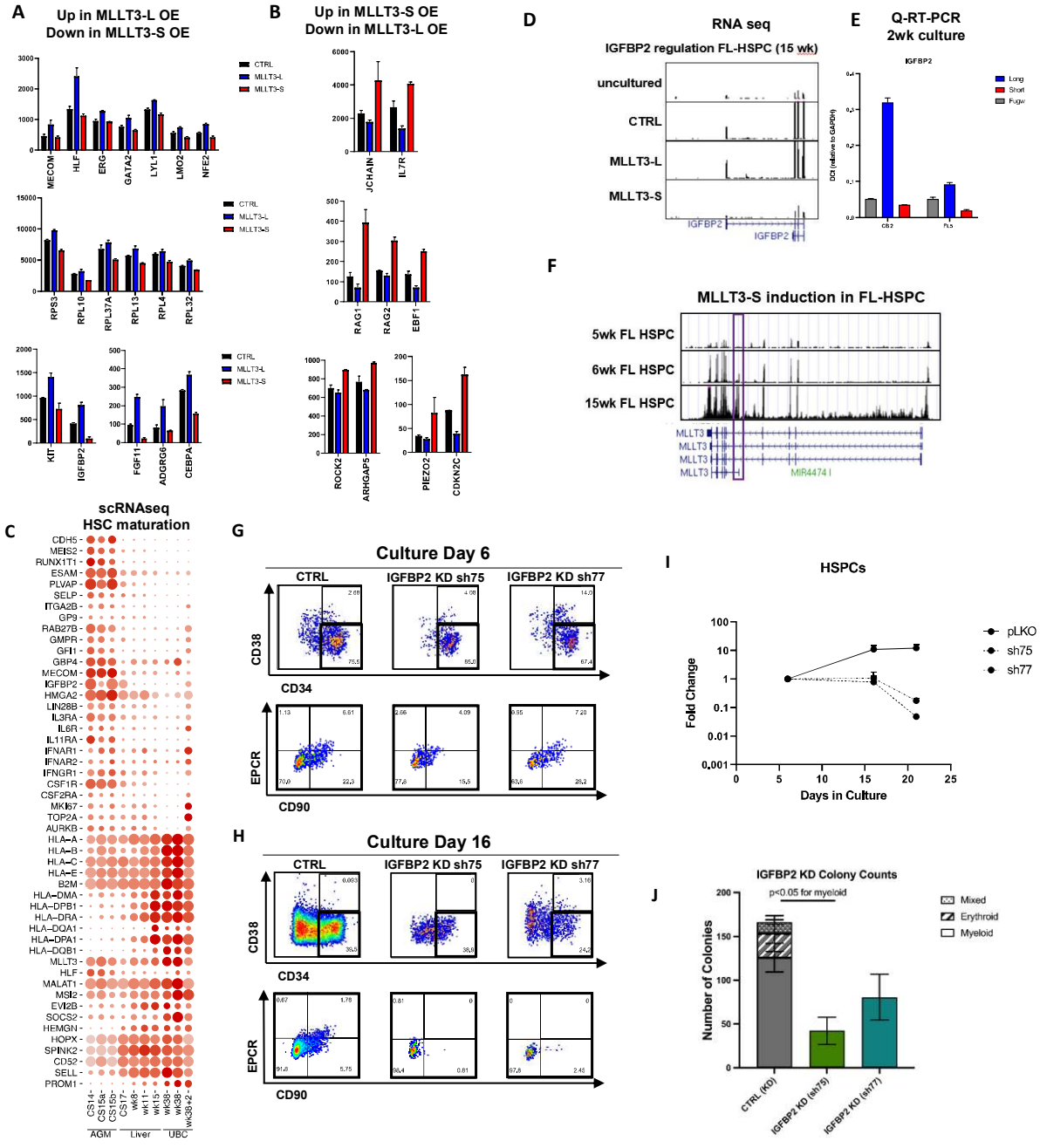
A
Up in MLLT3-L OE
Down in MLLT3-S OE
CTRL
MLLT3-L
MLLT3-S
MECOM
HLF
ERG
GATA2
LYL1
LMO2
NFE2
RPS3
RPL10
RPL37A
RPL13
RPL4
RPL32
KIT
IGFBP2
FGF11
ADGRG6
CEBPA
B
Up in MLLT3-S OE
Down in MLLT3-L OE
JCHAIN
IL7R
RAG1
RAG2
EBF1
ROCK2
ARHGAP5
PIEZO2
CDKN2C
C
scRNAseq
HSC maturation
CDH5
MEIS2
RUNX1T1
ESAM
PLVAP
SELP
ITGA2B
GP9
RAB27B
GMPR
GFI1
GBP4
MECOM
IGFBP2
HMGA2
LIN28B
IL3RA
IL6R
IL11RA
IFNAR1
IFNAR2
IFNGR1
CSF1R
CSF2RA
MKI67
TOP2A
AURKB
HLA-A
HLA-B
HLA-C
HLA-E
B2M
HLA-DMA
HLA-DPB1
HLA-DRA
HLA-DQA1
HLA-DPA1
HLA-DQB1
MLLT3
HLF
MALAT1
MSI2
EVI2B
SOCS2
HEMGN
HOPX
SPINK2
CD52
SELL
PROM1
CS14-
CS15a-
CS15b-
CS17-
wk8-
wk11-
wk15-
wk38-
wk38-
wk38+2-
AGM
Liver
UBC
D
RNA seq
IGFBP2 regulation FL-HSPC (15 wk)
uncultured
CTRL
MLLT3-L
MLLT3-S
IGFBP2
E
Q-RT-PCR
2wk culture
IGFBP2
Long
Short
Fig4w
F
MLLT3-S induction in FL-HSPC
5wk FL HSPC
6wk FL HSPC
15wk FL HSPC
MLLT3
MIR4474 I
G
Culture Day 6
CTRL
IGFBP2 KD sh75
IGFBP2 KD sh77
CD38
CD34
EPCR
CD90
H
Culture Day 16
CTRL
IGFBP2 KD sh75
IGFBP2 KD sh77
CD38
CD34
EPCR
CD90
I
HSPCs
Fold Change
Days in Culture
pLKO
sh75
sh77
J
IGFBP2 KD Colony Counts
p<0.05 for myeloid
Number of Colonies
Mixed
Erythroid
Myeloid
CTRL (KD)
IGFBP2 KD (sh75)
IGFBP2 KD (sh77)

Figure 3.2. MLLT3 isoforms regulate IGFBP2 expression during HSC development and culture

A,B. Selected genes differentially regulated by lentiviral overexpression of MLLT3 isoforms in CD34+CD38^{lo}CD90+FL HSPCs after 4 week culture, as identified by bulk RNAseq (n=2). A. shows genes upregulated by MLLT3-L and downregulated by MLLT3-S, while B. shows genes downregulated by MLLT3-L and upregulated by MLLT3-S. **C**. scRNAseq of human fetal hematopoietic tissues showing IGFBP2 downregulation in FL HSCs during maturation and upregulation in cultured CB HSPCs **D.** Bulk RNAseq shows IGFBP2 is one of the genes most strongly differentially regulated by MLLT3 isoforms. **E**. Q-RT-PCR quantification of IGFBP2 expression in cultured human FL and CB following MLLT3 isoform overexpression. **F**. Bulk RNAseq of FL HSPCs demonstrates MLLT3-S induction during fetal HSC maturation. **G,H** Flow cytometry analysis of human HSPC marker expression in the progeny of CD34^{+}CD38^{lo}CD90^{+} CB HSPCs transduced with empty vector control or knockdown vectors targeting IGFBP2 on day 6 G. or day 16 H. of culture. **I.** Quantification of CD34+CD38loCD90+ HSPC expansion following IGFBP2 KD in CB HSPCs (data from n=2 independent experimental replicates, n=4 CB samples total). **J.** Quantification of CFU-C in methylcellulose colony assay (mean and sd, data from n=2 independent replicates, n=4 CB samples total)

IGFBP2 is a key target gene differentially regulated by MLLT3 isoforms that promotes sustained human HSPC culture expansion

Since IGFBP2 is upregulated by MLLT3-L and downregulated by MLLT3-S in cultured FL and CB HSPCs, we sought to define whether cell-intrinsic IGFBP2 acting downstream of MLLT3 has a function in sustaining human HSPC expansion. CD34+CD38^{lo}CD90+ human CB HSPCs were transduced with two different previously validated IGFBP2 shRNA vectors, sh75 and sh77 and empty vector control, and transduced cells were cultured and analyzed by FACS for markers of HSPC identity (CD34+ CD38^{lo}CD90+ EPCR+) over time. Both IGFBP2 knockdown vectors severely limited CB HSPC expansion in culture **(Figure 3.2G-I)**: while control HSPCs continued to expand until day 22, IGFBP2 knockdown HSPCs showed minimal expansion of CD34+CD38^{lo}CD90+ cells from day 6 to day 16 and day 22 **(Figure 3.2I)**. This lack of expansion mimicked the phenotype of MLLT3-L KD **(Figure 3.2F)**. These data show that loss of cell-intrinsic IGFBP2 severely limits human HSPC expansion in culture.

To assess whether IGFBP2 is required for the pro-expansion effect observed with MLLT3-L overexpression in cultured human HSCs, CB HSCs were transduced with various combinations of MLLT3-L OE, IGFBP2 KD, and control vectors (pLKO for KD control, FUGW vector for OE control) and CD34+CD38^{lo}CD90+ HSPC were quantified by FACS over time. As previously observed, IGFBP2 knockdown samples showed minimal HSC expansion between day 6 and day 16 or day 21, whereas control (FUGW/pLKO) and MLLT3-L OE samples showed the greatest fold change in expansion. Importantly, IGFBP2 KD/MLLT3-L OE cells showed no significant difference in fold change ($p>0.05$, unpaired t test) compared to IGFBP2 KD samples, demonstrating that knockdown of IGFBP2 eliminates the HSC expansion effect induced by

MLLT3-L OE **(Figure S3.2C,D)**. Together, these data show that IGFBP2 expression is required for MLLT3-L to promote robust CB HSPC expansion in culture.

To assess whether IGFBP2 may contribute to the establishment of differentiated blood cell lineages (as previously observed with MLLT3), we next assessed the effect of IGFBP2 knockdown on myelo-erythroid lineage differentiation using a methylcellulose CFU assay. After 15 days in culture, IGFBP2 knockdown samples formed fewer colonies and, similar to knockdown of MLLT3-L, showed defects in mixed and erythroid colony production as compared to control samples **(Figure 3.2J)**. These data indicate that like MLLT3-L, IGFBP2 is important for maintenance of multipotent human HSPCs in culture.

To investigate the mechanism by which MLLT3 isoforms can regulate the expression of IGFBP2, ChIP-seq was performed on cultured CB CD34 cells with and without MLLT3-L overexpression. This verified that MLLT3-L, the isoform able to bind chromatin, binds the IGFBP2 gene (**Figure 3.3A**). Furthermore, ChIP-seq for H3K79me2 revealed the increased MLLT3-L binding to IGFBP2 also leads to increased H3K79me2 within this gene (**Figure 3.3A**), explaining its increased expression with MLLT3-L OE. MLLT3 is a known component of the DOT1L complex, which carries out H3K79 methylation. Thus, we can infer that increased MLLT3-L binding to IGFBP2 increases DotCom activity at this site, leading to increased IGFBP2 expression. Although MLLT3-S is missing the YEATS domain and is unable to modulate gene expression by binding chromatin, it does retain the protein-binding ANC1 homology domain (which MLLT3-L also contains). To assess whether the AHD allows MLLT3 isoforms to bind the same protein partners in the DotCom and SEC, we carried out Co-IP followed by mass spectrometry in K562 cells with either MLLT3-L or MLLT3-S OE. We found that both MLLT3-S and MLLT3-L bound the SEC components AFF1, AFF4, and ELL, as well as the DotCom

components DOT1L and MLLT10[29,31-33,40] (**Figure 3.3B,C; Table 3.1**). MLLT3-S IP also pulled down MLLT6, also a known MLLT3 protein partner and DotCom component. Based on this, we propose a model of gene expression regulation by MLLT3 isoforms competing for the same protein binding partners. When the SEC and DotCom contain MLLT3-L, they bind chromatin and carry out their respective transcription-promoting functions, while when they contain MLLT3-S, they are sequestered from chromatin and inactive **(Figure 3.3C)**. In this way, MLLT3-S can act as the dominant negative isoform of MLLT3, countering MLLT3-L-driven gene expression.

Figure 3.3

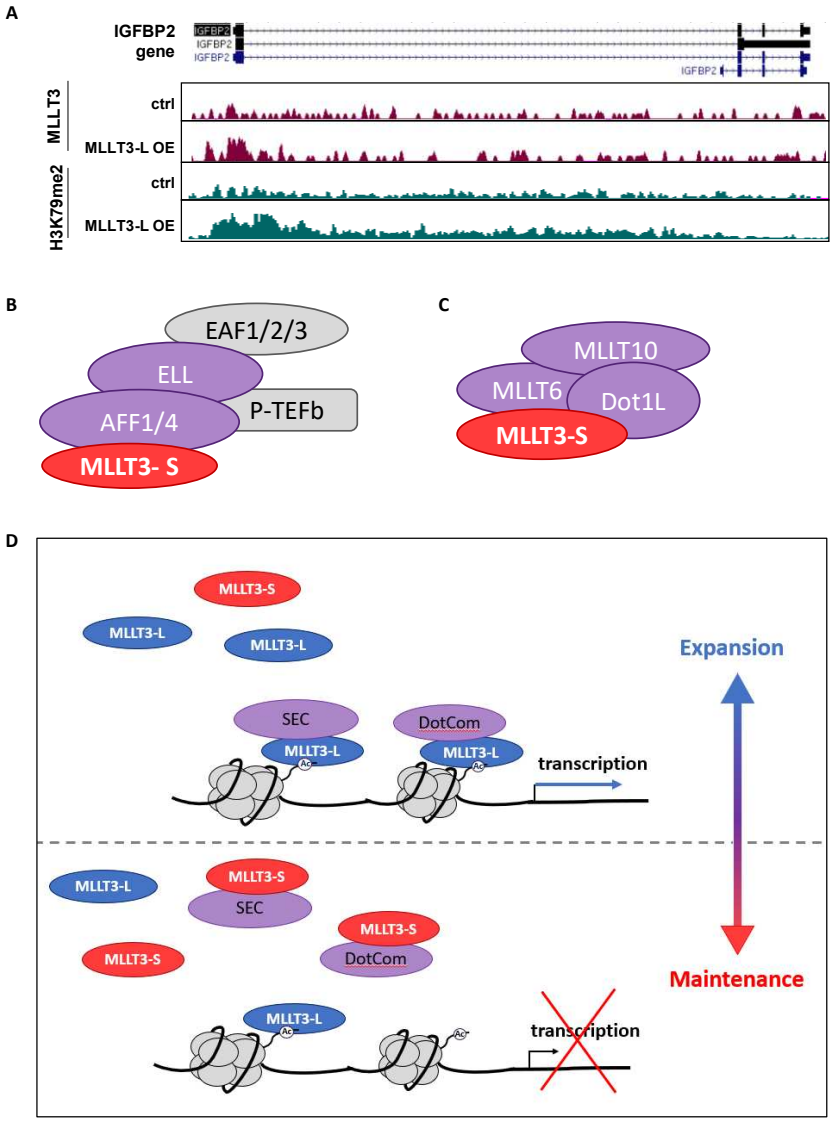

Figure 3.3

A. ChIP-seq for MLLT3-L binding and H3K79me2 in cultured human FL HSCs with an MLLT3-L overexpression or empty vector controls shows MLLT3-L binding in IGFBP2 gene and corresponding increase in H3K79me2, demonstrating increased DotCom activity. **B, C.** Diagrams showing known protein partners of MLLT3-L in the SEC (B) and DotCom (C) identified by co-IP of K562 cells with MLLT3-S OE followed by mass spectrometry. **D**. Proposed mechanism by which MLLT3-S acts as a dominant negative isoform to counteract gene expression promoted by MLLT3-L

Conclusions

We demonstrate that the distinct TSS and regulatory elements of the long and short MLLT3 isoforms allow them to be regulated separately during development. While MLLT3-L is induced early during HSC specification in the arterial hemogenic endothelium, MLLT3-S is induced in HSCs later during fetal development, at a time when developing HSCs are transitioning from an immature, highly proliferative state to a mature, homeostatic state. However, MLLT3-S regulatory elements become accessible already in pre-HE, HE and nascent HSCs in the AGM, preceding MLLT3-S expression, providing an opportunity for "epigenetic lineage tracing" to pinpoint the precursors of HSCs based on the intend to activate MLLT3-S expression.

One example of the way by which MLLT3 isoforms modulate HSC developmental regulation is through controlling IGFBP2 expression. IGFBP2 is a component of the insulin growth factor signaling pathway, although it may also have other functions such as through interactions with integrins[72,76]. IGFBP2 is highly expressed is rapidly proliferating tissues during embryonic and fetal development, and declines during development, for example as HSCs undergo developmental maturation[24]. Exogenously added IGFBP2 has been shown to stimulate HSC culture expansion [72,73], while endogenous IGFBP2 expression is known to facilitate the survival of human leukemia cells in culture and in mouse models[74], but its role as a cell-autonomous factor in normal hematopoiesis has been unclear. We demonstrate that IGFBP2 expression is upregulated again in human FL and CB HSCs as they are placed in expansion culture, and required for efficient HSPC culture expansion. While MLLT3-L binds to IGFBP2 TSS and its overexpression in culture human HSCs further upregulates its expression, MLLT3-S downregulates it. During fetal development, IGFBP2 downregulation in late embryonic period in human HSCs coincides with the induction of MLLT3-S expression, suggesting that downregulation of IGFBP2 may be a key

mechanism by which MLLT3-S contributes to fetal HSC maturation that is required for robust bone marrow engraftment ability.

The importance of both isoforms is further underscored by our finding that they not only modulate each other's function likely by sharing the ANC-homology domain that enables protein interactions, but they can also influence each other's expression. MLLT3-L, the only isoform with the chromatin binding YEATS domain, binds to both its own TSS and MLLT3-S TSS, helping to maintain MLLT3-S expression and ensure the faithful function of human HSCs that are expanded ex vivo with MLLT3-L overexpression. Furthermore, these MLLT3 isoforms also likely modulate each other's function through their shared ANC1 homology domain, which allows both isoforms to interact with known MLLT3 protein partners in the SEC and Dot1L complexes. By binding with MLLT3-L protein partners but failing to bind chromatin and target the DotCom to genes normally targeted by the MLLT3-L YEATS domain, MLLT3-S may act as a dominant negative isoform of MLLT3, opposing the effects of MLLT3-L on HSC gene expression and behavior.

Figure S3.1

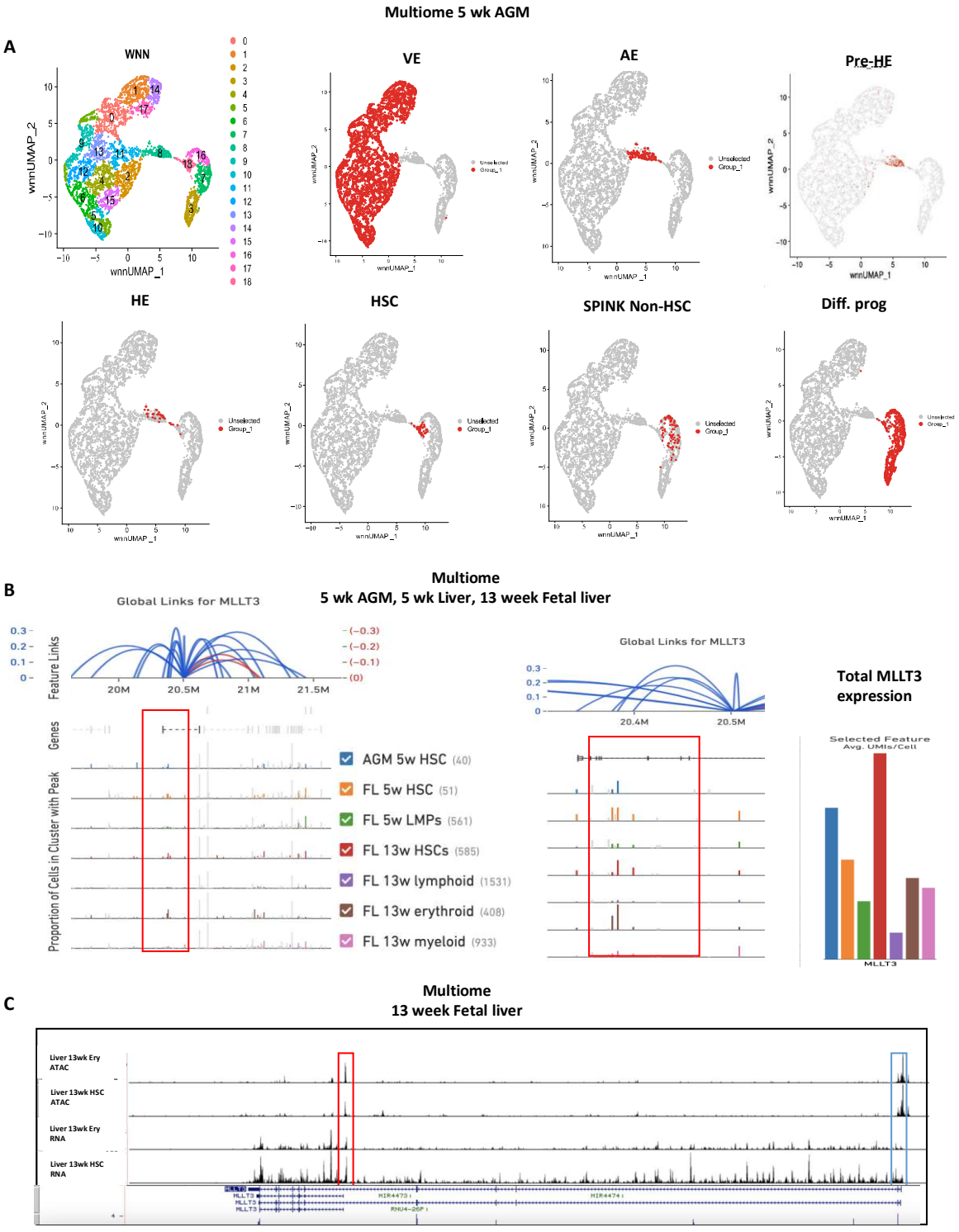
Multiome 5 wk AGM
A
WNN
VE
AE
Pre-HE
HE
HSC
SPINK Non-HSC
Diff. prog
wnnUMAP_1
wnnUMAP_2
Unselected
Group_1
B
Multiome
5 wk AGM, 5 wk Liver, 13 week Fetal liver
Global Links for MLLT3
Feature Links
Genes
Proportion of Cells in Cluster with Peak
AGM 5w HSC (40)
FL 5w HSC (51)
FL 5w LMPs (561)
FL 13w HSCs (585)
FL 13w lymphoid (1531)
FL 13w erythroid (408)
FL 13w myeloid (933)
Total MLLT3 expression
Selected Feature
Avg. UMIs/Cell
MLLT3
C
Multiome
13 week Fetal liver
Liver 13wk Ery ATAC
Liver 13wk HSC ATAC
Liver 13wk Ery RNA
Liver 13wk HSC RNA

Figure S3.1. Multiome analysis of human HSC specification. Related to Figure 3.1

A. Single cell 10X Multiome analysis of 5 week human AGM cells. UMAP clustering and selecting specific cell types using EHT markers documents progression from arterial pre-HE and HE to HSCs. **B.** Single cell 10X Multiome analysis comparing human 5 week AGM and liver, and 13 week liver tissues documents accessibility of MLLT3-S regulatory elements already in nascent HSCs in 5 week AGM and liver, preceding MLLT3-S transcript. **C.** Single cell 10X Multiome analysis comparing of HLF+ HSCs and GATA1+ erythroid cells in 13 week liver documents accessibility of MLLT3-S TSS and cell type specific putative enhancer in erythroid cells and HSCs in ATAC seq, and RNA peak corresponding to MLLT3-S 5'UTR in both erythroid cells and HSCs.

Figure S3.2

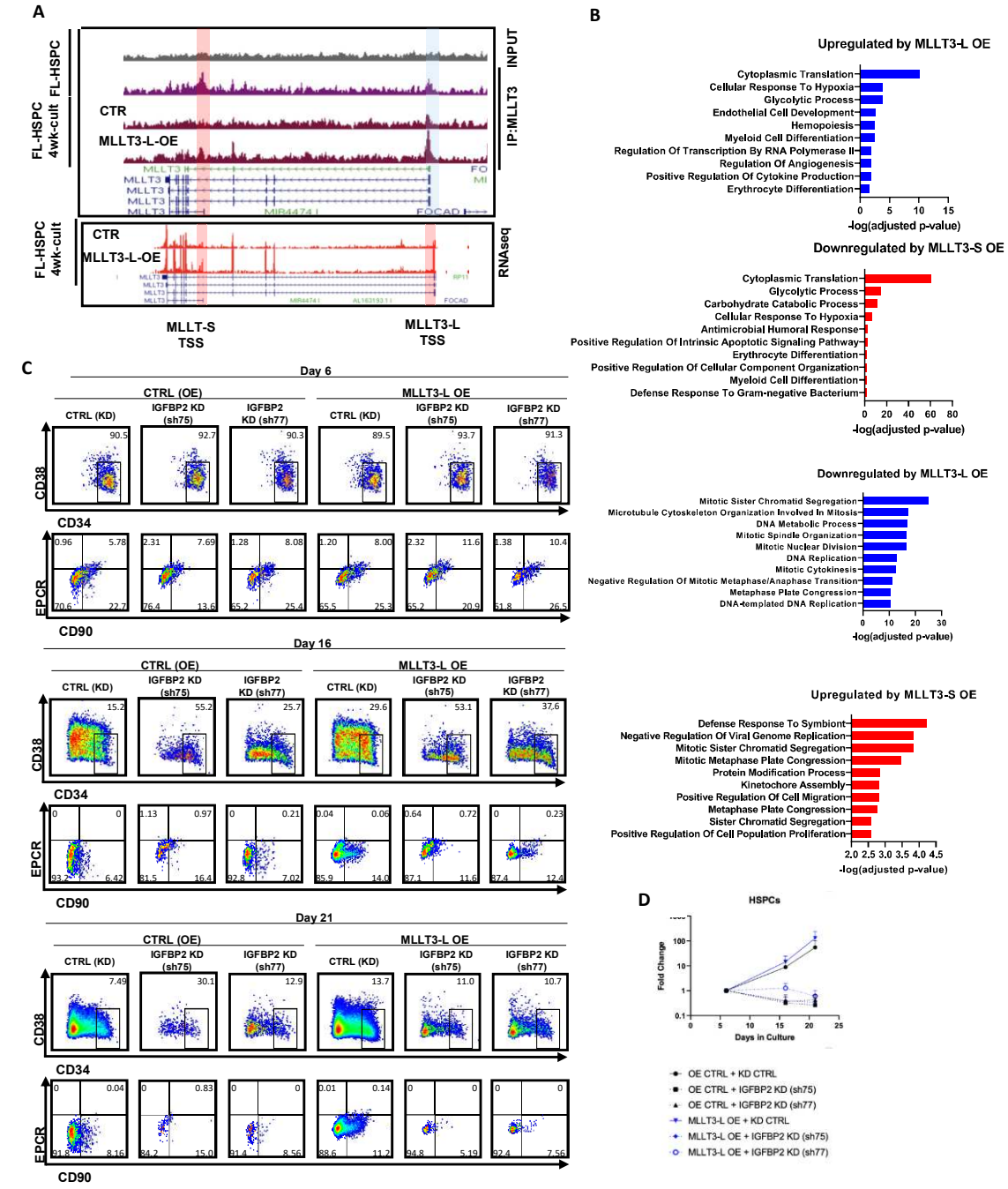
A
FL-HSPC
FL-HSPC 4wk-cult
CTR
MLLT3-L-OE
INPUT
IP:MLLT3
RNAseq
MLLT-S TSS
MLLT3-L TSS
B
Upregulated by MLLT3-L OE
Cytoplasmic Translation
Cellular Response To Hypoxia
Glycolytic Process
Endothelial Cell Development
Hemopoiesis
Myeloid Cell Differentiation
Regulation Of Transcription By RNA Polymerase II
Regulation Of Angiogenesis
Positive Regulation Of Cytokine Production
Erythrocyte Differentiation
-log(adjusted p-value)
Downregulated by MLLT3-S OE
Cytoplasmic Translation
Glycolytic Process
Carbohydrate Catabolic Process
Cellular Response To Hypoxia
Antimicrobial Humoral Response
Positive Regulation Of Intrinsic Apoptotic Signaling Pathway
Erythrocyte Differentiation
Positive Regulation Of Cellular Component Organization
Myeloid Cell Differentiation
Defense Response To Gram-negative Bacterium
-log(adjusted p-value)
Downregulated by MLLT3-L OE
Mitotic Sister Chromatid Segregation
Microtubule Cytoskeleton Organization Involved In Mitosis
DNA Metabolic Process
Mitotic Spindle Organization
Mitotic Nuclear Division
DNA Replication
Mitotic Cytokinesis
Negative Regulation Of Mitotic Metaphase/Anaphase Transition
Metaphase Plate Congression
DNA-templated DNA Replication
-log(adjusted p-value)
Upregulated by MLLT3-S OE
Defense Response To Symbiont
Negative Regulation Of Viral Genome Replication
Mitotic Sister Chromatid Segregation
Mitotic Metaphase Plate Congression
Protein Modification Process
Kinetochore Assembly
Positive Regulation Of Cell Migration
Metaphase Plate Congression
Sister Chromatid Segregation
Positive Regulation Of Cell Population Proliferation
-log(adjusted p-value)
C
Day 6
Day 16
Day 21
CTRL (OE)
MLLT3-L OE
CTRL (KD)
IGFBP2 KD (sh75)
IGFBP2 KD (sh77)
CD38
CD34
EPCR
CD90
D
HSPCs
Fold Change
Days in Culture
OE CTRL + KD CTRL
OE CTRL + IGFBP2 KD (sh75)
OE CTRL + IGFBP2 KD (sh77)
MLLT3-L OE + KD CTRL
MLLT3-L OE + IGFBP2 KD (sh75)
MLLT3-L OE + IGFBP2 KD (sh77)

Figure S3.2. Identification and validation of programs downstream of MLLT3 isoforms. Related to Figure 3.2.

A. CHIP-seq and Bulk RNA-seq of MLLT3 binding and isoform expression in cultured FL HSPCs documents MLLT3-L binding to both its own TSS and MLLT3-S TSS, and inducing MLLT3-S expression. Representative examples of n=3. **B.** Top GO categories that are either up- or downregulated following overexpression of MLLT3-S C. or MLLT3-L **C.** Flow cytometry analysis of human HSPC marker expression in the progeny of $CD34^{+}CD38^{lo}CD90^{+}$ CB HSPCs transduced with empty vector control or vectors for MLLT3-L overexpression and knockdown vectors targeting IGFBP2 isoforms on day 6, day 16, or day 21 of culture. **D.** Quantification of HSPC expansion upon IGFBP2-KD and MLLT3 overexpression. (data from n=2 independent experimental replicates, n=4 CB samples total).

Table S3.1 SEC and DotCom proteins bound by MLLT3 isoforms identified by co-immunoprecipitation and mass spectrometry in K562 cells

MLLT3-L	MLLT3-S
AFF1	AFF1
AFF4	AFF4
MLLT10	MLLT10
DOT1L	DOT1L
ELL	ELL
	MLLT6

Bibliography

1. Calvanese, V., Capellera-Garcia, S., Ma, F., Fares, I., Liebscher, S., Ng, E.S., Ekstrand, S., Aguadé-Gorgorió, J., Vavilina, A., and Lefaudeux, D. (2022). Mapping human haematopoietic stem cells from haemogenic endothelium to birth. Nature *604*, 534-540.
2. Calvanese, V., and Mikkola, H. (2023). The genesis of human hematopoietic stem cells. Blood Journal, blood. 2022017934.
3. Ding, J., Li, Y., and Larochelle, A. (2023). De Novo Generation of Human Hematopoietic Stem Cells from Pluripotent Stem Cells for Cellular Therapy. Cells *12*, 321. 10.3390/cells12020321.
4. Ditadi, A., Sturgeon, C.M., and Keller, G. (2017). A view of human haematopoietic development from the Petri dish. Nature Reviews Molecular Cell Biology *18*, 56-67. 10.1038/nrm.2016.127.
5. Ivanovs, A., Rybtsov, S., Ng, E.S., Stanley, E.G., Elefanty, A.G., and Medvinsky, A. (2017). Human haematopoietic stem cell development: from the embryo to the dish. Development *144*, 2323-2337. 10.1242/dev.134866.
6. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631-644. 10.1016/j.cell.2008.01.025.
7. Zhang, Y., Gao, S., Xia, J., and Liu, F. (2018). Hematopoietic hierarchy–an updated roadmap. Trends in cell biology *28*, 976-986.
8. Van Handel, B., Prashad, S.L., Hassanzadeh-Kiabi, N., Huang, A., Magnusson, M., Atanassova, B., Chen, A., Hamalainen, E.I., and Mikkola, H.K. (2010). The first trimester human placenta is a site for terminal maturation of primitive erythroid cells. Blood, The Journal of the American Society of Hematology *116*, 3321-3330.

9. Zeng, Y., He, J., Bai, Z., Li, Z., Gong, Y., Liu, C., Ni, Y., Du, J., Ma, C., and Bian, L. (2019). Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. Cell research *29*, 881-894.

10. Ghosn, E., Yoshimoto, M., Nakauchi, H., Weissman, I.L., and Herzenberg, L.A. (2019). Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. Development *146*, dev170571.

11. Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., and Li, Z. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. Nature *533*, 487-492.

12. Tavian, M., Hallais, M.-F., and Péault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. Development *126*, 793-803.

13. Müller, A.M., Medvinsky, A., Strouboulis, J., Grosveld, F., and Dzierzakt, E. (1994). Development of hematopoietic stem cell activity in the mouse embryo. Immunity *1*, 291-301.

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

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

16. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., Crooks, G.M., et al. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281-286. 10.1038/s41586-019-1790-2.

17. Loughran, S.J., Kruse, E.A., Hacking, D.F., De Graaf, C.A., Hyland, C.D., Willson, T.A., Henley, K.J., Ellis, S., Voss, A.K., and Metcalf, D. (2008). The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nature immunology *9*, 810-819.

18. Taoudi, S., Bee, T., Hilton, A., Knezevic, K., Scott, J., Willson, T.A., Collin, C., Thomas, T., Voss, A.K., and Kile, B.T. (2011). ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. Genes & development *25*, 251-262.

19. Tipping, A.J., Pina, C., Castor, A., Hong, D., Rodrigues, N.P., Lazzari, L., May, G.E., Jacobsen, S.E.W., and Enver, T. (2009). High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. Blood, The Journal of the American Society of Hematology *113*, 2661-2672.

20. De Pater, E., Kaimakis, P., Vink, C.S., Yokomizo, T., Yamada-Inagawa, T., van der Linden, R., Kartalaei, P.S., Camper, S.A., Speck, N., and Dzierzak, E. (2013). Gata2 is required for HSC generation and survival. Journal of Experimental Medicine *210*, 2843-2850.

21. Lim, K.-C., Hosoya, T., Brandt, W., Ku, C.-J., Hosoya-Ohmura, S., Camper, S.A., Yamamoto, M., and Engel, J.D. (2012). Conditional Gata2 inactivation results in HSC loss and lymphatic mispatterning. The Journal of clinical investigation *122*, 3705-3717.

22. Vicente, C., Conchillo, A., García-Sánchez, M.A., and Odero, M.D. (2012). The role of the GATA2 transcription factor in normal and malignant hematopoiesis. Critical reviews in oncology/hematology *82*, 1-17.

23. El Omari, K., Hoosdally, S.J., Tuladhar, K., Karia, D., Vyas, P., Patient, R., Porcher, C., and Mancini, E.J. (2011). Structure of the leukemia oncogene LMO2: implications for the

assembly of a hematopoietic transcription factor complex. Blood, The Journal of the American Society of Hematology *117*, 2146-2156.

24. Kjeldsen, E., Veigaard, C., Aggerholm, A., and Hasle, H. (2018). Congenital hypoplastic bone marrow failure associated with a de novo partial deletion of the MECOM gene at 3q26. 2. Gene *656*, 86-94.

25. Voit, R.A., and Sankaran, V.G. (2023). MECOM Deficiency: from Bone Marrow Failure to Impaired B-Cell Development. Journal of Clinical Immunology *43*, 1052-1066.

26. Zhang, Y., Stehling-Sun, S., Lezon-Geyda, K., Juneja, S.C., Coillard, L., Chatterjee, G., Wuertzer, C.A., Camargo, F., and Perkins, A.S. (2011). PR-domain–containing Mds1-Evi1 is critical for long-term hematopoietic stem cell function. Blood, The Journal of the American Society of Hematology *118*, 3853-3861.

27. Lehnertz, B., Chagraoui, J., MacRae, T., Tomellini, E., Corneau, S., Mayotte, N., Boivin, I., Durand, A., Gracias, D., and Sauvageau, G. (2021). HLF expression defines the human hematopoietic stem cell state. Blood, The Journal of the American Society of Hematology *138*, 2642-2654.

28. Komorowska, K., Doyle, A., Wahlestedt, M., Subramaniam, A., Debnath, S., Chen, J., Soneji, S., Van Handel, B., Mikkola, H.K., and Miharada, K. (2017). Hepatic leukemia factor maintains quiescence of hematopoietic stem cells and protects the stem cell pool during regeneration. Cell reports *21*, 3514-3523.

29. Kent, D., Copley, M., Benz, C., Dykstra, B., Bowie, M., and Eaves, C. (2008). Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway. Clinical Cancer Research *14*, 1926-1930.

30. Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., and Meira, A.R. (2021). Transitions in lineage specification

and gene regulatory networks in hematopoietic stem/progenitor cells over human development. Cell Reports *36*.

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

32. Huynh, H., Iizuka, S., Kaba, M., Kirak, O., Zheng, J., Lodish, H.F., and Zhang, C.C. (2008). Insulin-like growth factor-binding protein 2 secreted by a tumorigenic cell line supports ex vivo expansion of mouse hematopoietic stem cells. Stem cells *26*, 1628-1635.

33. Zhang, C.C., Kaba, M., Iizuka, S., Huynh, H., and Lodish, H.F. (2008). Angiopoietin-like 5 and IGFBP2 stimulate ex vivo expansion of human cord blood hematopoietic stem cells as assayed by NOD/SCID transplantation. Blood, The Journal of the American Society of Hematology *111*, 3415-3423.

34. Chen, X., Zheng, J., Zou, Y., Song, C., Hu, X., and Zhang, C.C. (2013). IGF binding protein 2 is a cell-autonomous factor supporting survival and migration of acute leukemia cells. Journal of Hematology & Oncology *6*, 1-10.

35. Kühnl, A., Kaiser, M., Neumann, M., Fransecky, L., Heesch, S., Radmacher, M., Marcucci, G., Bloomfield, C.D., Hofmann, W.-K., Thiel, E., and Baldus, C.D. (2011). High expression of IGFBP2 is associated with chemoresistance in adult acute myeloid leukemia. Leukemia Research *35*, 1585-1590. 10.1016/j.leukres.2011.08.006.

36. He, N., Chan, C.K., Sobhian, B., Chou, S., Xue, Y., Liu, M., Alber, T., Benkirane, M., and Zhou, Q. (2011). Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. Proceedings of the National Academy of Sciences *108*, E636-E645.

37. Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S.A., and Bushweller, J.H. (2013). Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding. Structure *21*, 176-183.

38. Mohan, M., Herz, H.-M., Takahashi, Y.-H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes & development *24*, 574-589.

39. Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nature reviews Molecular cell biology *13*, 543-547.

40. Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output. Molecular and Cellular Biology *32*, 2608-2617. 10.1128/mcb.00182-12.

41. Li, T., Forbes, M.E., Fuller, G.N., Li, J., Yang, X., and Zhang, W. (2020). IGFBP2: integrative hub of developmental and oncogenic signaling network. Oncogene *39*, 2243-2257.

Chapter 4

Materials and Methods

Methods

Human hematopoietic tissue collection and processing

Human HSPC from cord blood, fetal liver and AGM and were used to assess the expression and function of MLLT3 isoforms in human HSC. Cord blood (CB) units were obtained from full-term pregnancies following informed consent and de-identified upon collection. Second trimester (14-18 weeks) fetal livers (FL) and first trimester AGM and embryonic liver (5-8 weeks) were de-identified, discarded material obtained from elective terminations of pregnancy following informed consent. Specimen age is denoted as developmental age, two weeks less than gestational age, as determined by ultrasound or estimated from the last menstrual period. Human developmental tissues were mechanically dissociated using scalpels and syringes, digested in 2.5 U dispase (Thermofisher Scientific 17105-041), 90 mg collagenase A (Worthington LS004176) and 0.075 mg DNase I (Sigma-Aldrich D4513) per ml in DPBS containing 10% FBS, for 20–45 min at 37 °C. Cells were disaggregated by pipetting and filtered through a 70 μm cell strainer. Single cell suspension from FL and CB were enriched for mononuclear cells by Lymphoprep (Stem Cells Technologies, SCT) and centrifugation following manufacturer instructions. $CD34^+$ cells were magnetically isolated from mononuclear cell fraction using anti-CD34 microbeads (Miltenyi Biotech).

Ethical Statement

First trimester tissues were obtained from University of Tubingen and delivered to UCLA within 48 h after the procedure. The Ethics Committee at the Medical Faculty of the Eberhard Karls University Tübingen and at the University Hospital in Tübingen approved the use of human embryo tissues from elective terminations for hematopoietic stem cell research (#290/2016BO1).

Second trimester fetal liver tissues were obtained from elective terminations performed at UCLA, or at Family Planning Associates (FPA) and provided to the UCLA CFAR Cell and Gene Therapy core for distribution to UCLA investigators. All human fetal tissue used were discarded material from elective terminations that were obtained following informed consent. The donated human fetal tissues were anonymized and did not carry any personal identifiers. In all cases, the decision to terminate the pregnancy occurred before the decision to donate tissue. No payments were made to donors and the donors knowingly and willingly consented to provide research materials without restrictions for research and for use without identifiers. The UCLA IRB determined that the provision of anonymized fetal material for research does not constitute human subjects research per the US Federal regulations because the tissues are anonymized, e.g., provided without any direct or indirect identifiers that could be linked back to a living individual. As a result, investigators using such material are not engaged in research subject to IRB oversight. All donors gave informed consent in compliance with US Public Health Service Act, Sections 498A and 498B for the use of fetal material in research. All human tissue materials were treated as Biosafety level 2 and approved by UCLA Institutional Biosafety Committee (IBC) (BUA-2016-142-001, BUA-2019-186-001). Studies using human iPSC were approved by The Royal Children's Hospital Human Research Ethics Committee (reference 33001A).

Serum free expansion medium (SFEM) culture for human HSPC expansion

To assess human HSPC function in serum-free, stroma-free culture conditions, HSPCs were plated in StemSpan SFEM II (SCT) supplemented with human SCF (100 ng/ml), human FLT3-L (100 ng/ml), human TPO (50 ng/ml), human low-density lipoprotein (10 μg/ml, SCT), P/S/G, 500nM StemRegenerin1 (SR1, SCT) and 35nM UM171 (SCT). Cells were cultured at 37^{o}C

and 5% CO_2 and re-plated or analyzed/sorted by flow cytometry for human HSPC markers every 7-14 days. Half of the HSC-medium was replaced every other day.

OP9M2 stroma co-culture for HSPC expansion

Human HSC co-culture on OP9M2 (subclone of OP94[4]) stroma was used to for maintenance/expansion of immunophenotypic HSPC in some experiments. OP9M2 cells were irradiated (20 Gy) and pre-plated (50,000 cells/cm^2) 24 hours before start of co-culture in OP9-medium, which includes α-MEM (Invitrogen), 20% FBS (Omega), 1X Penicillin/Streptomycin/Glutamine (P/S/G). Human HSPC were plated on stromal layer in OP9-medium supplemented with human HSC cytokines SCF (25 ng/ml, Peprotech or Invitrogen), FLT3-L (25 ng/ml, Peprotech) and TPO (25 ng/ml, Peprotech) (HSC-medium). Cells were co-cultured at 37°C and 5% CO_2 and re-plated or analyzed/sorted by flow cytometry for HSPC markers ($CD34^+CD38^{-/low}CD90^+GPI80^+$) every 7-14 days. Half of the HSC-medium was replaced every 2-3 days. Where indicated, 500nM StemRegenerin1 (SR1) and 35nM UM171 (SCT) were added to the cultures to improve human HSPC expansion.

Flow cytometry and cell sorting

Flow cytometry analysis was performed using single cell suspensions from human embryonic hematopoietic tissues, FL and CB, or obtained from NSBGW mice transplanted with human hematopoietic cells.

For identification of human HSPCs, cells were stained with mouse anti-human monoclonal antibodies against CD34-APC cl. 581 (555824; BD, 1:10) or -BV605 (343529; Biolegend, 1:25), CD90-FITC cl. 5E10 (555595; BD, 1:25) or -APC (555595; BD, 1:25) or -BV421 cl. 5E10 (562556, BD, 1:25), CD38-PE-Cy7 or -BUV496cl., HIT2 (560677 and 564657; BD, 1:100), GPI-

80-PE cl. 3H9 (D087-5; MBL, used at 1:25), and rat anti-human monoclonal antibodies against CD201-APC cl RCR-401 (351906, Biolegend, 1:20) Cells were assayed on a BD-LSRII or BD-Fortessa flow cytometer and data were analyzed with FlowJo software (Tree Star Inc.). Cell sorting was performed using a BD FACS Aria II.

To monitor human hematopoietic engraftment *in vivo* in NBSGW mice, BM cells were stained with Rat anti-mouse-CD45-APC-H7 cl. 30-F11 (557659; BD, 1:100) and mouse anti-human monoclonal antibodies against human-CD45-BV711. Hematopoietic differentiation was assessed *in vivo* and *in vitro* using mouse anti-human monoclonal antibodies against human CD19-PE or BV605 cl. 1D3 or HIB19 (12-0193, 12-0199; eBiosciences, 1:50), CD66b-BV421 cl. G10F5 (555724; BD, 1:50), CD14-V500 cl. WM53 (561816; BD, 1:100), CD3-PE-Cy7 cl. SK7 (557851, BD) (eBiosciences, 1:50), CD4-APC (MHCD0405; Invitrogen, 1:50), CD8-PE cl. HIT8A (555635; BD, 1:50), CD235a-PE or -APC (Glycophorin A, HIR2, BD 1:100), CD71-AF647 (DF1513, Santa Cruz 1:100), CD41a-APC cl. HIP8 (579777; BD, 1:20), CD42b-PEcy7 cl. HIP1 (303916; Biolegend, 1:100). Dead cells were excluded with 7-amino-actinomycin D (7AAD) (BD Biosciences, used at 1:50).

RNA isolation, cDNA synthesis and quantitative reverse transcriptase PCR

RNA isolation was performed using the RNeasy Mini kit (Qiagen) with additional DNase step using manufacturer's protocol. cDNAs were prepared using High-Capacity cDNA Reverse Transcription Kit (Thermo-Fisher), and qPCR for MLLT3 isoforms, GAPDH, IGFBP2, was performed with the LightCycler 480 SYBR Green I Master Mix (Roche) or TaqMan Gene Expression Master Mix on the Lightcycler 480 (Roche). Primers are presented in Table S4.1.

Production of lentiviral shRNA and overexpression vectors

shRNA experiments were performed with pLKO.1 lentiviral vectors from the TRC library containing puromycin resistance gene. shRNA TRCN0000005793 (#93, Sigma) and shRNA TRCN0000005792 (#92, Sigma) were selected for *in vitro* and *in vivo* experiments after testing for isoform specific knockdown efficacy. shRNA3 and shRNA4 for MLLT3-S knockdown were designed using InvivoGen siRNA/shRNA Wizard (https://www.invivogen.com/sirna-wizard), custom ordered in the pLKO.1 vector from Sigma-Aldrich, and selected for *in vitro* and *in vivo* experiments after testing for knockdown efficacy against MLLT3-S. Previously validated shRNA for IGFBP2 knockdown (TRCN0000006575 and TRCN0000006575, Millipore Sigma) were used.

Human MLLT3 was cloned from human FL-HSPC full-length cDNA into the constitutive FUGW (Addgene plasmid # 14883, from David Baltimore) lentiviral vector, where its expression is controlled by the *UBC* promoter. MLLT3 cDNA with a C-terminal V5-tag was inserted downstream and in frame with the eGFP sequence with the synthetic addition a P2A sequence between the 2 ORFs, through two rounds of PCR using PfuUltra II Ultra Fusion HS (Agilent). The MLLT3-S ORF was cloned in the FUGW vector using the same strategy. For the FUGW-GFP-P2A-MLLT3-L-V5 vector, eGFP was swapped for mApple using the InFusion cloning kit (Takara)

Human MLLT3-L and MLLT3-S were cloned from human FL-HSPC full-length cDNA into the constitutive pLVX-EF1α-IRES-mCherry (631987, Takara Bio) and pLVX-EF1α-IRES-ZsGreen1 (631982, Takara Bio) lentiviral vectors, where their expression is controlled by the EF1α promoter. Amplified MLLT3-S with a C-terminal V5 tag and MLLT3-L with a C-terminal FLAG tag was inserted into the multiple cloning site upstream of the IRES and mCherry/ZsGreen.

A summary of the plasmids included in this paper and their usage can be found in TableS4.2

For lentiviral vector production, 20 million 293T cells were co-transfected with deltaR8.2 packaging plasmid, VSVG-envelope plasmid, the lentiviral vector plasmid of choice and Turbo DNAfectin 3000 (Lamda), following manufacturer's instructions, in DMEM supplemented with 10% FBS. After incubation for 48 hours supernatant was filtered and concentrated by ultracentrifugation, and pelleted viruses were resuspended in SFEM and stored at -80ºC. Alternatively, viral supernatants were produced by the UCLA Center for Systems Biomedicine Vector Core.

Lentiviral transduction

Sorted FL and CB HSPC were prestimulated 24 h in StemSpan SFEM II supplemented with SCF, FLT3-L, TPO, human LDL (SCT), and Antibiotic-Antimycotic (LT). Wells were pre-coated with 50 ug/mL RetroNectin (Takara) and lentivirus was added. After centrifugation at 2000g for 2 hours at 32ºC, wells were washed with PBS containing 5% BSA, and HSPCs in 100ul StemSpan II SFEM supplemented with Lentiboost (SB-P-LV-101-10, Mayflower Bioscience, 1:100) were seeded and centrifuged at 800g for 1.5 hours. After 24 hours from cell seeding, transduced cells were washed and seeded in the indicated conditions for each assay. For shRNA-lentiviral vectors, puromycin (1.0 μg/ml) treatment was used for selection of transduced HSPC and maintained throughout culture. Cells infected with overexpression vectors were selected by cell sorting for fluorescent protein expression.

5'RACE for identification of MLLT3-S variants

5'RACE was performed using the SMARTer™ RACE 5'/3' Kit (634860, Takara Bio) according to the manufacturer's instructions. Briefly, a primer specific to Exon 7 of MLLT3 was designed and used for amplification of 5'RACE-ready cDNA from CB CD34+ CD38^{lo}CD90+ sorted cells.

The products were gel-purified and cloned into TOPO-TA vector using the TOPO-TA Cloning Kit (K4575J10, Invitrogen), and sequenced using the M13R primer (Laragen).

Hematopoietic Colony assays

Cultured cells were transduced with lentiviral knockdown vectors and selected via culture in StemSpan SFEM II supplemented with puromycin (1.0 μg/ml) for three days. At day 6, transduced cell numbers were quantified using flow cytometry, and 500 CD34+ cells from each condition plated in duplicate in 1.1mL of methylcellulose MethoCult H4434 Classic media. Colonies were manually counted under 5X magnification (Zeiss Axiovert 40 CFL light microscope) and scored as erythroid, myeloid, or mixed colonies on day 13 and day 15, and cells were harvested for final FACS analysis (BD LSRFortessa flow cytometer and FlowJo software) on day 17.

Culture of Human Hematopoietic Cell Lines

Human K562 cells were cultured in IMDM (Invitrogen) supplemented with P/S and 10% FBS. Human KG1 cells were cultured in RPMI 1640 (Invitrogen) supplemented with P/S and 10% FBS.

Subcellular Fractionation

Subcellular fractionation was carried out using Subcellular Protein Fractionation Kit for Cultured Cells (Thermo Scientific) in accordance with manufacturer instructions. After measuring protein concentration, the same amount of protein from each sample/fraction was used for protein quantification by Western blot.

Western Blot

Cultured cells were pelleted and washed twice with ice cold PBS, then lysed in RIPA Buffer (Invitrogen) with 1x Halt Protease Inhibitor Cocktail (Thermo Scientific). Protein was quantified by Pierce BCA assay (Thermo Scientific). Protein extracts were loaded on SDS-polyacrylamide (SDS-PAGE) gel and blotted to a polyvinylidene fluoride (PVDF) membrane. PVDF membrane was incubated at 4ºC overnight with the following antibodies: rabbit anti-human MLLT3 (only binds MLLT3-L) (ab154492, Abcam, 1:1000), mouse anti-V5-tag cl. SV5-Pk1 (R960-25, Invitrogen, 1:500), rabbit anti-Actin cl. EPR16769 ab179467 (ab179467, Abcam, 1:5000), rabbit anti-LaminB1 cl. PR8985B ab133741 (ab133741, Abcam, 1:5000), and rabbit anti-H3 cl. D1H2 #4499 (4499T, Cell Signalling Technology, 1:5000). Secondary staining was performed for 1hr at RT using either Amersham ECL Rabbit IgG HRP-linked whole Ab (NA934-1ML, Cytiva Life Sciences, 1:2500) or Amersham ECL Mouse IgG HRP-linked whole Ab (NA931-1ML, Cytiva Life Sciences, 1:2500). Blots were imaged using West Pico PLUS Chemiluminescent Substrate (Thermo Scientific) using the BioRad Chemidoc.

Transplantation assays in NBSGW mice

In vivo reconstitution ability of MLLT3 isoform KD HSPC was assessed in immunodeficient mice. Female NOD.Cg-KitW-41J Tyr + Prkdcscid Il2rgtm1Wjl/ThomJ (NBSGW, Jackson Laboratories) mice, 8-12 weeks old, were retro-orbitally injected with CB-derived cells in a volume of 100 µL of RPMI.

For MLLT3 shRNA knockdown experiments, 30,000 CB $CD34^+$ $CD38^{lo}$CD90+ cells were infected with MLLT3 isoform specific shRNA or control lentiviral vectors and cultured on HSC media for 2 days under puromycin selection. Prior to transplantation, a sample of the cells was

analyzed and counted by flow cytometry to ensure mice in all conditions received the same number of CD34+ cells. At 10 and/or 18 weeks, MLLT3 isoform KD mice were analyzed by performing BM biopsy. After 24 weeks, mouse blood was collected, and mice were sacrificed to harvest bone marrow and spleen. Collected cells were analyzed by flow cytometry to evaluate human engraftment (presence of human CD45 cells without mouse CD45).

Differentiation of hCD45 cells into myelo-lymphoid and erythroid-megakaryocytic lineages was evaluated through detection of myeloid (CD14 or CD66b), erythroid (CD71, GlycophorinA), megakaryocytic (CD41a) B-lymphoid (CD19) and T-lymphoid (CD3, CD4 and CD8) markers on human $CD45^{+}$ cells. Preservation of the HSPC compartment ($CD34^{+}CD38^{lo}CD90$) was also recorded. All transplanted mice were included in the analysis unless they died before the experimental endpoint (in total 3 mice died before 24 weeks, one each from control, MLLT3-S KD [sh3], and MLLT3-S KD [sh4] conditions). All studies and procedures involving mice were conducted in compliance with all the relevant ethical regulations and were approved by the UCLA Animal Research Committee (Protocol 2005-109).

ChIP-sequencing analysis

Sorted FL HSPC ($CD34+CD38^{lo}CD90+$) (50000-100000 per IP) were crosslinked in 1% formaldehyde for 10 minutes, quenched with glycine 0.0125M and snap-frozen as a dry pellet. The pellet was re-suspended in lysis buffer (50mM Tris 8.2, 10mM EDTA, 1% Triton X 100, 0.1% NaDeoxycholate, 0.5% Sarkosyl) and sonicated 12 min at a 5% intensity using Misonix cup-horn sonicator. Chromatin was incubated overnight with 2 µg of antibody (anti-MLLT3, Genetex, GTX102835; anti-V5, Abcam, ab15828; H3K4me3, Abcam ab8580; H3K9ac, CST #9649; H3K27ac, AM #39133; H3K4me1, AM #39297) preloaded onto 20 µl of Protein G Dynabeads (Thermo Fisher) and washed twice with each of the solutions (low salt wash, high salt wash, LiCl

wash and TE) as previously described[77]. CB HSPC sorted after 2 weeks culture with or without MLLT3-OE were used to assess MLLT3 binding in CB HSPCs.

Libraries were prepared with the Nugen Ovation Ultralow kit v2 following manufacturer instructions and sequenced using HiSeq-4000 (Illumina) to obtain single end 50 bp long reads. Demultiplexing of the reads based on the barcoding was performed using in house Unix shell script. Mapping to the human genome (hg19) was performed using bowtie2. Samtools v1.3.1 package was uses to create a bam file, remove duplicates, blacklisted region and ChrM regions, sort and index. Deeptols BamCoverage was used to generate Bigwig files, for track visualization on the UCSC genome browser. MACS2 v2.1.1[78] was used to call MLLT3 peaks using default parameters for broadPeak calling.

ATAC-sequencing analysis

FL-HSPC (CD34+CD38^{lo}/-CD90+GPI80+), their downstream progeny (CD34+CD38^{lo}/-CD90+GPI80- and CD34+CD38^{lo}/-CD90-GPI80-), were processed according to the protocol[79], with minor adjustments. Nuclei were purified by the addition of 250 ul of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630) to sorted cells, pelleted and resuspended in the transposition reaction mix (Nextera DNA Library Prep Kit, Illumina) and incubated at 37^{o}C for 30 minutes. Transposed DNA was column purified and used for library amplification with custom made adaptor primers[79] using NEBNext High-Fidelity 2x PCR Master Mix (New England Labs). The amplification was interrupted after 5 cycles and a SyBR green qPCR was performed with 1/10 of the sample to estimate for each sample the additional number of cycles to perform before saturation was achieved. Total amplification was between 10 and 15 cycles. Purified libraries were sequenced using HiSeq-2000 (Illumina) to obtain paired end 50 bp long reads. Read mapping to the genome (hg19) was using Bowtie2 or v2.2.9[80] with parameters -

-local -X 2000 -N 1 --no-mixed. Bamcoverage tool from Deeptools was used to create the coverage .bw files for visualization[81]. Samtools v1.3.1 was used to remove duplicates and reads aligned to chrM.

Single cell RNA-sequencing

The effects of MLLT3 isoform knockdown were evaluated at single cell level using 10x Genomics single cell RNA sequencing. For the generation of single-cell gel beads in emulsion, cells were loaded on a Chromium single cell instrument (10x Genomics) with an average estimated targeted cell recovery of ~6,000 cells. Single-cell suspensions of cells in 0.4% BSA-PBS were added to each channel on the 10X chip. Cells were partitioned with Gel Beads into emulsion in the Chromium instrument where cell lysis and barcoded reverse transcription of RNA occurred following amplification. Single-cell RNA-seq libraries were prepared by using the Chromium single cell 3' library and gel bead kit v3 (10X Genomics). Sequencing was performed on Illumina NovaSeq 6000.

Single-cell data analysis

After sequencing, fastq files were generated using cellranger mkfastq (version 2.1.1). The raw reads were mapped to human reference genome (refdata-cellranger-GRCh38-1.2.0) using cellranger count. Digital expression matrix was extracted from the "filtered_gene_bc_matrices" folder output by the cellranger count pipeline. To identify different cell types and find signature genes for each cell type, the R package Seurat (version 3.1.2) was used to analyze the digital expression matrix. Cells with less than 500 unique molecular identifiers (UMIs) or greater than 5% mitochondrial expression were removed from further analysis. The Seurat function NormalizeData was used to normalize the raw counts. Variable genes were identified using the

FindVariableGenes function. The ScaleData function was used to scale and center expression values in the dataset, the number of unique molecular identifiers (UMI) was regressed against each gene. Principal component analysis (PCA), t-distributed stochastic neighbor embedding (tSNE), and uniform manifold approximation and projection (UMAP) were used to reduce the dimensions of the data, and the first 2 dimensions were used in the plots. The FindClusters function was used to cluster the cells. Marker genes were found using the FindAllMarkers function for each cluster, which employs the Wilcoxon Rank Sum Test to determine the significance and the Benjamini-Hochberg Procedure to correct for multiple comparisons. Cell types were annotated based on the marker genes and their match to canonical markers. For differential expression analyses between cell selections, the FindMarkers function was used.

Gene ontology analysis (using Enrichr[82]) was applied to the positively or negatively correlated gene set to identify enriched pathways. The DotPlot function was used to illustrate the expression pattern of chosen genes in selected cells. Using dot plots, module scores or scorecards were created to highlight specific developmental processes regulated by MLLT3 isoforms.

Bulk RNA-sequencing analysis

Total RNA from 50000 sorted HSPC cultured with MLLT3 isoform OE was extracted using the RNeasy Mini kit (Qiagen) and library was constructed using KAPA RNA HyperPrep Kit with RiboErase (HMR). Libraries were sequenced using HiSeq-4000 (Illumina) to obtain paired end 50 bp long reads. Mapping to the human genome (hg19) was performed using TopHat v2.0.9 or v2.0.14[83] with the parameters --no-coverage-search -M -T -x 1. Coverage files were created with the Genomecov tool from Bedtools[84] with the parameters -bg -split -ibam. For abundance estimations (FPKMs) the aligned read files were further processed with HOMER coupled to edgeR on the hg19 annotation. Gene ontology of differentially expressed genes was calculated using

ENRICHR. Gene expression changes were considered significantly up- and down-regulated when adjusted p-value < 0.05 and -logFC > 0.322. Heatmaps were generated using Morpheus (Broad institute) using a color scale normalized by the minimum (MIN) and maximum (MAX) value for each gene.

Multiome

Sorted hematopoietic tissues from 5 week human embryo and 13 week second trimester fetal liver were processed using the Chromium Nuclei Isolation kit and processed using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit in accordance with manufacturer's protocol. Comparison of gene expression among Multiome data was carried out using ArchR and Seurat. Principal component analysis (PCA), t-distributed stochastic neighbor embedding (tSNE), and uniform manifold approximation and projection (UMAP) were used to reduce the dimensions of the data, and the first 2 dimensions were used in the plots. The FindClusters function was used to cluster the cells. Marker genes were found using the FindAllMarkers function for each cluster, which employs the Wilcoxon Rank Sum Test to determine the significance and the Benjamini-Hochberg Procedure to correct for multiple comparisons. Cell types were annotated based on the marker genes and their match to canonical markers. For differential expression analyses between clusters or cell selections, the FindMarkers function was used. scATAC-seq data was plotted using ArchR, Seurat and Loupe browser, while scRNA-seq data was plotted using the Loupe browser and Seurat.

Co-Immunoprecipitation

Cultured K562 cells were pelleted and washed twice with ice cold PBS, then lysed in RIPA Buffer (Invitrogen) with 1x Halt Protease Inhibitor Cocktail (Thermo Scientific). Protein was quantified

by Pierce BCA assay (Thermo Scientific), and 2ug of protein from each sample was combined with Dynabeads (Invitrogen) and either rabbit anti-human MLLT3 (only binds MLLT3-L) (ab154492, Abcam, 1:1000) or mouse anti-V5-tag cl. SV5-Pk1 (R960-25, Invitrogen, 1:500) antibody in with added 150 mM NaCl, 1.0% IGEPAL® CA-630, 0.5% sodium deoxycholate, 0.1% SDS, 1x Halt Protease Inhibitor Cocktail (Thermo Scientific), 1U Benzonase nuclease (EMD millipore) /µl Buffer, and 1mM MgCl2. and incubated overnight with gentle agitation at 4C. Beads were washed 2X with 50 mM Tris, pH 8.0 buffer containing 150mM NaCL and protein was eluted into wash buffer by heating to 90C for 10min.

Statistics and Reproducibility

Graphs were generated and statistical significance was calculated with GraphPad PRISM software. Statistical significance was assessed using 2-sided Student's *t*-test, unless otherwise stated. The null hypothesis of the medians/means being equal was rejected at $\alpha = 0.05$ and significant p-values are shown in each graph.

Data availability

Sequence data that support the findings of this study have been deposited in GEO with the accession codes GSE250349 and GSE23347. Data from published reference[63] are available in the GEO superseries GSE111484. There is no restriction in data availability.

Code availability

Custom codes, R objects and metadata of these R objects are available on GitHub page (https://github.com/mikkolalab).

Bibliography

1. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., Crooks, G.M., et al. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281-286. 10.1038/s41586-019-1790-2.
2. Org, T., Duan, D., Ferrari, R., Montel-Hagen, A., Van Handel, B., Kerenyi, M.A., Sasidharan, R., Rubbi, L., Fujiwara, Y., Pellegrini, M., et al. (2015). Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. EMBO J *34*, 759-777. 10.15252/embj.201490542.
3. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. 10.1186/gb-2008-9-9-r137.
4. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218. 10.1038/nmeth.2688.
5. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359. 10.1038/nmeth.1923.
6. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191. 10.1093/nar/gku365.

7. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97. 10.1093/nar/gkw377.

8. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111. 10.1093/bioinformatics/btp120.

9. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842. 10.1093/bioinformatics/btq033.

Chapter 5

Summary and Discussion

Conclusions:

How long-term repopulating HSCs manage to balance fate decisions to respond to the body's varying needs throughout life without exhausting the stem cell pool has remained a major unresolved question in HSC biology. The inability to regulate HSC fate in culture leads to premature differentiation or death, or expansion of immunophenotypic HSCs that lack the ability to function in vivo, which has compromised our ability to expand and manipulate human HSCs ex vivo for broader therapeutic applications. Similarly, our poor understanding of the processes that effect HSC specification and maturation has proved a barrier to developing processes for generation of mature, fully functional HSCs from PSCs that are sufficiently reliable and scalable for clinical use. Overcoming these limitations is critical for expanding the number of patients that can be treated for blood cancers or inherited disorders. Building on previous work identifying MLLT3 as a crucial factor for HSC expansion and maturation, this thesis sought to define how a novel, truncated MLLT3 isoform cooperates with full-length MLLT3 to regulate HSC maturation and fate decisions.

In this work, we demonstrate that:

1. Human HSCs express both the full length MLLT3 and a truncated MLLT3 isoform that lacks chromatin binding ability but is able to bind MLLT3 protein partners in the SEC and DotCom.
2. Long and short MLLT3 isoforms control human HSC behavior and stemness programs in opposite directions, but are both required for bone marrow repopulation ability
3. Opening of MLLT3-S TSS and enhancer during HSC emergence precede MLLT3-S induction during HSC maturation in the liver

4. MLLT3 isoforms differentially regulate IGFBP2, which supports HSPC proliferation in nascent HSCs and during *ex vivo* culture

Chapter 2

In this section, we show that HSCs also express a truncated isoform of MLLT3 that retains the Anc1 homology domain and the ability to bind MLLT3 protein partners, but is missing the YEATS domain of full length MLLT3 and is unable to bind chromatin. The differences in their protein structures allow MLLT3-S to act as a dominant negative isoform of MLLT3-L, and play an opposing but complementary role in HSC behavior and gene expression. Rescue of decreased MLLT3-L expression in culture leads to the expansion of transplantable HSCs[1], while rescue of MLLT3-S expression suppresses HSC expansion. Furthermore, the two isoforms have opposing effects on the expression of key HSC programs, such as protein synthesis, splicing, and metabolism, which are themselves known regulators of HSC fate decisions[2-12]. Despite their differing roles in HSC fate decisions, MLLT3-L and MLLT3-S are both required for full HSC function, as shown by dramatically decreased HSC engraftment following knockdown of either of the two isoforms. Expression of both MLLT3-S and MLLT3-L gives HSCs more precise and dynamic control over MLLT3-mediated gene expression, allowing them to rapidly respond to changing physiological requirements. This capacity to balance between different fate decisions is what allows HSCs to engraft upon transplantation and maintain long-term hematopoiesis.

Our findings show that MLLT3 isoforms can regulate diverse sets of genes depending on the context and developmental state. As such, MLLT3 should not be viewed simply as a transcription factor or component of a pathway with a narrow set of targets, but rather as a "stemness operating system" where the different isoforms enable fine-tuned control to HSC fate

decisions and stage-related programs, and at least to some degree in reversible fashion. MLLT3 is not the only gene for which a dominant negative truncated isoform has been shown to play a role in hematopoiesis. However, many of the other examples of functionally distinct isoforms, such as in IKAROS, CD44, and BPRF1, are regulated at the level of alternative splicing, not through an alternative TSS [13-16], which offers a different kind of molecular control to how distinct isoforms can be regulated. Other examples of key regulators in which two isoforms are expressed from separate TSSs and have differing roles during development, hematopoietic differentiation, and leukemic cell survival, is LEF1, a key target gene within the WNT signaling pathway, [17], and RUNX1, where developmentally regulated isoforms from two different promoters are also involved in trisomy 21 associated leukemia pathogenesis. However, despite their diverse roles in hematopoiesis, neither RUNX1 nor LEF1 are enriched specifically in HSCs or critically required to sustain HSC self-renewal and repopulation ability, whereas both the long and short MLLT3 isoforms are required for human HSC in vivo function. While HSCs that lose MLLT3-L rapidly lose HSC identity and begin to differentiate, the function of MLLT3-S is more nuanced in balancing HSCs stemness programs required for BM engraftment and HSC transition to homeostatic state required for life-long hematopoiesis. While many of the known MLLT3 interaction partners (AFF1, AFF4, DOT1L, MLLT10) are expressed more broadly in different cell types, the expression of MLLT3 isoforms is much more restricted, in particular that of MLLT3-S, further demonstrating their unique role in HSCs. Thus, MLLT3 isoforms serve also as important "quality control" markers when assessing preservation of stemness programs that are required for functionality in cultured HSCs. Alongside the above-mentioned examples, the identification of a dominant negative isoform of MLLT3 that is essential for regulating different flavors of HSC stemness, shows that it is important to adjust our analysis of stem cell gene expression to the level of isoforms, not just individual genes.

Chapter 3

In this section, we show that expression of the two MLLT3-S isoform is sequentially induced during human development. Although MLLT3-L is already expressed during the EHT in the hemogenic endothelium and emerging HSPCs in human 5 week AGM, MLLT3-S expression is induced after HSPC emergence and may be part of the HSC maturation process. However, MLLT3-S regulatory elements (TSS and HSC-specific enhancer) are accessible prior to its induction, through the EHT and HSC emergence in the AGM, and are inaccessible in venous endothelium and non-HSC hematopoietic cells. Thus, MLLT3-S TSS and HSC-specific enhancer accessibility can be viewed as an early indicator of specification of the definitive HSC lineage, even in the absence of MLLT3-S expression.

The expression of MLLT3-L in cell types that do not express MLLT3-S suggests that the two isoforms are regulated independently, at least to some extent. Further evidence for this can be found in our examination of MLLT3-L binding in cultured FL HSPCs following MLLT3-L OE. This data reveals that MLLT3-L binds to the TSS of MLLT3-S, partially sustaining its expression when it would normally decline. If MLLT3-S expression was solely dependent on the expression of MLLT3-L, one would expect both isoforms to be expressed at the same stages of HSC specification. However, the lag between the expression of MLLT3-L during EHT and the induction of MLLT3-S later in HSC maturation, despite the accessibility of MLLT3-S regulatory elements in hemogenic endothelium, strongly suggests that the developmental regulation of MLLT3-S expression is dependent upon additional, as of yet unidentified, factors. These may be revealed by future investigation and comparison of transcription factor binding motifs in the MLLT3-L TSS, the MLLT3-S TSS, and the MLLT3-S HSC-specific enhancer. Another potential point of comparison could be a distal MLLT3-S regulatory element that becomes accessible in erythroid

cells, which have an accessible MLLT3-S TSS and MLLT3-S expression, but in which the HSC-specific MLLT3-S enhancer is inaccessible. Understanding the factors and signaling governing MLLT3-S induction could be key for manipulating its expression during *ex vivo* HSC culture or *de novo* generation of PSC-derived HSCs without reliance on lentiviral overexpression vectors.

Our work also provides evidence that not only is MLLT3-S expression developmentally regulated, but MLLT3-S itself may also directly facilitate HSC maturation. Increased MLLT3-S expression in fetal liver HSCs closely corresponds to the downregulation of IGFBP2, an onco-fetal gene expressed in highly-proliferative fetal tissues and certain cancers. Decrease of IGFBP2 expression marks the broader transition from a highly-proliferative fetal state to a more mature, homeostatic one. IGFBP2 expression is re-activated in cultured HSCs, and we demonstrate that this renewed expression of IGFBP2 is required for HSPC *ex vivo* expansion, as well as for MLLT3-L-mediated expansion. Importantly, IGFBP2 is a direct MLLT3-L target whose expression is regulated in opposing directions by MLLT3 isoforms in culture; its expression is increased by MLLT3-L OE and decreased by MLLT3-S OE. This suggests that MLLT3-S upregulation may be a critical determinant of IGFBP2 downregulation in the fetal liver during development, and that MLLT3-S thus contributes to the establishment of homeostatic mechanisms in maturing HSCs. If so, future work may find that achieving stable expression of MLLT3-S in PSC-derived HSCs at the right moment in their differentiation may facilitate their maturation to the fetal liver stage, which has previously been unsuccessful.

Bibliography

1. Calvanese, V., Nguyen, A.T., Bolan, T.J., Vavilina, A., Su, T., Lee, L.K., Wang, Y., Lay, F.D., Magnusson, M., Crooks, G.M., et al. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature *576*, 281-286. 10.1038/s41586-019-1790-2.
2. Signer, R.A., Magee, J.A., Salic, A., and Morrison, S.J. (2014). Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature *509*, 49-54.
3. Buszczak, M., Signer, R.A., and Morrison, S.J. (2014). Cellular differences in protein synthesis regulate tissue homeostasis. Cell *159*, 242-251.
4. Rimmelé, P., Liang, R., Bigarella, C.L., Kocabas, F., Xie, J., Serasinghe, M.N., Chipuk, J., Sadek, H., Zhang, C.C., and Ghaffari, S. (2015). Mitochondrial metabolism in hematopoietic stem cells requires functional FOXO 3. EMBO reports *16*, 1164-1176.
5. Bonora, M., Ito, K., Morganti, C., Pinton, P., and Ito, K. (2018). Membrane-potential compensation reveals mitochondrial volume expansion during HSC commitment. Experimental hematology *68*, 30-37. e31.
6. Moon, J., Kim, H.R., and Shin, M.G. (2018). Rejuvenating aged hematopoietic stem cells through improvement of mitochondrial function. Annals of Laboratory Medicine *38*, 395.
7. Papa, L., Djedaini, M., and Hoffman, R. (2019). Mitochondrial role in stemness and differentiation of hematopoietic stem cells. Stem cells international *2019*.
8. Qiu, J., and Ghaffari, S. (2022). Mitochondrial deep dive into hematopoietic stem cell dormancy: not much glycolysis but plenty of sluggish lysosomes. Experimental hematology.

9. Takubo, K., Nagamatsu, G., Kobayashi, C.I., Nakamura-Ishizu, A., Kobayashi, H., Ikeda, E., Goda, N., Rahimi, Y., Johnson, R.S., and Soga, T. (2013). Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. Cell stem cell *12*, 49-61.

10. Goldstein, O., Meyer, K., Greenshpan, Y., Bujanover, N., Feigin, M., Ner-Gaon, H., Shay, T., and Gazit, R. (2017). Mapping whole-transcriptome splicing in mouse hematopoietic stem cells. Stem Cell Reports *8*, 163-176.

11. Tan, D.Q., Li, Y., Yang, C., Li, J., Tan, S.H., Chin, D.W.L., Nakamura-Ishizu, A., Yang, H., and Suda, T. (2019). PRMT5 modulates splicing for genome integrity and preserves proteostasis of hematopoietic stem cells. Cell reports *26*, 2316-2328. e2316.

12. Wang, F., Tan, P., Zhang, P., Ren, Y., Zhou, J., Li, Y., Hou, S., Li, S., Zhang, L., and Ma, Y. (2022). Single-cell architecture and functional requirement of alternative splicing during hematopoietic stem cell formation. Science advances *8*, eabg5369.

13. Igoe, N., Bayle, E.D., Fedorov, O., Tallant, C., Savitsky, P., Rogers, C., Owen, D.R., Deb, G., Somervaille, T.C., and Andrews, D.M. (2017). Design of a biased potent small molecule inhibitor of the bromodomain and PHD finger-containing (BRPF) proteins suitable for cellular and in vivo studies. Journal of medicinal chemistry *60*, 668-680.

14. Beer, P.A., Knapp, D.J., Kannan, N., Miller, P.H., Babovic, S., Bulaeva, E., Aghaeepour, N., Rabu, G., Rostamirad, S., and Shih, K. (2014). A dominant-negative isoform of IKAROS expands primitive normal human hematopoietic cells. Stem Cell Reports *3*, 841-857.

15. Davis, K.L. (2011). Ikaros: master of hematopoiesis, agent of leukemia. Therapeutic advances in hematology *2*, 359-368.

16. Ghaffari, S., Smadja-Joffe, F., Oostendorp, R., Lévesque, J.-P., Dougherty, G., Eaves, A., and Eaves, C. (1999). CD44 isoforms in normal and leukemic hematopoiesis. Experimental hematology *27*, 978-993.

17. Feder, K., Edmaier-Schröger, K., Rawat, V.P., Kirsten, N., Metzeler, K., Kraus, J.M., Döhner, K., Doehner, H., Kestler, H.A., and Feuring-Buske, M. (2020). Differences in expression and function of LEF1 isoforms in normal versus leukemic hematopoiesis. Leukemia *34*, 1027-1037.

Appendix 1

Article

MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment

https://doi.org/10.1038/s41586-019-1790-2

Received: 10 March 2018

Accepted: 9 October 2019

Published online: 27 November 2019

Vincenzo Calvanese[1,2*], Andrew T. Nguyen[1], Timothy J. Bolan[1], Anastasia Vavilina[1], Trent Su[3], Lydia K. Lee[4], Yanling Wang[1], Fides D. Lay[1], Mattias Magnusson[1,2], Gay M. Crooks[2,5,6], Siavash K. Kurdistani[2,3,6,7] & Hanna K. A. Mikkola[1,2,6,7*]

Limited knowledge of the mechanisms that govern the self-renewal of human haematopoietic stem cells (HSCs), and why this fails in culture, have impeded the expansion of HSCs for transplantation[1]. Here we identify MLLT3 (also known as AF9) as a crucial regulator of HSCs that is highly enriched in human fetal, neonatal and adult HSCs, but downregulated in culture. Depletion of MLLT3 prevented the maintenance of transplantable human haematopoietic stem or progenitor cells (HSPCs) in culture, whereas stabilizing MLLT3 expression in culture enabled more than 12-fold expansion of transplantable HSCs that provided balanced multilineage reconstitution in primary and secondary mouse recipients. Similar to endogenous MLLT3, overexpressed MLLT3 localized to active promoters in HSPCs, sustained levels of H3K79me2 and protected the HSC transcriptional program in culture. MLLT3 thus acts as HSC maintenance factor that links histone reader and modifying activities to modulate HSC gene expression, and may provide a promising approach to expand HSCs for transplantation.

HSCs can self-renew throughout their lifetime while replenishing all blood lineages, making HSC transplantation a life-saving treatment for many blood diseases. However, a lack of HLA-matched bone marrow donors and a low yield of HSCs in cord blood limit the number of patients that can be treated[1]. A better understanding of HSC self-renewal is required to expand human HSCs in culture or to generate them from pluripotent stem cells.

HSCs develop during embryogenesis from haemogenic endothelium in large arteries and expand in the fetal liver before colonizing the bone marrow[2]. Although many factors that drive the specification of haemogenic endothelium and HSCs have been identified, we know less about those that maintain HSC self-renewal. Here we identify MLLT3 as a crucial regulator of human HSC maintenance, and show that restoring MLLT3 levels in cultured human HSCs protects stemness and enables the ex vivo expansion of transplantable HSCs.

MLLT3 is enriched and required in human HSCs

To define the molecular machinery that governs human HSC self-renewal and determine why it fails in culture, we compared the transcriptomes of highly self-renewing HSPCs from human fetal liver to their immediate progeny[3] and to dysfunctional, cultured HSPCs, derived from fetal liver or embryonic stem cells[4,5]. From the 12 nuclear regulators correlating with self-renewal, MLLT3 was selected for further study (Fig. 1a, Extended Data Fig. 1a, b). MLLT3 is a component of the superelongation complex[6] and co-operates with DOT1L, which di/trimethylates H3K79 to promote transcription[7–9]. MLLT3 localizes to active transcription start sites (TSSs) through the YEATS domain, which recognizes active histone marks such as H3K9 acetylation and crotonylation[8,10]. A truncated MLLT3 that lacks the YEATS domain forms a leukaemic fusion protein with the N terminus of MLL1, which misdirects MLLT3-interacting complexes to induce aberrant gene transcription[11–14]. MLLT3 also regulates erythroid or megakaryocytic progenitors[15] and was identified as a definitive HSC hub gene during mouse development[16].

MLLT3 expression was enriched in undifferentiated human HSPCs in fetal liver, cord blood and bone marrow (Extended Data Fig. 1c). RNA-sequencing (RNA-seq) analysis of developmental haematopoietic tissues showed that *MLLT3* was upregulated in fetal liver[17] (Extended Data Fig. 1d), whereas genes specific to the development of haemogenic endothelium and HSCs such as *TAL1* (also known as *SCL*), *RUNX1*, *SOX17* and *HOXA* genes were already highly expressed in 5 week aorta–gonad–mesonephros, which suggests that MLLT3 is involved in HSC maturation and maintenance.

To determine whether human HSCs require MLLT3, two validated *MLLT3* short hairpin RNAs (shRNAs) (Extended Data Fig. 1e, f) were tested in an HSPC expansion culture system using the OP9M2 stromal stem-cell line[4]. Both of the shRNAs resulted in premature depletion of fetal liver HSPCs (FL-HSPCs) in vitro (Fig. 1b, c, Extended Data Fig. 1g–j). When FL-HSPCs transduced with MLLT3-knockdown (KD) or control vector were transplanted into immunodeficient NSG (NOD–SCID *Il2rg*-null) mice, only the control cells showed multilineage (myelo/lymphoid) human haematopoietic reconstitution (Fig. 1d, e, Extended Data Fig. 1k, Supplementary Table 1), which indicates an important regulatory function for MLLT3.

[1]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA. [2]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA, USA. [3]Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA, USA. [4]Department of Obstetrics and Gynecology, University of California Los Angeles, Los Angeles, CA, USA. [5]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. [6]Jonsson Comprehensive Cancer Center, University of California Los Angeles, Los Angeles, CA, USA. [7]Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA. *e-mail: vincalv@gmail.com; hmikkola@mcdb.ucla.edu

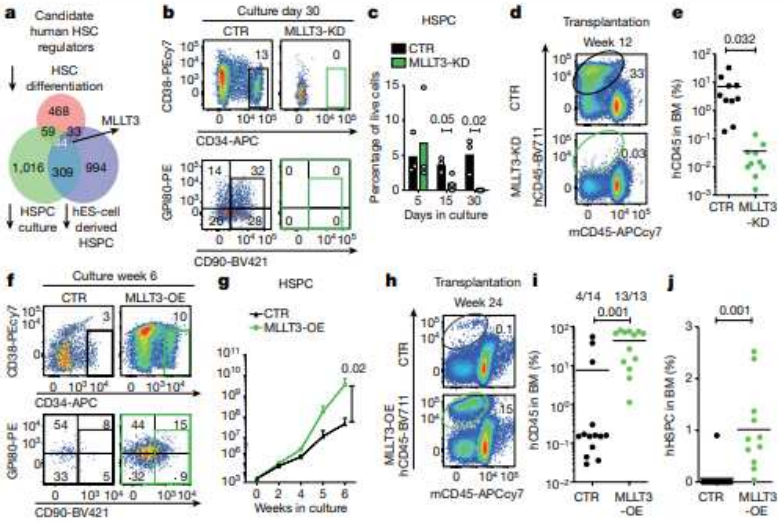

Fig. 1 | MLLT3 regulates human HSPC expansion. a, Venn diagram of microarray gene expression data, identifying genes enriched in self-renewing human FL-HSPCs. Number of genes downregulated after differentiation (pink) of fetal liver $CD34^+CD38^{-/lo}CD90^+GPI80^+$ HSCs to $CD34^+CD38^{-/lo}CD90^+GPI80^-$ progenitors[3]; number of genes downregulated in FL-HSPCs during 5-week culture on OP9M2 stroma (green)[4]; and number of genes suppressed in human embryonic stem (ES)-cell-derived HSPCs (purple)[5] are shown. **b**, FACS analysis 30 days after transduction of $CD34^+CD38^{-/lo}CD90^+$ HSPCs with *MLLT3* shRNA (MLLT3-KD) or empty vector control (CTR) (representative of three plots). **c**, Quantification of cells as in **b** after 5, 15 and 30 days in culture ($n = 3$). **d**, FACS analysis of bone marrow from NSG mice 12 weeks after transplantation of FL-HSPCs transduced with MLLT3-KD or empty vector control (representative of 10 mice). **e**, Quantification of human (h) $CD45^+$ cells in bone marrow (BM) from NSG mice treated as in **d** ($n = 10$ mice, two independent experiments). **f**, FACS analysis of $CD34^+CD38^{-/lo}CD90^+$ FL-HSPCs transduced with control or MLLT3-OE lentiviral vector (representative of six experiments). **g**, Expansion of HSPCs as in **f** ($n = 6$ independent experiments). **h**–**j**, FACS analysis showing human haematopoietic reconstitution (hCD45 expression) (**h**) and quantification of total hCD45 cells (**i**) or human HSPCs ($hCD45^+CD34^+CD38^{-/lo}$) (**j**) in bone marrow from NSG mice. In **i** and **j**, the number of mice showing multi-lineage reconstitution versus the number of total transplanted mice is shown ($n = 13$ or 14 mice, 4 independent experiments). APCcy7, PEcy7, BV421 and BV711 denote fluorochrome dyes. Data in **c** denote mean values; data in **e**, **i** and **j** denote mean and individual values; data in **g** are mean ± s.e.m. All *P* values determined by two-sided *t*-test.

Sustaining MLLT3 levels improves HSPC culture

As studies had shown expansion of multipotent human HSPCs on OP9M2 stroma without measurable expansion of transplantable HSCs[4], we asked whether maintaining MLLT3 expression in cultured HSPCs improves their function. Restoring physiological MLLT3 levels using an overexpression lentiviral vector (MLLT3-OE) significantly enhanced the expansion of $CD34^+CD38^{-/lo}CD90^+$ FL-HSPCs on OP9M2 stroma (78-fold greater than controls at 6 weeks) (Fig. 1f, g, Extended Data Fig. 2a, b). Targeting MLLT3-OE in the most undifferentiated $CD34^+CD38^{-/lo}CD90^+GPI80^+$ FL-HSPCs[3] recapitulated the expansion phenotype, whereas MLLT3-OE in $CD34^+CD38^{-/lo}CD90^+GPI80^-$ progenitors did not confer the HSC immunophenotype or expand them (Extended Data Fig. 2c–e). Withdrawing OP9M2 stroma depleted undifferentiated MLLT3-OE HSPCs, demonstrating a dependence on a HSC-supportive microenvironment (Extended Data Fig. 2f–h). MLLT3-OE also enhanced FL-HSPC expansion when cultured in clinically suitable conditions using serum-free expansion medium (SFEM) and the small molecules SR1[18] and UM171[19], or on OP9M2 stroma with both SR1 and UM171 (Extended Data Fig. 2i, j), showing the beneficial effects on HSPC expansion on all HSC-supportive conditions tested.

BrdU incorporation assays did not show enhanced proliferation of MLLT3-OE HSPCs in culture (Extended Data Fig. 3a, b). Staining of annexinV and 7-aminoactinomycin D (7AAD) showed greater viability of MLLT3-OE FL-HSPCs than empty-vector-transduced FL-HSPCs, but reduced cell viability compared with uncultured FL-HSPCs (Extended Data Fig. 3c). Differentiation of MLLT3-OE HSPCs after a 4-week expansion showed comparable monocytic, granulocytic, erythroid, megakaryocytic, T and B lymphoid differentiation potential to non-expanded FL-HSPCs (Extended Data Fig. 3d, e). Hence, sustaining MLLT3 expression in cultured HSPCs maintains their identity and viability without causing excessive proliferation or resistance to apoptosis, or blocking differentiation.

MLLT3 enhances transplantation of cultured HSPCs

To assess whether maintaining MLLT3 levels in cultured HSPCs improves the reconstitution potential in NSG mice, $CD34^+GFP^+$ fetal liver cells sorted at day 5 and transduced with MLLT3-OE or control vector were expanded in SFEM containing SR1 and UM171 and transplanted on day 15 (Extended Data Fig. 4a). Mice transplanted with MLLT3-OE cells showed more frequent long-term (24 week) human multilineage engraftment in the bone marrow, and higher levels of engraftment than mice transplanted with control HSPCs (Fig. 1h, i, Extended Data Fig. 4b, Supplementary Table 2a). Most mice transplanted with MLLT3-OE cells, but not with control cells, also contained human HSPCs in the bone marrow and generated myeloid and lymphoid cells in the peripheral blood and spleen (Fig. 1j, Extended Data Fig. 4c–e). Increased haematopoietic reconstitution with MLLT3-OE HSPCs was not explained by altered proliferation of MLLT3-OE HSPCs or differentiated cells in recipient mice or in culture before transplantation (Extended Data Fig. 4f–j).

To ascertain the role of extended culture in MLLT3-mediated enhanced in vivo reconstitution, the engraftment levels were compared between $CD34^+GFP^+$ cells sorted 5 days after transduction and their progeny sorted at 15 days. Culture with MLLT3-OE, but not control vector, significantly increased total human haematopoietic reconstitution and HSPC reconstitution (Extended Data Fig. 4k–m). When bone marrow from primary mice was transplanted into secondary mouse

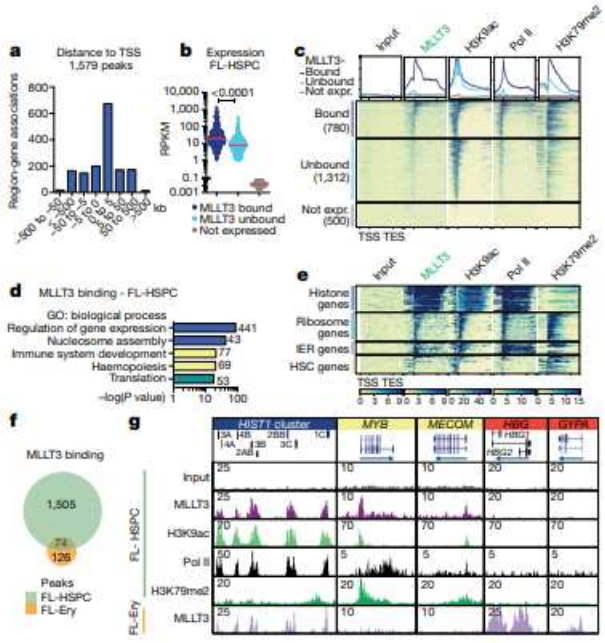

Fig. 2 | MLLT3 binds to the TSSs of active genes in a cell-type-specific manner. **a**, Distribution of MLLT3 peaks in FL-HSPCs. **b**, Genes expressed in FL-HSPCs (RPKM > 1, $n = 3$) were divided into MLLT3-bound genes (780, peak < 5 kb from the TSS) and MLLT3-unbound genes (1,312, no detectable peak), and compared to 500 randomly selected non-expressed genes (RPKM < 0.1). Red lines denote median values. *P* value determined by two-sided *t*-test. **c**, Average profile and heat map for MLLT3, H3K9ac, RNA Pol II and H3K79me2 ChIP–seq in FL-HSPCs. Metagene plot ± 2 kb is shown. **d**, GO analysis of MLLT3-bound genes in FL-HSPCs with gene numbers for each category. **e**, Heat maps for selected MLLT3-bound gene groups. IER, immediate early response. **f**, Venn diagram comparing MLLT3 peaks in FL-HSPCs and fetal liver erythroblasts (FL-Ery). **g**, UCSC genome browser tracks showing ChIP–seq of MLLT3 and epigenetic marks in representative MLLT3-bound genes in FL-HSPCs and FL-erythroblasts. In **a**–**g**, $n = 3$ MLLT3 in FL-HSPCs, $n = 2$ others.

recipients, only MLLT3-OE cells showed multilineage reconstitution after 18 weeks (Extended Data Fig. 4n, o). These data indicate that maintaining MLLT3 expression in FL-HSPCs during culture enhances human multilineage haematopoietic reconstitution in primary and secondary recipients.

MLLT3 binds to TSSs of active genes in human HSPCs

To understand how MLLT3 regulates human HSC stemness, we assessed the MLLT3 chromatin-binding pattern in FL-HSPCs. Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq) showed MLLT3 binding at 1,579 sites, with strongest enrichment around TSSs and within 5 kb downstream (Fig. 2a). MLLT3 peaks associated with 889 genes, 96.4% of which were expressed in FL-HSPCs, with reads per kilobase of transcript per million mapped reads (RPKM) values greater than one (Supplementary Table 3). The *k*-means clustering analysis showed co-localization of MLLT3 peaks with marks of active TSSs (assay for transposase-accessible chromatin using sequencing (ATAC-seq) peaks, H3K4me3, H3K9ac, H3K9cr, H3K27ac and RNA polymerase (Pol) II) (Extended Data Fig. 5a). The gene body histone mark H3K79me2 partially overlapped with MLLT3, whereas H3K36me3 was not enriched. There was minimal overlap with the enhancer mark H3K4me1 or the repressive marks H3K27me3 and H3K9me3. Thus, MLLT3 predominantly localizes to active promoters in HSPCs.

Comparing MLLT3-bound genes to other expressed genes showed that, although both contained active epigenetic marks at TSSs, MLLT3-bound genes featured higher median expression, higher H3K79me2 enrichment and higher RNA Pol II occupancy (Fig. 2b, c, Extended Data Fig. 5b). Gene Ontology (GO) analysis of MLLT3-bound genes in FL-HSPCs revealed enrichment of biological processes involved in regulation of gene expression and nucleosome assembly (for example, histone genes), immune system development and haemopoiesis (for example, HSC transcription factors RUNX1, MYB, MECOM and HOXA9) and translation (for example, ribosomal proteins) (Fig. 2d, Supplementary Table 3a–c). Analysis of epigenetic marks in distinct MLLT3-bound gene groups revealed differential enrichment for H3K79me2 and Pol II: the histone genes and immediate early response genes (such as *JUN* and *FOS*) showed high Pol II occupancy but low H3K79me2, whereas HSC genes and ribosomal protein genes showed high enrichment for H3K79me2 (Fig. 2e, Extended Data Fig. 5c). These data suggest that MLLT3 may regulate distinct target genes in HSPCs by influencing H3K79me2 and/or Pol II activity.

Analysis of MLLT3 binding in erythroblasts from fetal liver (Fig. 2f, Supplementary Table 3d) revealed cell-type specificity: 200 MLLT3 peaks were identified around TSSs, partially overlapping with FL-HSPC peaks (Extended Data Fig. 6a–c). Common GO categories included nucleosome assembly, whereas erythroid-specific categories included oxygen transport and haem metabolic processes (Extended Data Fig. 6d). Genome browser tracks demonstrated the correlation of MLLT3 binding with epigenetic marks of active TSS in each cell type (Fig. 2g, Extended Data Fig. 6a), and enrichment of H3K79me2 in MLLT3-bound HSC transcription factor genes.

MLLT3 protects HSC gene expression in cultured HSPCs

We next asked how sustaining MLLT3 expression in cultured FL-HSPCs modulates their transcriptional program. ChIP–seq data showed similar distribution of MLLT3-bound peaks and genes in uncultured HSPCs and MLLT3-OE-vector transduced HSPCs after a 4-week culture (Extended Data Fig. 7a–c). RNA-seq showed that modest differences in MLLT3 expression (2.77-fold MLLT3-OE versus control vector in HSPCs after 4-week culture) resulted in significant differential expression of 541 upregulated and 717 downregulated genes (Fig. 3a, Supplementary Table 4). Genes that regulate translation and glycolysis, and several HSC transcription factors (*MECOM* (also known as *EVI1*), *HLF*, *MYB* and *GFI1*) and HSC surface proteins (*c-KIT* (also known as *KIT*), *CXCR4*, *ROBO4*, *EMCN* and *PROM1* (also known as *CD133*)) were significantly upregulated in MLLT3-OE HSPCs (Fig. 3b). Programs related to immune response and apoptosis were suppressed (Fig. 3b). Although MLLT3 binding was equally distributed between upregulated and downregulated genes, MLLT3 binding was enriched in specific gene categories such as MLLT3-OE-upregulated HSC transcription factor genes, and downregulated nucleosome assembly (for example, histone) genes. Immune response and apoptosis genes suppressed in MLLT3-OE HSPCs showed minimal binding (Fig. 3b–d). Comparison of cultured and uncultured HSPCs suggested that MLLT3-OE may help to diminish culture-associated drift in gene expression, either directly (HSC factors) or indirectly (immune response genes)[4] (Fig. 3d, e, Extended Data Fig. 7d). Functional assessment of 2 of the 12 candidate HSC factors identified in Extended Data Fig. 1a—*MECOM* and *HLF*—both bound and upregulated by MLLT3, validated them as important MLLT3 downstream effectors that sustain HSC stemness. Knockdown of either factor resulted in premature HSPC exhaustion and diminished the effects of MLLT3-OE on HSPC expansion (Extended Data Fig. 7e–h).

Given the strong association between MLLT3 binding and H3K79me2 deposition in genes that encode HSC regulators, we asked whether MLLT3-OE protects the expression of HSC genes in cultured HSPCs

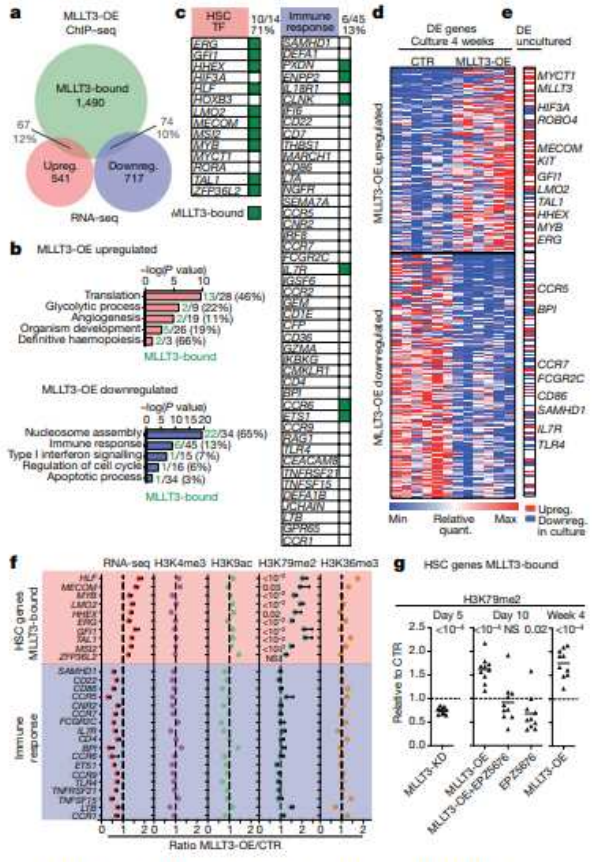

Fig. 3 | MLLT3 protects HSC stemness program through DOT1L and H3K79me2. a, Venn diagram showing the overlap between MLLT3-bound ($n = 6$ MLLT3 ChIPs combined; Extended Data Fig. 7b) and MLLT3-OE up- or downregulated genes. Total number and percentage of genes are shown. **b**, GO analysis of genes up- or downregulated in MLLT3-OE FL-HSPCs at 4 weeks. Numbers and percentage of MLLT3-bound and total genes are shown. **c**, Examples of gene groups up- or downregulated by MLLT3-OE with MLLT3-bound genes marked in green. Numbers and percentage of MLLT3-bound and total genes are shown. **d**, Heat map showing differentially expressed genes in HSPCs transduced with MLLT3-OE or control vector after 4-week expansion ($n = 6$ independent experiments). $P < 0.05$, Benjamini–Hochberg adjusted t-test. **e**, Differential expression (DE) of MLLT3-regulated genes between uncultured HSPCs ($n = 3$) and 4-week expanded FL-HSPCs ($n = 6$). Selected MLLT3-regulated genes that are similarly regulated in uncultured HSPCs are highlighted. MLLT3-bound genes are in bold. **f**, Quantification of RNA-seq and ChIP–seq signals in MLLT3-bound, upregulated HSC transcription factor genes, and unbound, downregulated immune response genes. RNA-seq represents fold change between expanded HSPCs transduced with MLLT3-OE or control vector. ChIP–seq shows ratio of MLLT3-OE and control signal normalized to non-MLLT3-bound housekeeping gene (*VCL*) ($n = 6$ RNA-seq; $n = 5$ H3K79me2 ChIP, $n = 2$ other ChIPs). Data are mean ± s.e.m. P values determined by two-sided t-test. **g**, Ratio of H3K79me2 signal in MLLT3-bound upregulated HSC genes between MLLT3-KD and control vector after 5-day culture (data are mean and individual values, $n = 10$ genes, two experiments) (left), and between MLLT3-OE, MLLT3-OE plus EPZ5676 (DOT1L inhibitor), or CTR plus EPZ5676 and control vector after 10-day culture (middle). H3K79me2 signal is normalized to S2 cell chromatin spike-in. MLLT3-OE versus control from 4-week culture is also shown (right). P values in **g** determined by one-sample two-sided t-test. NS, not significant.

through DOT1L and H3K79me2. A significant MLLT3-dependent increase in H3K79me2 was observed in MLLT3-bound haematopoietic regulators but not in immune response genes that were indirectly downregulated by MLLT3-OE (Fig. 3f, Extended Data Fig. 8a). Other active marks tested (H3K9ac, H3K4me3 and H3K36me3) showed little change with MLLT3-OE. Dependence of H3K79me2 on MLLT3 levels was verified in MLLT3-KD HSPCs (Fig. 3g). Culture with the DOT1L inhibitor EPZ5676[20] reduced MLLT3-OE-associated increase in H3K79me2 in HSC genes (Fig. 3g). EPZ5676 decreased H3K79me2 in both MLLT3-bound and non-bound genes, whereas MLLT3-KD and MLLT3-OE only affected H3K79me2 in MLLT3-bound genes (Extended Data Fig. 8b). These data suggest that, although MLLT3 is not required for DOT1L activity and H3K79me2 deposition per se, MLLT3 cooperates with DOT1L in human HSCs to enhance H3K79me2 deposition in HSC regulatory genes and to maintain their activity during culture expansion.

MLLT3 enables ex vivo expansion of cord blood HSCs

The low number of HSCs in cord blood limits their use for transplantation, despite their better availability and more permissive HLA-matching than bone marrow. We therefore asked whether MLLT3 can be used to control the self-renewal of cord-blood HSCs (CB-HSCs). MLLT3-KD in cord-blood HSPCs (CB-HSPCs) severely impaired HSPC maintenance in co-culture experiments with OP9M2 stroma cells (Extended Data Fig. 9a, b). Conversely, MLLT3-OE improved CB-HSPC expansion both in co-culture with OP9M2 cells and in serum-free conditions with UM171 and SR1 (Fig. 4a, b, Extended Data Fig. 9c, d).

A limiting-dilution transplantation assay (Fig. 4a) was used to quantify the expansion of transplantable HSCs (repopulating units) during CB-HSC culture. Several doses of MLLT3-OE-transduced and control-transduced cells were transplanted either at day 5 after transduction or day 15 after culture and compared to uncultured CB-HSPCs. When high doses were transplanted, multilineage human haematopoietic reconstitution was observed in mice transplanted with uncultured cells or with HSPCs transduced with control vector or MLLT3-OE vector (Fig. 4c). Quantification of lineage differentiation in engrafted mice confirmed a comparable differentiation ability of MLLT3-OE and uncultured HSPCs (Fig. 4d), whereas HSPCs expanded by control vector showed myeloid bias. Moreover, mice engrafted with MLLT3-OE cells showed both increased HSPCs and total reconstitution compared with equal cell doses of control cells (Fig. 4e, f). Analysis of mice transplanted at limiting dilutions verified robust expansion of MLLT3-OE HSCs at 15 days of culture; 12.5-fold increase in repopulating units compared with uncultured HSPCs and a 6.8-fold increase compared with MLLT3-OE HSPCs transplanted at day 5 (Fig. 4f–h). Therefore, increased reconstitution ability was not caused merely by the expansion of MLLT3-OE cells in recipient mice. Although cells transduced with control vector showed some expansion (2.4-fold compared with uncultured HSPCs) at 15 days, the expansion of repopulating units with MLLT3-OE was 5.2-fold higher. These data corroborate the importance of maintaining MLLT3 levels during culture, both to achieve greater HSC expansion and preserve differentiation potential.

Blood and spleen showed solid human haematopoietic reconstitution by MLLT3-OE-expanded cells without evidence of lineage bias, whereas control-vector-expanded cells showed inconsistent, lower-level engraftment (Extended Data Fig. 10a–d). Secondary transplantation of bone marrow from primary mice resulted in human haematopoietic reconstitution with uncultured and MLLT3-OE cells, but not with control-vector-expanded cells (Extended Data Fig. 10e, f). Neither primary nor secondary mice transplanted with MLLT3-OE cells suffered increased mortality or expansion of immature populations (Extended Data Fig. 10g, h), unlike mice transplanted with haematopoietic cells expressing the oncogenic fusion protein MLL1–MLLT3 that rapidly developed leukaemia[21,22]. Thus, sustaining MLLT3 expression in CB-HSPCs during culture enables a more than 12-fold expansion of

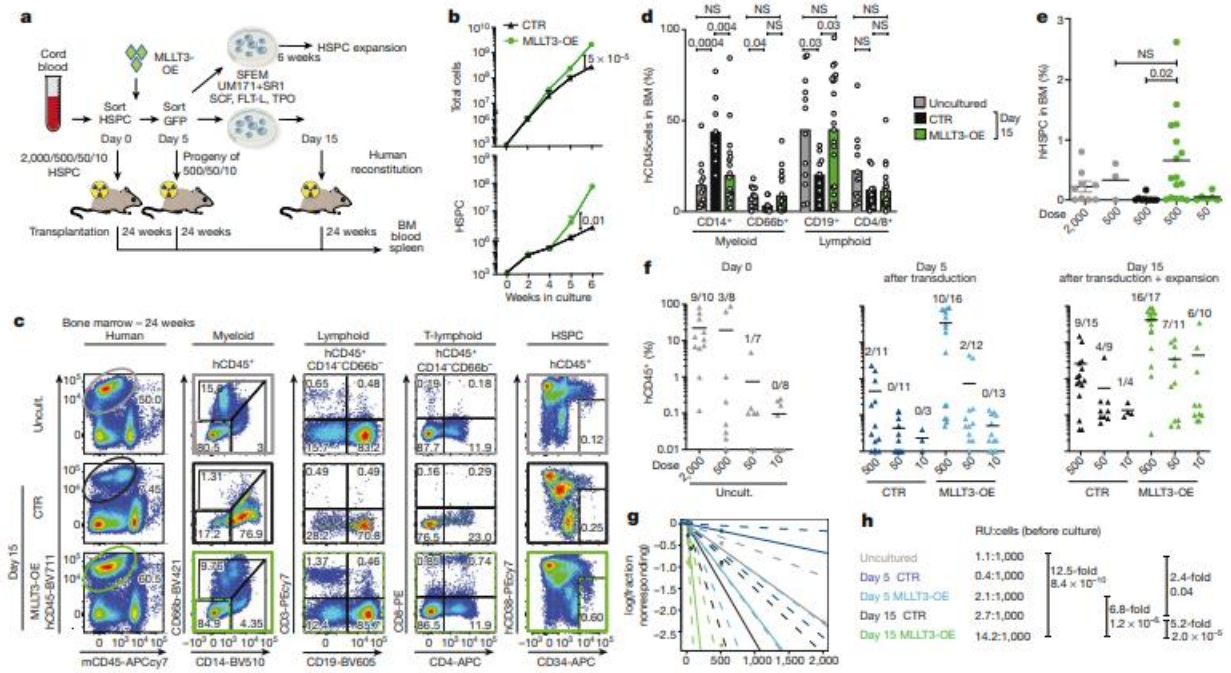

Fig. 4 | MLLT3-OE expands transplantable CB-HSCs in culture. a, Limiting dilution analysis for assessing MLLT3-OE effects on CB-HSPC expansion and in vivo reconstitution. **b**, Quantification of total cell and HSPC expansion in culture ($n = 3$). Data are mean values. *P* values determined by two-sided *t*-test. **c**, Representative FACS plots showing human haematopoietic reconstitution in mice transplanted with uncultured CB-HSPCs or CB-HSPCs transduced with control or MLLT3-OE vectors and transplanted at day 15 after culture and analysed 24 weeks after transplantation. FACS plots show total human haematopoietic reconstitution (CD45), multilineage differentiation (myeloid, B-lymphoid and T-lymphoid cells) and HSPCs in hCD45$^+$ cells. **d**, Quantification of differentiated populations as a percentage of hCD45$^+$ cells in the bone marrow of engrafted mice. **e**, Quantification of HSPCs as a percentage of live bone marrow cells in engrafted mice. **f**, Quantification of human haematopoietic reconstitution in NSG mice at 24 weeks. Mice were transplanted in limited dilution doses with uncultured cells, or with transduced cells cultured for 5 or 15 days. The percentage of hCD45$^+$ cells in bone marrow and the number of multilineage engrafted versus total transplanted mice is indicated. Data in **d**–**f** are mean and individual values from three independent experiments. *P* values determined by two-sided *t*-test. In **f**, *n* denotes engrafted mice/total transplanted mice for each condition. **g**, **h**, Calculation of reconstituting units from limiting dilution assay. Data in **h** are expressed as reconstituting units (RU) per 1,000 uncultured CB-HSPCs transplanted directly or transduced with MLLT3-OE or control vector and cultured for 5 or 15 days before transplantation. Expansion factors and *P* values between reconstituting units in uncultured HSPCs and MLLT3-OE or control vector-expanded cells were calculated using ELDA software[29] and are shown.

transplantable HSCs that maintain balanced multilineage haematopoiesis, demonstrating self-renewal of human HSCs in culture.

Discussion

Our discovery that restoring MLLT3 levels improves the expansion of engraftable, multipotent human HSCs when cultured with SR1 and UM171—small molecules that show promise for clinical HSC expansion[19,23,24]—suggests that combining several strategies to support HSCs may be optimal for clinical translation. Notably, recent studies showed a marked increase in mouse HSC activity after replacing albumin in culture medium[25]. As harnessing HSC self-renewal in culture becomes a reality, the potential for transformation and clonal haematopoiesis needs to be monitored carefully, independent of the expansion protocol used. Because every cell division in itself poses a risk for mutations, the goal should not necessarily be maximal HSC expansion, but safe expansion that preserves HSC integrity. Although no adverse effects were observed with the MLLT3-OE vector, the use of transient, non-integrating methods to maintain MLLT3 levels would be more appropriate for clinical use.

Our work identified MLLT3 as an 'HSC maintenance factor' that preserves—rather than confers—HSC stemness, by reinforcing programs established by other factors. Localization of MLLT3 to active TSSs of HSC genes in human HSCs in a cell-type-specific manner, using its YEATS domain[8,10], provides a molecular basis for the accurate binding of MLLT3-OE in cultured HSPCs. This explains why full-length MLLT3 does not induce a differentiation block or ectopic activation of self-renewal, unlike the oncogenic MLL1–MLLT3 fusion protein, which rapidly converts haematopoietic progenitors to self-renewing leukaemia stem cells[12,21,22]. Our work identifies MLLT3 as a central regulator of transcription factors that individually control HSC function[26–28] and suggests that MLLT3–DOT1L-dependent regulation of H3K79me2 helps to maintain an active chromatin state in HSC regulatory genes during culture. With the ability to protect HSC stemness program as HSCs divide, without enhancing proliferation or imposing self-renewal program on progenitors, MLLT3 may help to improve HSC ex vivo expansion for clinical use.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1790-2.

1. Ballen, K. K., Gluckman, E. & Broxmeyer, H. E. Umbilical cord blood transplantation: the first 25 years and beyond. *Blood* **122**, 491–498 (2013).

2. Gritz, E. & Hirschi, K. K. Specification and function of hemogenic endothelium during embryogenesis. *Cell. Mol. Life Sci.* **73**, 1547–1567 (2016).
3. Prashad, S. L. et al. GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. *Cell Stem Cell* **16**, 80–87 (2014).
4. Magnusson, M. et al. Expansion on stromal cells preserves the undifferentiated state of human hematopoietic stem cells despite compromised reconstitution ability. *PLoS ONE* **8**, e53912 (2013).
5. Dou, D. R. et al. Medial *HOXA* genes demarcate haematopoietic stem cell fate during human development. *Nat. Cell Biol.* **18**, 595–606 (2016).
6. He, N. et al. Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. *Proc. Natl Acad. Sci. USA* **108**, E636–E645 (2011).
7. Steger, D. J. et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol* **28**, 2825–2839 (2008).
8. Li, Y. et al. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **159**, 558–571 (2014).
9. Bitoun, E., Oliver, P. L. & Davies, K. E. The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum. Mol. Genet.* **16**, 92–106 (2007).
10. Li, Y. et al. Molecular coupling of histone crotonylation and active transcription by AF9 YEATS domain. *Mol. Cell* **62**, 181–193 (2016).
11. Schoch, C. et al. AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogenetically analyzed AML cases. *Blood* **102**, 2395–2402 (2003).
12. Krivtsov, A. V. et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).
13. Bernt, K. M. et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66–78 (2011).
14. Wang, X., Chen, C. W. & Armstrong, S. A. The role of DOT1L in the maintenance of leukemia gene expression. *Curr. Opin. Genet. Dev.* **36**, 68–72 (2016).
15. Pina, C., May, G., Soneji, S., Hong, D. & Enver, T. MLLT3 regulates early human erythroid and megakaryocytic cell fate. *Cell Stem Cell* **2**, 264–273 (2008).
16. McKinney-Freeman, S. et al. The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell* **11**, 701–714 (2012).
17. Ng, E. S. et al. Differentiation of human embryonic stem cells to HOXA$^+$ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nat. Biotechnol.* **34**, 1168–1179 (2016).
18. Boitano, A. E. et al. Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. *Science* **329**, 1345–1348 (2010).
19. Fares, I. et al. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. *Science* **345**, 1509–1512 (2014).
20. Daigle, S. R. et al. Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* **122**, 1017–1025 (2013).
21. Horton, S. J. et al. MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. *Leukemia* **27**, 1116–1126 (2013).
22. Sontakke, P. et al. Modeling BCR-ABL and MLL-AF9 leukemia in a human bone marrow-like scaffold-based xenograft model. *Leukemia* **30**, 2064–2073 (2016).
23. Wagner, J. E. Jr et al. Phase I/II trial of stemregenin-1 expanded umbilical cord blood hematopoietic stem cells supports testing as a stand-alone graft. *Cell Stem Cell* **18**, 144–155 (2016).
24. Ngom, M. et al. UM171 enhances lentiviral gene transfer and recovery of primitive human hematopoietic cells. *Mol. Ther. Methods Clin. Dev.* **10**, 156–164 (2018).
25. Wilkinson, A. C. et al. Long-term *ex vivo* haematopoietic-stem-cell expansion allows nonconditioned transplantation. *Nature* **571**, 117–121 (2019).
26. Kataoka, K. et al. Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. *J. Exp. Med.* **208**, 2403–2416 (2011).
27. Rentas, S. et al. Musashi-2 attenuates AHR signalling to expand human haematopoietic stem cells. *Nature* **532**, 508–511 (2016).
28. Komorowska, K. et al. Hepatic leukemia factor maintains quiescence of hematopoietic stem cells and protects the stem cell pool during regeneration. *Cell Reports* **21**, 3514–3523 (2017).
29. 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).

Methods

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Human haematopoietic tissue collection and processing

Human HSPCs from fetal liver and cord blood were used to assess the function of MLLT3 in fetal and neonatal human HSC. Second trimester (14–18 weeks) fetal livers were de-identified, discarded material obtained from elective terminations of pregnancy after informed consent. Specimen age is denoted as developmental age, two weeks less than gestational age, and was determined by ultrasound or estimated by the date of the last menstrual period. Cord blood units were obtained from full-term pregnancies following informed consent and de-identified upon collection. Because these tissues are discarded material with no personal identifiers, this research does not constitute human subjects research.

Fetal liver samples were mechanically dissociated using scalpels and syringes and strained through a 70-µm mesh. Single-cell suspensions from fetal liver and cord blood were enriched in mononuclear cells by layering on Lymphoprep (Stem Cells Technologies) and centrifugation following manufacturer instructions. $CD34^+$ cells were magnetically isolated from mononuclear cell fraction using anti-CD34 microbeads (Miltenyi Biotech).

OP9M2 stroma co-culture for HSPC expansion

Human HSC co-culture on OP9M2 (subclone of OP9[4], derived in our laboratory and validated by gene expression analysis) stroma was used to test the maintenance and expansion of immunophenotypic HSPCs. OP9M2 cells were irradiated (20 Gy) and pre-plated (50,000 cells cm^{-2}) 24 h before the start of co-culture in OP9 medium, which includes α-MEM (Invitrogen), 20% fetal bovine serum (FBS; Omega) and 1× penicillin/streptomycin/glutamine (P/S/G). Human HSPCs were plated on stromal layer in OP9 medium supplemented with human HSC cytokines SCF (25 ng ml^{-1}, Peprotech or Invitrogen), FLT3-L (25 ng ml^{-1}, Peprotech) and TPO (25 ng ml^{-1}, Peprotech) (HSC medium). Cells were co-cultured at 37 °C and 5% CO_2 and re-plated or analysed/sorted by flow cytometry for HSPC markers ($CD34^+CD38^{-/lo}CD90^+GPI80^+$) every 7–14 days. Half of the HSC-medium was replaced every 2–3 days. Where indicated, 500 nM StemRegenerin1 (SR1) and 35 nM UM171 (SCT) were added to the cultures to improve human HSPC expansion.

SFEM culture for HSPC expansion

To test HSPC function in clinically relevant serum-free, stroma-free culture conditions, HSPCs were plated in StemSpan SFEM II (SCT) supplemented with human SCF (100 ng ml^{-1}), human FLT3-L (100 ng ml^{-1}), human TPO (50 ng ml^{-1}), human low-density lipoprotein (10 µg ml^{-1}, SCT), P/S/G, 500 nM SR1 and 35 nM UM171. Cells were cultured at 37 °C and 5% CO_2 and re-plated or analysed/sorted by flow cytometry for human HSPC markers every 7–14 days. Half of the HSC medium was replaced every other day. In some experiments, DOT1l inhibitor EPZ5676 (Cayman) was added at a concentration of 500 nM.

Flow cytometry and cell sorting

FACS analysis was performed using single-cell suspensions prepared from fetal liver and cord blood as described above, or obtained from NSG mice transplanted with human cells.

For identification of human HSPCs, cells were stained with mouse anti-human monoclonal antibodies against human CD34-APC cl. 581 (555824; BD, 1:20) or -BV605 (343529; Biolegend, 1:20), CD90-FITC cl. 5E10 (555595; BD, 1:100) or -APC (555595; BD, 1:100), CD38-PE-Cy7 or -BUV496cl. HIT2 (560677 and 564657; BD, 1:100), GPI-80-PE cl. 3H9 (D087-5; MBL, used at 1:50). Cells were assayed on a BD-LSRII flow cytometer and data were analysed with FlowJo software (Tree Star). Cell sorting was performed using a BD FACS Aria II.

To monitor human haematopoietic engraftment in vivo in NSG mice, bone marrow cells were stained with rat anti-mouse-CD45-APC-H7 cl. 30-F11 (557659; BD, 1:100) and mouse anti-human monoclonal antibodies against human-CD45-BV711 or -BV785 cl.HI30 (304050, 304048; Biolegend, 1:100). Haematopoietic differentiation was assessed in vitro and in vivo using mouse anti-human monoclonal antibodies against human CD19-PE or BV605 cl. 1D3 or HIB19 (12-0193, 12-0199; eBiosciences, 1:50), CD3-PE-Cy7 cl. SK7 (557851, BD) (eBiosciences, 1:50), CD4-APC (MHCD0405; Invitrogen, 1:50), CD8-PE cl. HIT8A (555635; BD, 1:50), CD13-APC cl. WM15 (557454; BD, 1:50), CD66b-BV421 cl. G10F5 (555724; BD, 1:50), CD14-V500 cl. WM53 (561816; BD, 1:100), CD235a-PE or -APC (Glycophorin A, HIR2, BD 1:100), CD71-AF647 (DF1513, Santa Cruz 1:100), CD41a-APC cl. HIP8 (579777; BD, 1:20), CD42b-PEcy7 cl. HIP1 (303916; Biolegend, 1:100), CD10-AF700 cl. HI10a (563509; BD, 1:100), CD33-PE cl. WM53 (555450; BD, 1:50), CD24-BV711 cl. ML5 (563401; BD, 1:100), CD20-BV650 cl. 2H7 (563780; BD, 1:100).

Dead cells were excluded with 7AAD (BD Biosciences, used at 1:50). For assessment of cell-cycle stages or apoptosis, 7AAD was combined with BrdU-PE (556029; BD, 1:100) or annexinV-PE (556422; BD, 1:50), respectively (see below).

RNA isolation, cDNA synthesis and qRT–PCR

RNA isolation was performed using the RNeasy Mini kit (Qiagen) with additional DNase step using manufacturer's protocol. cDNAs were prepared using High-Capacity cDNA Reverse Transcription Kit (ThermoFisher), and qPCR for *GAPDH*, *MLLT3*, *MECOM* and *HLF* was performed with the LightCycler 480 SYBR Green I Master Mix (Roche) or TaqMan Gene Expression Master Mix on the Lightcycler 480 (Roche). Primers are presented in Supplementary Table 5.

Production of lentiviral shRNA and overexpression vectors

shRNA experiments were performed with pLKO lentiviral vectors from the TRC library containing puromycin resistance gene. shRNA TRCN0000005793 (93, Sigma) was selected for in vitro and in vivo experiments after testing for knockdown efficacy compared to other shRNAs against *MLLT3* from the RNAi Consortium (TRCN0000005790 to TRCN0000005794) series. TRCN0000014790 and TRCN0000002528 were used to knockdown *HLF* and *MECOM*, respectively. shRNAs were tested on the K562 and Hep2G cell lines, obtained by ATCC and mycoplasma-free.

Human *MLLT3* was cloned from human FL-HSPC full-length cDNA into the constitutive FUGW (Addgene plasmid 14883, from D. Baltimore) lentiviral vector, in which its expression is controlled by the *UBC* promoter. *MLLT3* cDNA with a C-terminal V5-tag was inserted downstream and in frame with the GFP sequence with the synthetic addition a P2A sequence between the 2 ORFs, through two rounds of PCR (Supplementary Table 4) using PfuUltra II Ultra Fusion HS (Agilent).

For lentiviral vector production, 20 million 293T cells were co-transfected with deltaR8.2 packaging plasmid, VSVG-envelope plasmid, the lentiviral vector plasmid of choice and Turbo DNAfectin 3000 (Lamda), following manufacturer's instructions, in Opti-MEM (Life Technologies) and incubated for 5–6 h at 37 °C. After incubation for 48 h in Ultraculture medium (Lonza), supernatant was filtered and concentrated by ultracentrifugation and pelleted viruses were resuspended in SFEM and stored at −80 °C.

Lentiviral transduction

Sorted fetal liver and cord blood HSPCs were prestimulated 24 h in StemSpan SFEM II supplemented with SCF, FLT3-L, TPO and antibiotic-antimycotic (LT). Wells were pre-coated with 40 µg ml^{-1} RetroNectin (Takara) and seeded with pre-stimulated HSPC in 300 µl SFEM. Lentivirus was added twice, first after 24 h from cell seeding and again after additional 8 h. After 48 h from cell seeding, transduced cells

were washed and seeded in the indicated conditions for each assay. For shRNA-lentiviral vectors, puromycin (1.0 μg ml^{-1}) treatment was used for selection of transduced HSPCs and maintained throughout culture. Cells infected with overexpression vectors were selected by cell sorting for GFP expression.

Haematopoietic differentiation assays

FL-HSPCs sorted directly or after 4 weeks of expansion on OP9M2 were plated on differentiation assays. For myeloid assay, FL-HSPCs were plated in StemSpan SFEM II (SCT) supplemented with G-CSF (20 ng ml^{-1}, Peprotech), GM-CSF (20 ng ml^{-1}, Peprotech), TPO (25 ng ml^{-1}) and SCF (25 ng ml^{-1}). For the erythroid assay, HSPCs were plated in StemSpan SFEM II supplemented with IL-3 (20 ng ml^{-1}, Peprotech), EPO (2.5 IU, Thermo Fisher Scientific), SCF (25 ng ml^{-1}), L-glutamine and antibiotic-antimycotic. For the megakaryocyte differentiation assay, HSPCs were plated in IMDM supplemented with recombumin (0.4% Albumedix), 2-mercaptoethanol (100 μM, Gibco), SCF (100 ng ml^{-1}), TPO (50 ng ml^{-1}), IL-3 (10 ng ml^{-1}), IL-6 (7.5 ng ml^{-1}, Peprotech), IL-9 (13.5 ng ml^{-1}, Humanzyme), human low-density lipoprotein (4 μg ml^{-1}, SCT), ITS-X (Gibco), glutamax and antibiotic-antimycotic. For the T cell assay, HSPCs were plated on non-irradiated OP9-DLL1 stroma (25,000 cells cm^{-2}) in OP9 medium supplemented with SCF (25 ng ml^{-1}), FLT3-L (10 ng ml^{-1}) and IL-7 (20 ng ml^{-1}, Peprotech). For the B cell assay, HSPCs were seeded on irradiated OP9M2 in MEM-alpha 5% FBS supplemented with SCF (25 ng ml^{-1}), FLT3-L (10 ng ml^{-1}) and IL-7 (20 ng ml^{-1}). In each assay, cells were cultured in 24-well plates with 1 ml of the indicated medium and half medium changes were applied every 2–3 days. Cells were analysed by flow cytometry after 2 weeks.

Cell cycle and apoptosis assays

For the BrdU incorporation analysis in vitro, control- or MLLT3-OE-transduced FL-HSPCs were cultured for 4 weeks on OP9M2 with 500 nM SR1 and 35 nM UM171 and pulse-labelled with 10 μM BrdU for 35 min in culture. Cells were sorted for the indicated surface phenotypes and processed according to the PE-BrdU flow kit (BD) instructions and analysed by flow cytometry. Apoptosis from FL-HSPCs cultured in the same conditions was assessed by flow cytometry using annexin-V-PE (BD) and 7AAD incorporation, following manufacturer instructions.

Transplantation assays in NSG mice

In vivo reconstitution ability of cultured HSPCs was assessed in immunodeficient mice. Female NSG (Jackson Laboratories) mice, 8–12 weeks old, were sub-lethally irradiated (2.75 Gy) and retro-orbitally injected with cells derived from fetal liver or cord blood in a volume of 100 μl of RPMI.

For *MLLT3* shRNA knockdown experiments, 30,000 fetal liver CD34$^+$ cells were infected with *MLLT3* shRNA or control lentivirus and cultured on OP9M2 for 9 days under puromycin selection before transplantation. At 12 weeks, MLLT3-KD mice were euthanized to obtain bone marrow.

For MLLT3-OE experiments with FL-HSPCs, mice were transplanted with the progeny of 1,000 sorted GPI80$^+$ FL-HSPCs (CD34$^+$CD38$^{-/lo}$ CD90$^+$GPI80$^+$). Cells were transduced with the MLLT3-OE or control vector and re-sorted after 5 days for GFP$^+$CD34$^+$ cells. These cells were either injected retro-orbitally immediately after sorting (day 5) or expanded on SFEM for an additional 10 days (day 15) before injection. After 24 weeks, mice blood was collected, and mice were euthanized to obtain bone marrow and spleen. Collected cells were analysed by FACS to evaluate human engraftment (human CD45 and absence of mouse CD45). Reconstituted mice were defined by the presence of GFP$^+$ (MLLT3-OE and control vector) human CD45 cells that differentiated to myeloid and lymphoid lineages 24 weeks after transplantation. Differentiation into myelo-lymphoid lineages was evaluated by the detection of myeloid (CD14 or CD66b), B-lymphoid (CD19) and T-lymphoid (CD3, CD4 and CD8) markers on human CD45$^+$ cells. Of note, T-lymphoid engraftment was evaluated only at 24 weeks from transplantation, owing to the absence of reliable T-lymphoid differentiation at week 12 (MLLT3-KD, Extended Data Fig. 1k). Preservation of the HSPC compartment (CD34$^+$CD38$^{-/lo}$) was also recorded.

For BrdU incorporation analysis in vivo, NSG mice were transplanted with high doses of human FL-HSPCs (8,000 GPI80$^+$ HSPCs or their progeny at culture day 15 after transduction with control or MLLT3-OE vector). Then 14 weeks after transplantation, 2 mg of BrdU was injected intraperitoneally. Mice were euthanized 100 min after and bone marrow was collected. Cells were sorted for the indicated surface phenotypes and processed according to the PE-BrdU flow kit (BD) instructions and assessed by flow cytometry.

For secondary transplantations, viably frozen bone marrow from primary NSG mice transplanted with 15-day cultured HSPCs or uncultured HSPCs was used. Bone marrow from mice from the same experimental group was pooled and a dose equivalent of 1/2 femur of the primary mouse total bone marrow was injected retro-orbitally in sub-lethally irradiated secondary recipients. Human engraftment was assessed in bone marrow after 18 weeks from transplantation.

For the limiting dilution assay, CB-HSPCs were injected after FACS sorting, after transduction and sorting transduced cells (GFP$^+$CD34$^+$, day 5), or after an additional 10 days of expansion in culture (day 15). The number of cells reported (2,000, 500, 50, 10) is the number of sorted HSPCs (uncultured) that was transplanted directly, or the progeny of which was transplanted after culture (days 5 and 15). The criteria for multilineage engraftment was having hCD45 positive (GFP$^+$ with MLLT3 and CTR vector) cells that displayed at least one myeloid (CD14 CD66b or both) and the B-lymphoid (CD19) marker. HSC frequency was assessed using ELDA software[29].

A panel of markers covering the classical AML blast (CD34 and CD33) and B-cell precursor markers (CD34, CD10, CD20, CD24, CD38) altered in MLL–MLLT3 fusion gene-driven leukaemia in immunodeficient mice were tested in mice transplanted with day-15-expanded MLLT3-OE CB-HSPCs and uncultured CB-HSPCs.

All transplanted mice were included in the analysis unless they died before the experimental endpoint (in total from combined fetal liver and cord blood experiments, 14 out of 115 mice transplanted with MLLT3-OE, 17 out of 98 mice transplanted with control vector HSPCs, and 3 out of 36 mice with uncultured HSPCs died before 24 weeks).

All studies and procedures involving mice were conducted in compliance with all the relevant ethical regulations and were approved by the UCLA Animal Research Committee (protocol 2005-109).

ChIP–seq analysis

Sorted FL-HSPCs or erythroblasts (CD34$^-$CD235$^+$CD71$^+$) (50,000–10,0000 per immunoprecipitation) were crosslinked in 1% formaldehyde for 10 min, quenched with glycine 0.0125 M and snap-frozen as a dry pellet. The pellet was re-suspended in lysis buffer (50 mM Tris, pH 8.2, 10 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.5% sarkosyl) and sonicated 12 min at a 5% intensity using Misonix cup-horn sonicator. Chromatin was incubated overnight with 2 μg of antibody (anti-MLLT3, Genetex, GTX102835; anti-V5, Abcam, ab15828; H3K4me3, Abcam ab8580; H3K9ac, CST 9649; H3K9cr, PTM PTM-516; H3K36me3, Abcam ab9050; H3K79me2, Abcam, ab3594; RNA-polII Rbp1 (8WG16) Biolegend 664911, H3K27ac, AM 39133; H3K4me1, AM 39297; H3K27me3, AM 39155; H3K9m3, AM 39161) preloaded onto 20 μl of Protein G Dynabeads (Thermo Fisher) and washed twice with each of the solutions (low-salt wash, high-salt wash, LiCl wash and TE buffer) as previously described[30].

ChIP experiments for H3K79me2 were also performed from short-term cultures with FL-HSPCs transduced with either MLLT3-KD (no. 93) or MLLT3-OE vector and respective controls. Knockdown was collected at day 5, which included 3 days of puromycin selection. MLLT3-OE were sorted for GFP and collected at day 10, which included 4 days of EPZ5676 treatment at 500 nM, or DMSO as a control. Sonicated chromatin from these chips was spiked-in with S2 chromatin (Activ Motif)

following manufacturer instructions. Libraries were prepared with the Nugen Ovation Ultralow kit v2 following manufacturer instructions and sequenced using HiSeq-4000 (Illumina) to obtain single-end 50-bp long reads. Demultiplexing of the reads based on the barcoding was performed using in house Unix shell script. Mapping to the human genome (hg19) was performed using bowtie2. Samtools v.1.3.1 package was uses to create a .bam file, remove duplicates, blacklisted region and ChrM regions, sort and index. Bedtools multicov was used to quantify histone mark signals. GO analysis on genes bound by MLLT3 was calculated using Homer annotatePeaks.pl[31] using -go function and statistic is reported as *P* value, calculated with standard parameters by the algorithm. Each H3K79me2, H3K4me3, H3K36me3 and H3K9ac ChIP replicate on MLLT3-OE (Fig. 3f) is internally normalized using the non-MLLT3-bound, housekeeping gene *VCL* (Extended Data Fig. 8b). For H3K79me2 ChIP samples spiked in with S2 chromatin, sequences were aligned to both the human and *Drosophila melanogaster* genome (dm6) and the quantified reads, were normalized by total reads aligned to dm6 for each sample (Fig. 3g). Coverage files, average profiles heat maps were created with Deeptools packages[32]. MACS2 v.2.1.1[33] was used to call MLLT3 peaks using default parameters for broadPeak calling.

ATAC-seq analysis

FL-HSPCs or erythroblasts ($CD34^-CD235^+CD71^+$) (50,000 sorted cells) were processed according to the protocol[34], with minor adjustments. Nuclei were purified by the addition of 250 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% IGEPAL CA-630) to sorted cells, pelleted and resuspended in the transposition reaction mix (Nextera DNA Library Prep Kit, Illumina) and incubated at 37 °C for 30 min. Transposed DNA was column purified and used for library amplification with custom made adaptor primers[34] using NEBNext High-Fidelity 2× PCR Master Mix (New England Labs). The amplification was interrupted after 5 cycles and a SyBR green qPCR was performed with 1/10 of the sample to estimate for each sample the additional number of cycles to perform before saturation was achieved. Total amplification was between 10 and 15 cycles. Purified libraries were sequenced using HiSeq-2000 (Illumina) to obtain paired-end 50-bp long reads. Read-mapping to the genome (hg19) was done using Bowtie2 or v.2.2.9[35] with parameters –local -X 2000 -N1–no-mixed. The Bamcoverage tool from Deeptools was used to create the coverage .bw files for visualization[32]. Samtools v.1.3.1 was used to remove duplicates and reads aligned to chrM.

RNA-seq analysis

Total RNA from 50,000 sorted HSPCs was extracted using the RNeasy Mini kit (Qiagen) and library was constructed using KAPA RNA HyperPrep Kit with RiboErase (HMR). Libraries were sequenced using HiSeq-4000 (Illumina) to obtain paired-end 50-bp long reads. Mapping to the human genome (hg19) was performed using TopHat v.2.0.9 or v.2.0.14[36] with the parameters–no-coverage-search -M -T -x 1. Coverage files were created with the Genomecov tool from Bedtools[37] with the parameters -bg -split -ibam. For abundance estimations (FPKMs) the aligned read files were further processed with HOMER coupled to edgeR on the hg19 annotation. GO was calculated using DAVID[38] and statistic is reported as *P* value, calculated with standard parameters by the algorithm. Gene expression changes were considered significantly up- and downregulated when adjusted $P < 0.05$ (–log-transformed fold change > 0.322). The selection of differentially expressed genes is based on adjusted *P* value (significant across six replicates) rather than fold change because, as MLLT3 is required to maintain HSC identity and viability, the changes observed in the HSPC compartment reflect the beginning of MLLT3-dependent processes, not the end point. Heat maps were generated using Morpheus (Broad Institute) using a colour scale normalized by the minimum and maximum value for each gene.

Reporting summary

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

Data availability

Sequence data that support the findings of this study have been deposited in Gene Expression Omnibus (GEO) with the accession code GSE111484. Data from published reference are available in GEO GSE81080[17]. All other data are either available within the paper or from the corresponding author upon reasonable request. Custom codes for data analysis are also available upon request. There is no restriction in data availability.

30. Org, T. et al. Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence. *EMBO J.* **34**, 759–777 (2015).
31. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
32. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187-91 (2014).
33. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
34. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
36. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
37. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
38. Dennis, G. Jr et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).

Acknowledgements We thank BSCRC flow cytometry (F. Codrea, J. Scholes and J. Calimlim) and sequencing (S. Feng) cores, TCGB (NIH P30CA016042) and CFAR cores (NIH P30AI028697-21) at UCLA. We thank J. Zhao, T. Montoya and D. Dou for assistance with experiments; V. Rezek, D. Johnson and O. Witte for help with NSG mice and D. Kohn, Z. Romero and R. Hollis for help with lentiviral vectors. This work was supported by NIH RO1 DK100959 and RO1 DK121557; Broad Stem Cell Center at UCLA, Rose Hills Foundation, Jonsson Cancer Center Foundation, and UCLA David Geffen School of Medicine Regenerative Medicine Theme Award for H.K.A.M.; LLS Special Fellow Award and BSCRC post-doctoral fellow award for V.C.; CIRM GC1R-06673B for C.M.G. and H.K.A.M; Ruth L. Kirschstein National Research Service Award T32HL069766 for A.V., LLS Fellow Award and AACR post-doctoral award for F.L.; Beckman Scholars Program for A.T.N. and T.J.B; and NIH grant CA178415 to S.K.K.

Author contributions V.C. and H.K.A.M. designed experiments and interpreted data. V.C. performed and/or supervised all experiments and data analysis, including the generation of functional data documenting the effects of MLLT3 on HSC expansion in culture, the characterization of the expanded HSCs in transplantation and the molecular characterization of MLLT3-mediated gene regulation in HSCs. A.T.N., T.J.B. and A.V. contributed to MLLT3 knockdown and overexpression experiments. Other members of the laboratory also independently replicated these key functional data. A.V., T.J.B., L.K.L. and Y.W. contributed to the transplantation experiments in NSG mice. V.C., F.D.L. and T.S. performed bioinformatics analysis. M.M., G.M.C. and S.K.K. provided support with experimental design and discussed data analysis and results. V.C. and H.K.A.M. wrote the manuscript, which all authors edited and approved.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1790-2.

Correspondence and requests for materials should be addressed to V.C. or H.K.A.M.

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

Appendix 2

`

Article

Mapping human haematopoietic stem cells from haemogenic endothelium to birth

https://doi.org/10.1038/s41586-022-04571-x

Received: 31 December 2020

Accepted: 22 February 2022

Published online: 13 April 2022

Check for updates

Vincenzo Calvanese[1,2,3,16] ✉, Sandra Capellera-Garcia[1,2,16], Feiyang Ma[1,2,4,16], Iman Fares[1,2], Simone Liebscher[5], Elizabeth S. Ng[6], Sophia Ekstrand[1,2], Júlia Aguadé-Gorgorió[1,2], Anastasia Vavilina[1,2], Diane Lefaudeux[7,8], Brian Nadel[1], Jacky Y. Li[6], Yanling Wang[1], Lydia K. Lee[9], Reza Ardehali[2,10], M. Luisa Iruela-Arispe[11], Matteo Pellegrini[1,2], Ed G. Stanley[6,12,13], Andrew G. Elefanty[6,12,13], Katja Schenke-Layland[5,10,14,15] & Hanna K. A. Mikkola[1,2] ✉

The ontogeny of human haematopoietic stem cells (HSCs) is poorly defined owing to the inability to identify HSCs as they emerge and mature at different haematopoietic sites[1]. Here we created a single-cell transcriptome map of human haematopoietic tissues from the first trimester to birth and found that the HSC signature *RUNX1*$^+$*HOXA9*$^+$*MLLT3*$^+$*MECOM*$^+$*HLF*$^+$*SPINK2*$^+$ distinguishes HSCs from progenitors throughout gestation. In addition to the aorta–gonad–mesonephros region, nascent HSCs populated the placenta and yolk sac before colonizing the liver at 6 weeks. A comparison of HSCs at different maturation stages revealed the establishment of HSC transcription factor machinery after the emergence of HSCs, whereas their surface phenotype evolved throughout development. The HSC transition to the liver marked a molecular shift evidenced by suppression of surface antigens reflecting nascent HSC identity, and acquisition of the HSC maturity markers CD133 (encoded by *PROM1*) and HLA-DR. HSC origin was tracked to *ALDH1A1*$^+$*KCNK17*$^+$ haemogenic endothelial cells, which arose from an *IL33*$^+$*ALDH1A1*$^+$ arterial endothelial subset termed pre-haemogenic endothelial cells. Using spatial transcriptomics and immunofluorescence, we visualized this process in ventrally located intra-aortic haematopoietic clusters. The in vivo map of human HSC ontogeny validated the generation of aorta–gonad–mesonephros-like definitive haematopoietic stem and progenitor cells from human pluripotent stem cells, and serves as a guide to improve their maturation to functional HSCs.

Developmental haematopoiesis consists of multiple waves of blood cell production that culminate in the generation of self-renewing HSCs. The steps to generate human HSCs rather than HSC-independent progenitors are poorly defined, compromising the efforts to differentiate HSCs from human pluripotent stem (PS) cells for transplantation and disease modelling. Human HSCs emerge in the aorta–gonad–mesonephros (AGM) region between Carnegie stages 13 and 17 (CS13–17; 4–6 weeks) through intra-aortic haematopoietic clusters (IAHCs)[2]. Although IAHCs contain numerous haemogenic cells, transplantable HSCs are rare (one per AGM)[3], suggesting functional immaturity[4]. Nascent HSCs colonize the liver, where they mature and acquire robust bone marrow engraftment ability[4]. In contrast to mouse HSCs, there are no methods to recapitulate human HSC maturation in culture or pinpoint their maturation stage[5,6].

HSC emergence is preceded by several HSC-independent progenitor waves[7,8]. The yolk sac first generates primitive erythrocytes and macrophages, followed by erythro-myeloid progenitors (EMPs) that initiate fetal liver haematopoiesis, at least in mice[8,9]. In humans, yolk-sac-derived myeloid progenitors seed the liver by CS12 (ref. [10]). Recent lineage-tracing studies have challenged the dogmas by showing that progenitors generate long-lived progeny, including microglia and tissue-resident macrophages[11,12]. Moreover, lymphoid potential, which was previously considered to be an exclusive trait of HSCs, has been reported in HSC-independent progenitors[7,10,13,14]. Innate-like B1 B cells were linked to developmentally restricted haematopoietic stem and progenitor cells (HSPCs) that can acquire self-renewal ability after transplantation[15]. Single-cell technologies uncovered an early (CS10–11) HSC-independent intraembryonic human haematopoietic wave[16], and

[1]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, USA. [2]Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of California Los Angeles, Los Angeles, CA, USA. [3]Laboratory for Molecular Cell Biology, University College London, London, UK. [4]Chongqing International Institute for Immunology, Chongqing, China. [5]Institute of Biomedical Engineering, Department for Medical Technologies and Regenerative Medicine, Eberhard Karls University, Tübingen, Germany. [6]Murdoch Children's Research Institute, The Royal Children's Hospital, Parkville, Victoria, Australia. [7]Signaling Systems Laboratory, Department of Microbiology Immunology and Molecular Genetics (MIMG), University of California Los Angeles, Los Angeles, CA, USA. [8]Institute for Quantitative and Computational Biosciences (QCB), University of California Los Angeles, Los Angeles, CA, USA. [9]Department of Obstetrics and Gynecology, University of California Los Angeles, Los Angeles, CA, USA. [10]Department of Medicine/Cardiology, CVRL, University of California Los Angeles, Los Angeles, CA, USA. [11]Cell and Developmental Biology, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA. [12]Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, Victoria, Australia. [13]Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria, Australia. [14]Cluster of Excellence iFIT (EXC 2180) 'Image-Guided and Functionally Instructed Tumor Therapies', Eberhard Karls University Tübingen, Tübingen, Germany. [15]NMI Natural and Medical Sciences Institute, University Tübingen, Reutlingen, Germany. [16]These authors contributed equally: Vincenzo Calvanese, Sandra Capellera-Garcia, Feiyang Ma. ✉e-mail: v.calvanese@ucl.ac.uk; hmikkola@mcdb.ucla.edu

associated fetal haematopoiesis with unexpected differentiation trajectories and locations[13,17]. Although developmental haematopoiesis involves multiple anatomical sites, including the yolk sac, placenta, major arteries, and even the head and the heart[18–26], their contribution to HSC development is unknown, especially in humans[2]. Our understanding of human HSC development has remained incomplete without the ability to distinguish between HSCs and progenitors, and to identify their endothelial precursor[14,16,27–31]. The endothelial and haematopoietic markers) *CDH5* (which encodes VE-cadherin), *RUNX1* and *PTPRC* (which encodes CD45) associated with IAHCs[3] are not specific to HSCs.

By creating a single-cell transcriptome map of human HSC ontogeny, we documented HSC development from arterial haemogenic endothelium (HE) to transplantable HSCs using cell-type-specific molecular signatures, stage-specific gene expression scorecards and protein landmarks, and visualized HSC emergence in IAHCs using spatial transcriptomics. This map validated the generation of AGM-stage HSPCs from PS cells, and uncovered remaining bottlenecks to HSC generation in culture.

Molecular identity of nascent human HSCs

To identify nascent human HSCs, we performed single-cell RNA-sequencing (scRNA-seq) analysis of CD34$^+$-and/or CD31$^+$-enriched haemato-vascular cells from the AGM region of three CS14–15 (4.5–5 weeks) embryos (Supplementary Table 1). Grouping the clusters by cell-type-specific gene expression uncovered a cluster with the expected features of human HSCs: co-expression of transcription factors that regulate HSC specification (medial HOXA genes[32,33]) and self-renewal (*MLLT3*[34], *MECOM*[35], *HLF*[36]) and HSC surface markers *CD34*, *THY1* (which encodes CD90) and *ACE*[37] (Fig. 1a and Extended Data Fig. 1a–d). The expression of the highly HSC-enriched gene *HLF*[16,34,38,39] was most specific to the HSC cluster.

Reclustering AGM haematopoietic cells also distinguished an *HLF*$^+$ HSC cluster that was enriched for HSC regulatory genes, whereas non-HSCs expressed genes reflecting immune system development (Extended Data Fig. 1f–k and Supplementary Table 2). To identify markers of nascent HSCs, the HSC cluster was also compared with all of the other AGM clusters (Supplementary Table 2). The top 30 HSC-enriched genes identified from these two comparisons included known HSC regulators (*GFI1*, *MYB*) and potential new human AGM HSC markers (*SPINK2*, *RAB27B*) (Fig. 1b, c). *SPINK2* (ref. [40]), which encodes a serine protease inhibitor observed recently in HSPC datasets[16,38], was the most significantly enriched gene in HSC cluster. AGM HSCs also possessed a distinct endothelial signature (*PROCR*, *EMCN*). The genes enriched in the AGM HSC cluster were combined to generate a 'nascent HSC scorecard' (Fig. 1b).

HSC and progenitor distribution at CS14

To determine whether extraembryonic or other intraembryonic tissues contain HSCs, we performed scRNA-seq analysis of multiple tissues from a CS14 conceptus and searched for clusters co-expressing *RUNX1*$^+$ *HOXA9*$^+$ *MLLT3*$^+$ *MECOM*$^+$ *HLF*$^+$ *SPINK2*$^+$, herein termed the HSC signature. Selecting *HLF*$^+$ cells in such clusters identified candidate HSCs in CS14 AGM, placenta and yolk sac, some in umbilical and vitelline vessels, but minimal in the liver, head and heart, although they contained other haematopoietic cells (Extended Data Fig. 2a, b). The nascent HSC scorecard confirmed the resemblance of extraembryonic *HLF*$^+$ cells to AGM HSCs (Extended Data Fig. 2c).

To identify HSCs using a complementary method that minimizes dropouts in scRNA-seq, we used Markov affinity-based graph imputation of cells (MAGIC) imputation and calculated HSC module scores in CS14 tissues using HSC signature genes (Extended Data Fig. 2d–f). This analysis detected HSCs in the same locations (Fig. 1d). The nascent HSC scorecard documented comparable molecular properties between HSCs identified by HSC module score versus *HLF* expression (Extended Data Fig. 2g). These data imply that HSCs populate extraembryonic tissues before circulating systemically or colonizing the liver (Fig. 1e).

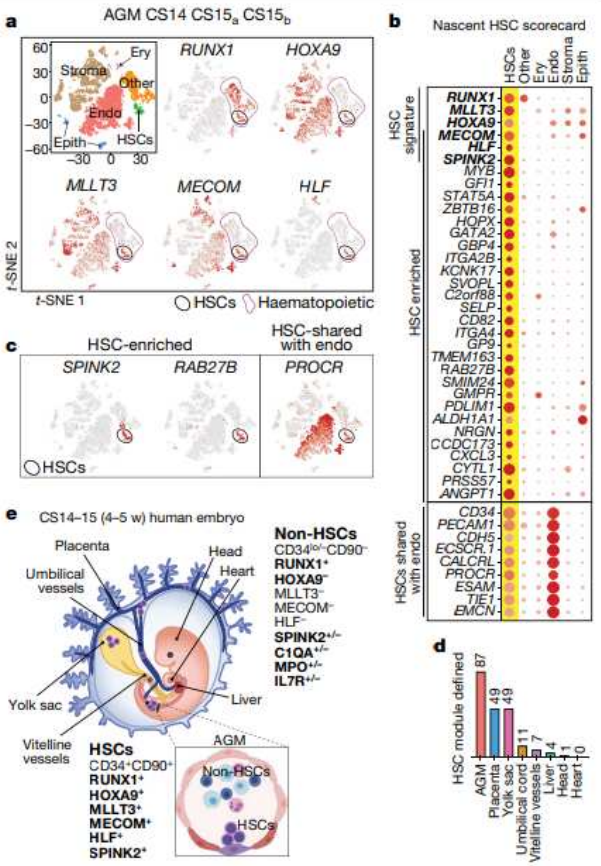

Fig. 1 | HSC molecular signature identifies nascent human HSCs. **a**, scRNA-seq analysis of CD34$^+$ and/or CD31$^+$ enriched cells from the AGM region of human embryos. $n = 3$ biologically independent samples: week 4.5 (CS14), week 5 (CS15$_a$) and week 5 (CS15$_b$). Cells are plotted using a *t*-distributed stochastic neighbour embedding (*t*-SNE) analysis and are categorized according to cell type. Clusters with haematopoietic cells (RUNX1$^+$CD45$^+$) and HSCs (cluster 12) are indicated by purple and black circles, respectively. Feature plots identify HSCs by the co-expression of HSC transcriptional regulators. Epith, epithelium; Other, other haematopoietic cells. **b**, The nascent HSC scorecard dot plot, which includes genes that are significantly enriched in the HSC cluster compared with all of the other clusters (in **a**), and all other haematopoietic cells (other) (Extended Data Fig. 1j) (expressed in <25% of cells in other populations). Statistical analysis was performed using a Wilcoxon rank-sum test; adjusted $P < 0.0001$. Selected HSC genes showing endothelial cell (endo) expression are also included. Ery, erythroid. Bold, HSC signature genes. **c**, Feature plots of HSC-enriched genes *SPINK2* and *RAB27B* (left), and *PROCR*, shared with HSC and endothelial cells (right). **d**, Quantification of module-selected HSCs in intra- and extraembryonic tissues in a week 4.5 (CS14) conceptus. **e**, Model scheme depicting nascent HSCs in the AGM and extraembryonic tissues in CS14–15 (4.5–5 weeks) embryos. The HSC molecular signature distinguishes nascent HSCs from progenitors and differentiated cells. **e** was partially created using BioRender.com.

Some haematopoietic cells in CS14 liver, head and heart expressed *SPINK2*, despite lacking other HSC landmarks (Extended Data Fig. 2a–h). The liver *SPINK2*$^+$ cells possessed a unique immune-system-related expression signature that was distinct from AGM *SPINK2*$^+$ HSCs, but shared with *SPINK2*$^+$ cells in the head and the heart (Extended Data Fig. 2i, j and Supplementary Table 3). Many CS14 liver *SPINK2*$^+$ cells co-expressed genes associated with progenitors (*CD34*, *KIT*), lymphoid differentiation (*IL7R*, *IL2RG*) and EMP/microglial development (*MRC1*, *CX3CR1*)[9,10] (Extended Data Fig. 2j–m). In contrast to prevalent myeloid

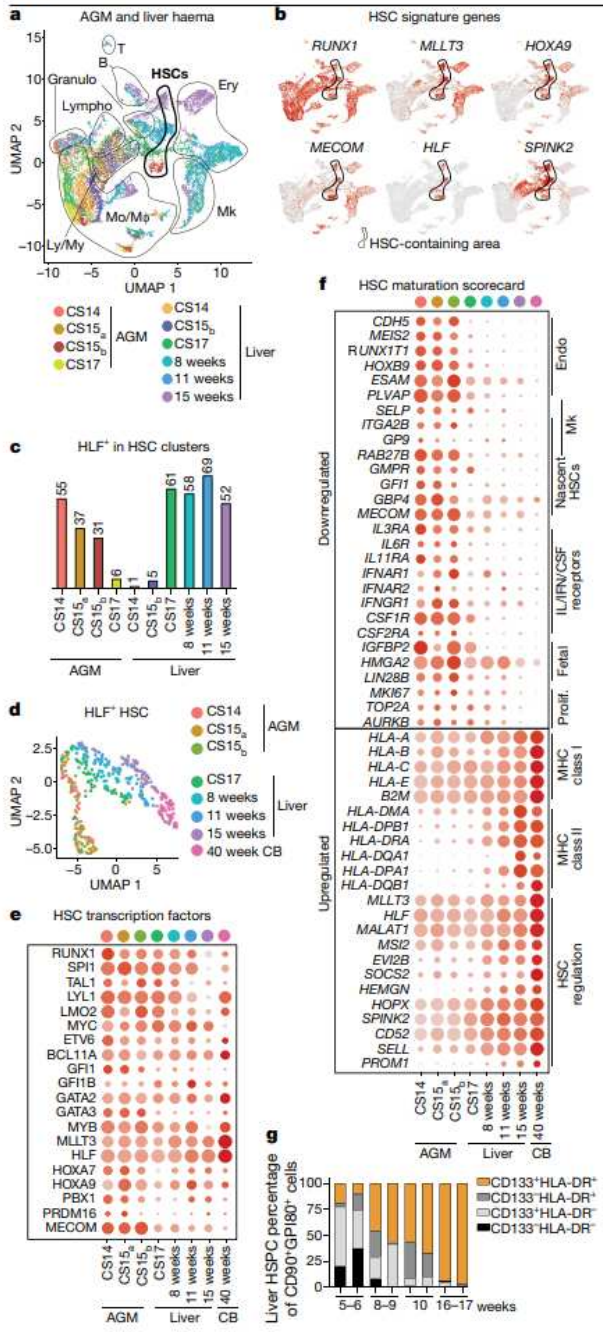

Fig. 2 | HSC developmental maturation associates with stage-specific molecular programs. a, scRNA-seq analysis of four AGM tissues (CS14–17) and six livers (CS14 to week 15) from eight concepti. $n = 10$ biologically independent samples. UMAP analysis of haematopoietic clusters (RUNX1$^+$CD45$^+$) combined from all tissues displaying haematopoietic cell (haema) types: HSC, lympho-myeloid (Ly/My), monocyte/macrophage (Mo/Mφ), granulocyte (Granulo), lymphoid (T lymphocyte (T), B lymphocyte (B)), erythroid, megakaryocytic (Mk). **b**, Feature plots showing the expression of HSC molecular signature genes (*HLF*$^+$ HSCs are circled). **c**, *HLF*$^+$ cells within HSC clusters in each tissue. **d**, *HLF*$^+$ HSC from tissues containing >10 HSCs, reclustered and shown in the UMAP analysis. **e**, HSC transcription factors in *HLF*$^+$ HSCs from different tissues. **f**, HSC maturation scorecard dot plot showing the selected genes that are upregulated or downregulated during HSC maturation (Prolif., proliferation). **g**, Flow cytometry quantification of the HSC maturation markers HLA-DR and PROM1 in fetal liver HSCs (CD43$^+$CD45^{mid}CD34$^+$CD38$^{low/-}$CD90$^+$GPI-80$^+$) at different stages. $n = 2$ biologically independent samples per stage.

differentiation, megakaryocytic/erythroid cells were rare. These data suggest that CS14 liver *SPINK2*$^+$ cells are lympho-myeloid progenitors (LMPs) that populate the liver before HSCs. Thus, lymphoid potential alone does not predict the presence of HSCs.

HSC maturation in the liver

To determine the timing of when HSCs colonize the liver, HSC-containing clusters expressing HSC signature genes (*RUNX1*$^+$*HOXA9*$^+$*MLLT3*$^+$*MECOM*$^+$*HLF*$^+$*SPINK2*$^+$) were identified in different-stage AGM (4.5–6 weeks) and livers (4.5–15 weeks), both in individual tissues and in combined *RUNX1*$^+$ cells (Fig. 2a–c and Extended Data Fig. 3a). Quantifying *HLF*$^+$ HSCs in such clusters revealed that the HSC transition to the liver occurred at around 6 weeks. As the nascent HSC scorecard revealed temporal changes among *HLF*$^+$ HSCs (Extended Data Fig. 3b) and uniform manifold approximation and projection (UMAP) analysis ordered *HLF*$^+$ HSCs by developmental age (Fig. 2d), we evaluated HSC maturation systematically. Although AGM HSCs already expressed 20 known HSC transcriptional regulators, documenting the transcriptional identity of HSCs (Fig. 2e), pseudotime analysis of *HLF*$^+$ HSCs identified temporally regulated programs highlighted in the HSC maturation scorecard (Fig. 2f, Extended Data Fig. 3c and Supplementary Table 4). Genes characteristic of AGM HSCs include those for endothelial and megakaryocytic surface antigens (*CDH5*, *ITGA2B*) and cytokine receptors (*IL3RA*, *CSF1R*) associated with myeloid progenitors and early HSC development[41,42]. Genes linked to fetal properties (*LIN28B*, *IGFBP2*, *HMGA2*[43]), and proliferative activity, oxidative phosphorylation, and glucose and nucleotide metabolism were gradually suppressed in liver HSCs, reflecting a transition to a homeostatic state (Fig. 2f and Extended Data Fig. 3d).

Genes in the antigen-presentation category, which was previously linked to maturation to transplantable HSCs in mice[44], were significantly upregulated during HSC maturation, especially MHC class II genes (Fig. 2f and Extended Data Fig. 3d). The expression of *MLLT3* and its targets (*HLF*, *MSI2*, *PROM1*)[34] also increased over time. Whereas *HOXA5–9* were maintained during HSC maturation, the posterior *HOXB7–9* genes declined (Extended Data Fig. 3e).

Selecting HSCs using HSC module score (*RUNX1*, *HOXA9*, *MLLT3*, *MECOM*, *HLF*, *SPINK2*) and MAGIC imputation confirmed HSC colonization to the liver at 6 weeks and organization by developmental age (Extended Data Fig. 3f–j). Module scores for immaturity (*CDH5*, *MEIS2*, *IGFBP2*, *HOXB9*) and maturity (*PROM1*, *HLA-DRA*, *HEMGN*, *MSI2*) endorsed a maturational switch in liver after 8 weeks (Extended Data Fig. 3k, l). Flow cytometry analysis of liver and cord blood (CB) CD34$^+$CD38^{low}CD90$^+$ HSPCs and liver GPI-80$^+$ HSCs[45,46] confirmed that there was a temporal increase in HLA-DR and PROM1 expression, pinpointing landmarks for human HSC maturation (Fig. 2g, Extended Data Fig. 3m, n and Supplementary Table 4).

HSC origin in arterial endothelium

To define the cellular precursor of human HSCs, we evaluated connections between haemato-vascular populations in CS14–15 AGM using UMAP analysis (Fig. 3a, b and Extended Data Fig. 4a). A direct developmental trajectory was observed from GJA5$^+$ arterial endothelial cells (ECs) to HSCs, but not other haematopoietic cells. Selecting CDH5$^+$RUNX1$^+$CD45$^-$ putative HE identified cells bridging the arterial and HSC clusters (Fig. 3b). Genes significantly enriched in HE or upregulated or downregulated before or after were

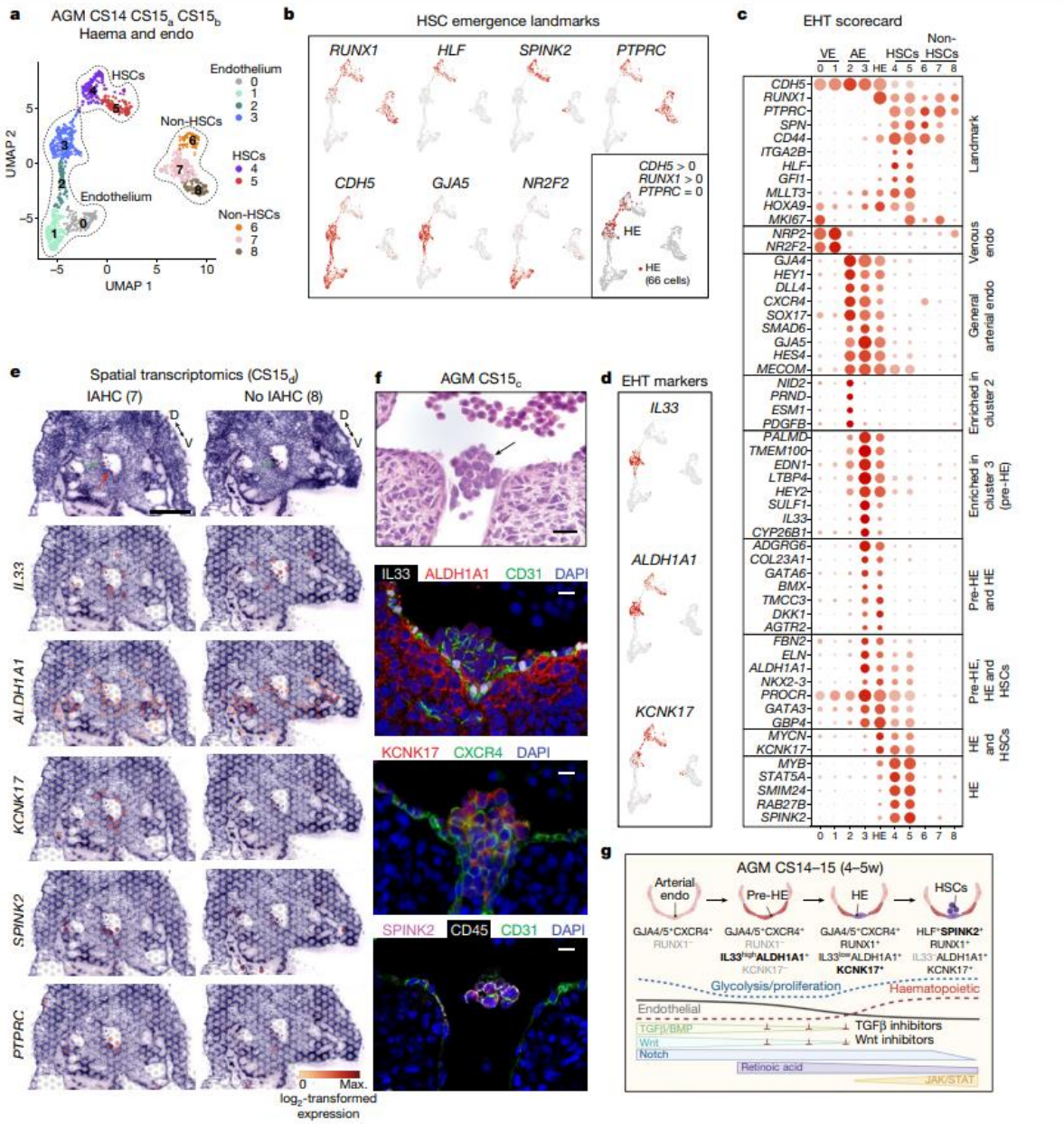

Fig. 3 | HSCs emerge from distinct arterial endothelial cells. a, UMAP analysis of haemato-vascular populations (*CDH5*$^+$ endothelium, *RUNX1*$^+$*HLF*$^+$ HSCs and *RUNX1*$^+$*HLF*$^-$ other haematopoietic cells) from CS14–15 AGM tissues ($n = 3$ biologically independent samples). The contribution of cells in venous EC and non-HSC clusters was balanced. **b**, UMAP feature plots displaying the expression of landmark genes for HSC emergence. HE were selected on the basis of co-expression of *RUNX1* and *CDH5* and the absence of CD45 (bottom right; 66 cells). **c**, EHT scorecard dot plot showing EHT landmark genes and genes that are co-regulated at different stages of EHT in each cluster (clusters 0–8) and HE. Genes that were significantly enriched in HE compared with in other populations, or up- or downregulated during the transition to and from HE, were selected. **d**, UMAP feature plots displaying pre-HE (*IL33* and *ALDH1A1*) and HE (*ALDH1A1* and *KCNK17*) EHT markers. **e**, Spatial transcriptomics of CS15$_d$ (5 weeks) embryo transverse sections. Top, H&E staining of two sections between the vitelline and umbilical arteries, focused on the dorsal aorta and the surrounding region. Red arrow, IAHC; green arrows, red blood cells; D, dorsal; V, ventral. Scale bars, 500 µm. Bottom, spatial expression of EHT genes, with the default colour scale from the Loupe browser, which represents $\log_2$-transformed expression from 0 to the maximum value in the spots. Max., maximum. Each dot is 55 µm and shows combined expression of 1–10 cells. **f**, H&E section (section 240) of CS15$_c$ (5 weeks) aorta at the intersection with the vitelline artery. The arrow indicates the IAHC. Immunofluorescence staining of the aorta for IL33, ALDH1A1, CD31 and DAPI (section 251); CXCR4, KCNK17 and DAPI (section 254); and SPINK2, CD45, CD31 (also known as PECAM-1) and DAPI (section 239). Scale bars, 20 µm. Individual antibody staining experiments were performed at least three times in independent embryos with a comparable staining pattern. **g**, Schematic summarizing the model for EHT involving the specification of pre-HE and HE from arterial EC and HSC emergence. Stage-specific markers and signalling switches specific for each stage are shown (bold, distinctive markers; gray, absent markers). **g** was created using BioRender.com.

displayed across the populations in the endothelial-to-haematopoietic transition (EHT) scorecard (Fig. 3c and Supplementary Table 5). A similar HOXA pattern between the arterial endothelium, HE and HSCs suggested a shared origin (Extended Data Fig. 4b). Classical arterial genes (*GJA5*, *CXCR4*, *SOX17*) were expressed across arterial clusters and downregulated after HE (Fig. 3c). Arterial cluster 3 and HE displayed a mature arterial phenotype (*LTBP4*, *TMEM100*) and exclusive expression of several regulatory molecules (Wnt inhibitor *DKK1*, angiotensin-II receptor *AGTR2*). These related populations could be distinguished by the expression of *IL33* and *SULF1* in arterial cluster 3, and *KCNK17*, *RUNX1* and *MYCN* in HE. *ALDH1A1* (RA signalling, essential for HSC fate[10,47]) was expressed from arterial cluster 3 to HSCs. Pseudotime analysis aligned these markers in a continuum with HSCs, supporting *IL33*$^+$*ALDH1A1*$^+$ arterial ECs (herein termed pre-HE) as the precursor for HSC-forming HE (Extended Data Fig. 4c–f).

Unsupervised clustering of variable genes across pseudotime uncovered precisely timed regulatory switches during the EHT (Supplementary Table 6 and Extended Data Fig. 4g–k). HE displayed a transition from endothelial to haematopoietic transcriptional programs, and were preceded by metabolically quiescent arterial pre-HE displaying TGFβ/BMP and Notch signalling that declined after HSC emergence. TGFβ/BMP inhibitors (*SMAD6*, *SMAD7*) and Wnt inhibitors (*DKK1*, *DKK2*) peaked in pre-HE and HE. Emerging HSCs maintained *ALDH1A1* and induced cytokine receptors and JAK–STAT signalling (Fig. 3g).

Localization of HSC emergence to IAHCs

To visualize HSC emergence, we performed Visium spatial transcriptomics analysis of a CS15$_d$ (5 weeks) embryo. Seven transverse sections, including liver, AGM, gut and vitelline and umbilical vessels, were sequenced (Fig. 3e, Extended Data Figs. 5 and 6 and Supplementary Tables 1 and 7). Cluster analysis identified regions with expected tissue/organ-specific genes (*NEUROD1* in the neural tube, *MYOD1* in the myotome), whereas the expression of haematopoietic genes (*GYPA* in erythroid cells, *RUNX1* and *HLF* in HSPCs) was scattered (Extended Data Fig. 5a–c and Supplementary Table 7). Many aorta sections showed evidence of *RUNX1* expression (sections 2, 4, 6 and 7) and some expressed *HLF* (sections 2 and 7). Haematoxylin and eosin (H&E) staining revealed IAHCs in these sections (Extended Data Fig. 6b).

The localization of EHT markers in CS15$_d$ embryo sections showed strong but not exclusive expression of *GJA5* in arteries, whereas the pre-HE marker *IL33* was mainly concentrated in the aorta and umbilical and vitelline arteries (Fig. 3e and Extended Data Fig. 6a, b). The pre-HE/HE marker *ALDH1A1* was expressed on the ventral side of the aorta, and in the mesonephros (kidney) and liver epithelia. The HE marker *KCNK17* was expressed in aorta sections with IAHCs (sections 2, 4, 6 and 7), similar to *RUNX1*. *SPINK2* was detected in IAHCs (sections 2, 4, 6 and 7) and the liver (sections 1 and 2), of which the liver, but not IAHCs, expressed the LMP marker *IL7R*. These data localize pre-HE along the dorsal aorta and other major arteries, and HE and HSC markers with IAHCs.

Immunofluorescence analysis of CS15$_c$ AGM revealed nuclear *IL33* expression in CD31$^+$ aortic ECs that co-expressed *ALDH1A1* and associated ventrally with ALDH1A1$^+$ stromal cells (Fig. 3f and Extended Data Fig. 6c). While IAHC cells maintained *CD31* and low *ALDH1A1* expression, they downregulated *IL33*, which has been associated with resting ECs[48]. *KCNK17* was induced in *CXCR4*$^+$ IAHCs and adjacent aortic ECs, confirming the arterial identity of HE and IAHCs. Together with *SPINK2* and *CD45* expression in IAHCs, these markers indicate HSC emergence (Fig. 3f and Extended Data Fig. 6c).

Haematopoietic waves preceding HSCs

To evaluate the specificity of pre-HE and HE markers to HSC formation, we compared CS13–17 AGM EHT data with HE from earlier (CS10–11) embryos[16] and yolk sac[10]. Unsupervised clustering of CDH5$^+$ ECs and *RUNX1*$^+$ haematopoietic cells documented two waves of *HLF*$^+$*SPINK2*$^+$ cells (CS13–17 AGM versus CS10–11 embryo and yolk sac) that associated with distinct EC populations (Extended Data Fig. 7a, b).

Surprisingly, *HLF*$^+$*SPINK2*$^+$ HPCs in CS10–11 embryo and yolk sac also expressed several nascent HSC scorecard genes (Extended Data Fig. 7d). To pinpoint the differences between the two haemogenic waves, previously termed early (HSC-independent) and late (HSC-forming)[16], we generated 'HSPC wave' and 'endo wave' scorecards and displayed wave-specific and common genes in each population (Supplementary Table 8 and Extended Data Fig. 7e, f). In contrast to *KCNK17*, which was expressed in HE and HSPCs in both waves, CS13–17 HE and HSCs uniquely expressed transcriptional regulators from the 'misregulated in cancer' Gene Ontology category (*RUNX1T1*, *MECOM*, *MLLT3*, *NKX2-3*), whereas CS10–11 HE and HPCs expressed *LIN28A*, *GAD1* and *FGF23*. *HOXA7* and *HOXA9* were absent from early HPCs, whereas *HOXB7* and *HOXB9* were observed in both waves. CS13–17 HSCs also expressed higher levels of the HSC surface markers *ITGA4*, *PROCR* and *EMCN*.

The mature arterial phenotype[16] was observed only with HSC-forming HE and associated ECs (Extended Data Fig. 7f). Although NOTCH1 signalling was active in both HE populations, there were differences in the target genes. Pre-HE/HE markers were specific to the HSC-forming wave, starting with *IL33* induction in the CS11 embryo (Extended Data Fig. 7b, c). Co-expression of *ALDH1A1* and *CYP26B1* in CS14–15 pre-HE/HE imply fine-tuned RA signalling in this stage, which coincides with Wnt and TGFβ/BMP inhibition and diminished glycolytic activity. These findings show that *ALDH1A1*$^+$*KCNK17*$^+$*RUNX1*$^+$ HSC-forming HE exist in a defined window and develop from metabolically quiescent *IL33*$^+$*ALDH1A1*$^+$ arterial endothelium (Fig. 3g).

Extraembryonic HSCs link to the liver

To position extraembryonic HSPCs within human haematopoietic ontogeny, we analysed intra- and extraembryonic haematopoietic cells from all tissues together. HSC module score analysis linked CS14 extraembryonic HSPCs to the HSC compartment, while immaturity and maturity module scores and UMAP analysis placed extraembryonic *HLF*$^+$ HSPCs between CS14–15 AGM and CS17 liver (Extended Data Fig. 8a–f). The association of CS14 placental and yolk sac HSCs with CS17 liver HSCs suggests that there is a relationship between HSCs in extraembryonic tissues and liver colonization.

The HSPC waves scorecard placed CS14 AGM and extraembryonic HSPCs to the HSC wave, as they lack early HPC genes and express robustly most HSC regulators (*MLLT3*, *MECOM*) and HSC surface markers (*ACE*[37]), although *HOXA9*, *NKX2-3* and *EMCN* expression was weaker in extraembryonic HSPCs (Extended Data Fig. 8g). The HSC maturation scorecard also positioned extraembryonic HSCs between the AGM and liver (Extended Data Fig. 8h).

Analysis of differentiation trajectories implied that there is minimal differentiation of HSCs in CS14–15 AGM, whereas CS10–11 embryo and yolk sac HPCs displayed megakaryocytic/erythroid differentiation (Extended Data Fig. 8j, k). At CS14, lymphoid and myeloid cells (mainly macrophages) associated with liver *SPINK2*$^+$*IL7R*$^+$ LMPs, which disappeared by week 8. Macrophages clustered by anatomical location even within the same conceptus.

The first evidence of true multilineage haematopoiesis and robust erythropoiesis in the liver was observed at CS17, with a trajectory from CS17 liver and CS14 placental HSPCs. *HOXA9* expression in liver erythroid precursors at 6–15 weeks and in a second lymphoid wave appearing at 8 weeks suggests that they originate from HSCs. Contrasting the stable *HOXA9* expression in HSCs and their immediate progeny throughout ontogeny, *HOXB9* was expressed only in immature HSCs and CS10–11 HPCs (Fig. 2e and Extended Data Fig. 8g, k). *SPINK2*$^+$ LMPs expressed neither *HOXA9* nor *HOXB9* (Supplementary Table 3). The distinct regulation of *HOXA* and *HOXB* genes provides molecular landmarks for tracking the HSC lineage, their maturation stage and immediate progeny

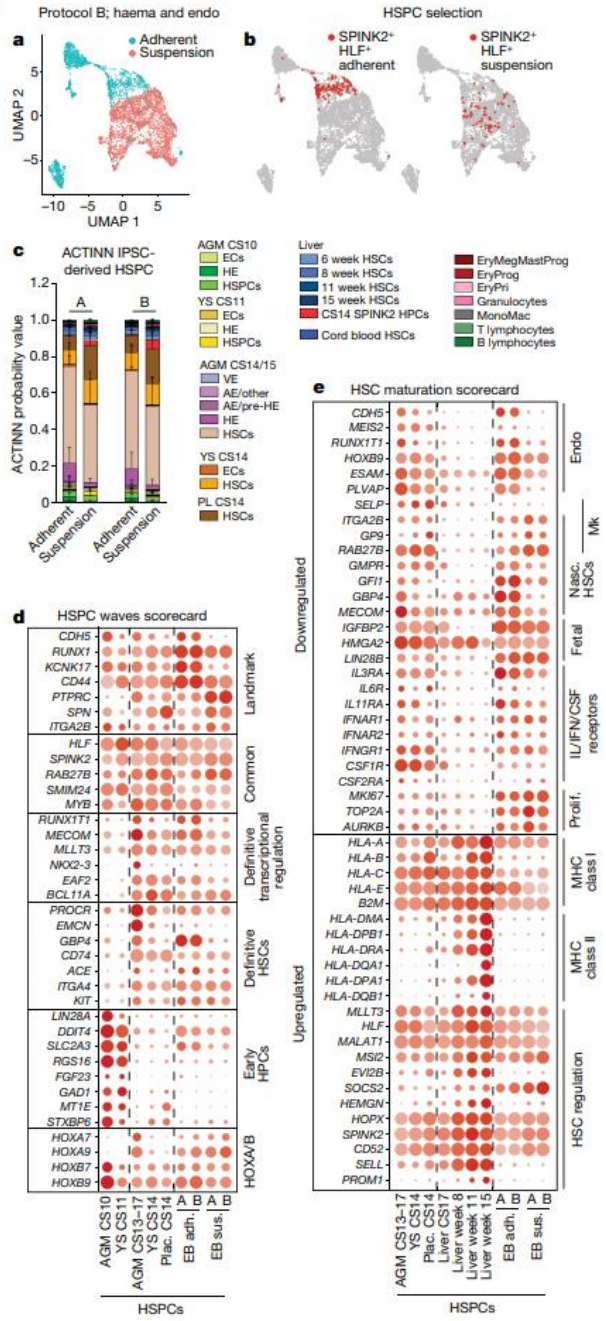

Fig. 4 | The map of human HSC ontogeny uncovers the developmental stage of PS-cell-derived HSPCs. a, UMAP analysis of adherent (adh.) and suspension (sus.) fractions in *CDH5*$^+$*RUNX1*$^+$ haemato-vascular cells (PS cell differentiation protocol B). **b**, UMAP analyses highlighting *SPINK2*$^+$*HLF*$^+$ HSPCs in PS-cell-derived *CDH5*$^+$ *RUNX1*$^+$ cells from adherent and suspension fractions. **c**, Summary of ACTINN probability values for PS-cell-derived *SPINK2*$^+$*HLF*$^+$ HSPCs from adherent and suspension fractions (protocols A and B). Data are mean ± s.d. $n = 349$ *SPINK2*$^+$*HLF*$^+$ adherent cells and $n = 68$ *SPINK2*$^+$*HLF*$^+$ suspension cells (protocol A); and $n = 141$ *SPINK2*$^+$*HLF*$^+$ adherent cells and $n = 71$ *SPINK2*$^+$*HLF*$^+$ suspension cells (protocol B). VE, venous endothelium; AE, arterial endothelium; EryMegMastProg, erythroid/megakaryocytic/mastocytic precursors; EryProg, erythroid precursors; EryPri, primitive erythroid; MonoMac, monocytes/macrophages. **d**, **e**, HSPC wave scorecard genes (**d**) and HSC maturation scorecard genes (**e**) in selected PS-cell-derived populations from differentiation protocols A and B compared with their in vivo counterparts. Nasc, nascent; Plac., placenta; YS, yolk sac.

Generation of AGM-like HSPCs in vitro

To evaluate human PS cell differentiation towards HSCs and identify their in vivo counterparts, we analysed scRNA-seq data from haemato-vascular cells that were generated using two closely related protocols that promote intraembryonic-type multilineage haematopoiesis[49]. These protocols (A and B) use the swirler embryoid body method with HSC supportive cytokines, small molecules, and modifications of WNT and ACTIVIN/BMP signalling (Extended Data Fig. 9a). Fluorescence-activated cell sorting (FACS) and scRNA-seq analysis of the adherent and suspension fractions from the embryoid body identified cells with surface phenotype of human HSPCs (CD34$^+$CD38^{low}CD90$^+$CD45$^+$) and arterial and haemogenic endothelia (Supplementary Table 1).

Using the scRNA-seq data of human haematopoietic cell types as a reference, the ACTINN neural network-based program matched induced PS (iPS)-cell-derived haematopoietic cells in the adherent fraction with nascent AGM HSCs. The suspension fraction included cells that were reminiscent of AGM, placental and yolk sac HSCs, as well as CS14 liver *SPINK2*$^+$ progenitors (Extended Data Fig. 9b, c). A close match of *HLF*$^+$*SPINK2*$^+$ HSPCs to immature AGM and extraembryonic HSCs was confirmed using ACTINN and nascent HSC, HSPC waves and HSC maturation scorecards (Fig. 4b–e and Extended Data Fig. 9d).

ACTINN matched PS-cell-derived ECs to CS14/15 AGM arterial pre-HE and HE (Extended Data Fig. 9e). However, scorecards for the EHT and endo waves identified differences compared with the EHT in vivo, including undetectable or weak expression of *ALDH1A1* and Wnt and TGFβ/BMP inhibitors in pre-HE and HE in vitro (Extended Data Fig. 9f, g). These signalling discrepancies during EHT and a lack of HSC maturation pinpoint molecular bottlenecks for the generation of functional human HSCs in vitro.

Discussion

We established a scRNA-seq map of human HSC ontogeny from early first trimester to birth to enable the study of human HSC and progenitor hierarchies in developmental tissues that are otherwise difficult to access. We defined a six-gene signature (*RUNX1*$^+$*HOXA9*$^+$*MLLT3*$^+$*MECOM*$^+$ *HLF*$^+$*SPINK2*$^+$) that distinguishes human HSCs from lineage-restricted progenitors in all stages, even when traditionally used assays to identify HSCs are too stringent (transplantability) or non-specific (multilineage haematopoiesis or lymphoid potential). We curated cell-type- and stage-specific scorecards to provide a lens for evaluating haematopoietic cells generated in vivo or in vitro across human developmental timeline (Extended Data Fig. 10).

These methods detected nascent HSCs in the AGM region, the placenta and yolk sac before HSCs populate the liver. Our analysis positions extraembryonic HSCs one step downstream from the most immature AGM HSCs, and suggests they are on track to colonize the liver to initiate multilineage haematopoiesis, including HSC-driven erythropoiesis. Our finding that transient *SPINK2*$^+$*IL7R*$^+$ LMPs colonize the liver before HSCs may elucidate the origin of tissue-resident macrophages and innate-like lymphoid cells in human embryo. Comparing in vivo- and in vitro-generated haematopoietic cells using scorecards and neural-network-based label transfer validated human iPS cell differentiation to AGM and placental-stage HSC-like cells, and HPCs reminiscent of liver LMPs. These discoveries will fuel future studies to decipher the development and disease relevance of these diverse populations, and help to understand the aetiology of blood diseases that develop in utero.

Our comprehensive map of human HSC ontogeny elucidates the HSC maturation process that is necessary for robust engraftment ability.

Although HSC transcriptional identity is apparent after HSC emergence, maturation to transplantable HSCs in the liver involves amplification of the MLLT3-driven HSC self-renewal program and changes in HSC surface phenotype and functional properties. As HSCs in the liver suppress fetal programs (*LIN28B, IGFBP2*) and transition towards a homeostatic, post-natal HSC state, they acquire surface expression of PROM1 and MHC-class II molecules (Extended Data Fig. 10). Our finding that in vitro-generated HSPCs were unable to complete maturation to liver stages highlights the importance of understanding the molecular underpinnings of the HSC maturation process and developing protocols that mimic liver HSC niches.

Single-cell analysis of the AGM during the developmental window when HSCs emerge (CS14–15) documented a connection between HSCs and arterial endothelium, clarifying the cellular origin of HSCs. Using the molecular landmarks *IL33*, *ALDH1A1*, *KCNK17* and *SPINK2* of the EHT (Extended Data Fig. 10), spatial transcriptomics and immunofluorescence analysis confirmed that IAHCs are sites of HSC emergence. Our analysis implied that the activation of the haemogenic program in ECs before the establishment of arterial identity generates differentiation-primed progenitors that express embryonic genes (*LIN28A*) and lack HOXA patterning and a robust HSC self-renewal program (Extended Data Fig. 10). Thus, the outcome of EHT depends on the signalling environment in which the haemogenic endothelial precursor was specified. Previous findings from human PS cell differentiation studies corroborate this model[4,50]. Moreover, identification of landmark genes for EHT and coordinated signalling switches that govern HSC specification and emergence helped to pinpoint potential shortcomings in the efficiency/timing of establishing bona fide arterial pre-HE and HE in vitro. Access to the detailed molecular map of human HSC development that is informed by the human embryo will help to uncover the correct instructions for generating fully functional HSCs in vitro.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04571-x.

1. Tavian, M., Hallais, M. F. & Peault, B. Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development* **126**, 793–803 (1999).
2. Ivanovs, A. et al. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *J. Exp. Med.* **208**, 2417–2427 (2011).
3. Boisset, J. C. et al. Progressive maturation toward hematopoietic stem cells in the mouse embryo aorta. *Blood* **125**, 465–469 (2015).
4. Ivanovs, A. et al. Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337 (2017).
5. Hadland, B. K. et al. Endothelium and NOTCH specify and amplify aorta-gonad-mesonephros–derived hematopoietic stem cells. *J. Clin. Invest.* **125**, 2032–2045 (2015).
6. Ivanovs, A., Rybtsov, S., Anderson, R. A. & Medvinsky, A. Vast self-renewal potential of human AGM region HSCs dramatically declines in the umbilical cord blood. *Stem Cell Rep.* **15**, 811–816 (2020).
7. Ghosn, E., Yoshimoto, M., Nakauchi, H., Weissman, I. L. & Herzenberg, L. A. Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes. *Development* **146**, dev170571 (2019).
8. Palis, J. Hematopoietic stem cell-independent hematopoiesis: emergence of erythroid, megakaryocyte, and myeloid potential in the mammalian embryo. *FEBS Lett.* **590**, 3965–3974 (2016).
9. 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**, e20201729 (2021).
10. Bian, Z. et al. Deciphering human macrophage development at single-cell resolution. *Nature* **582**, 571–576 (2020).
11. Ginhoux, F. et al. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* **330**, 841–845 (2010).
12. Gomez Perdiguero, E. et al. Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature* **518**, 547–551 (2015).
13. Zeng, Y. et al. Single-cell RNA sequencing resolves spatiotemporal development of pre-thymic lymphoid progenitors and thymus organogenesis in human embryos. *Immunity* **51**, 930–948 (2019).
14. Zhou, F. et al. Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* **533**, 487–492 (2016).
15. 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).
16. Zeng, Y. et al. Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res.* **29**, 881–894 (2019).
17. Popescu, D.-M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
18. 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).
19. Gekas, C., Dieterlen-Lievre, F., Orkin, S. H. & Mikkola, H. K. The placenta is a niche for hematopoietic stem cells. *Dev. Cell* **8**, 365–375 (2005).
20. Ottersbach, K. & Dzierzak, E. The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. *Dev. Cell* **8**, 377–387 (2005).
21. Rhodes, K. E. et al. The emergence of hematopoietic stem cells is initiated in the placental vasculature in the absence of circulation. *Cell Stem Cell* **2**, 252–263 (2008).
22. Li, Z. et al. Mouse embryonic head as a site for hematopoietic stem cell development. *Cell Stem Cell* **11**, 663–675 (2012).
23. Nakano, H. et al. Haemogenic endocardium contributes to transient definitive haematopoiesis. *Nat. Commun.* **4**, 1564 (2013).
24. Bárcena, A., Muench, M. O., Kapidzic, M. & Fisher, S. J. A new role for the human placenta as a hematopoietic site throughout gestation. *Reprod. Sci.* **16**, 178–187 (2009).
25. Robin, C. et al. Human placenta is a potent hematopoietic niche containing hematopoietic stem and progenitor cells throughout development. *Cell Stem Cell* **5**, 385–395 (2009).
26. Van Handel, B. et al. The first trimester human placenta is a site for terminal maturation of primitive erythroid cells. *Blood* **116**, 3321–3330 (2010).
27. Heck, A. M., Ishida, T. & Hadland, B. Location, location, location: how vascular specialization influences hematopoietic fates during development. *Front. Cell Dev. Biol.* **8**, 602617 (2020).
28. Zovein, A. C. et al. Fate tracing reveals the endothelial origin of hematopoietic stem cells. *Cell Stem Cell* **3**, 625–636 (2008).
29. Zhu, Q. et al. Developmental trajectory of prehematopoietic stem cell formation from endothelium. *Blood* **136**, 845–856 (2020).
30. Crosse, E. I. et al. Multi-layered spatial transcriptomics identify secretory factors promoting human hematopoietic stem cell development. *Cell Stem Cell* **27**, 822–839 (2020).
31. Ditadi, A. et al. Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nat. Cell Biol.* **17**, 580–591 (2015).
32. Dou, D. R. et al. Medial HOXA genes demarcate haematopoietic stem cell fate during human development. *Nat. Cell Biol.* **18**, 595–606 (2016).
33. Ng, E. S. et al. Differentiation of human embryonic stem cells to HOXA$^+$ hemogenic vasculature that resembles the aorta–gonad–mesonephros. *Nat. Biotechnol.* **34**, 1168–1179 (2016).
34. Calvanese, V. et al. MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. *Nature* **576**, 281–286 (2019).
35. Kataoka, K. et al. Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. *J. Exp. Med.* **208**, 2403–2416 (2011).
36. Komorowska, K. et al. Hepatic leukemia factor maintains quiescence of hematopoietic stem cells and protects the stem cell pool during regeneration. *Cell Rep.* **21**, 3514–3523 (2017).
37. Jokubaitis, V. J. et al. Angiotensin-converting enzyme (CD143) marks hematopoietic stem cells in human embryonic, fetal, and adult hematopoietic tissues. *Blood* **111**, 4055–4063 (2008).
38. Pellin, D. et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* **10**, 2395 (2019).
39. Lehnertz, B. et al. HLF expression defines the human hematopoietic stem cell state. *Blood* **138**, 2642–2654 (2021).
40. Lee, B. et al. Impaired spermatogenesis and fertility in mice carrying a mutation in the Spink2 gene expressed predominantly in testes. *J. Biol. Chem.* **286**, 29108–29117 (2011).
41. McKinney-Freeman, S. et al. The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell* **11**, 701–714 (2012).
42. Robin, C. et al. An unexpected role for IL-3 in the embryonic development of hematopoietic stem cells. *Dev. Cell* **11**, 171–180 (2006).
43. Copley, M. R. et al. The Lin28b–let-7–Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat. Cell Biol.* **15**, 916–925 (2013).
44. Kieusseian, A., Brunet de la Grange, P., Burlen-Defranoux, O., Godin, I. & Cumano, A. Immature hematopoietic stem cells undergo maturation in the fetal liver. *Development* **139**, 3521–3530 (2012).
45. Prashad, S. L. et al. GPI-80 defines self-renewal ability in hematopoietic stem cells during human development. *Cell Stem Cell* **16**, 80–87 (2014).
46. Vanuytsel, K. et al. Multi-modal profiling of human fetal liver hematopoietic stem cells reveals the molecular signature of engraftment. *Nat. Commun.* **13**, 1103 (2022).
47. Chanda, B., Ditadi, A., Iscove, N. N. & Keller, G. Retinoic acid signaling is essential for embryonic hematopoietic stem cell development. *Cell* **155**, 215–227 (2013).
48. Ali, S. et al. The dual function cytokine IL-33 interacts with the transcription factor NF-κB to dampen NF-κB-stimulated gene transcription. *J. Immunol.* **187**, 1609–1616 (2011).
49. Motazedian, A. et al. Multipotent RAG1$^+$ progenitors emerge directly from haemogenic endothelium in human pluripotent stem cell-derived haematopoietic organoids. *Nat. Cell Biol.* **22**, 60–73 (2020).
50. Uenishi, G. I. et al. NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells. *Nat. Commun.* **9**, 1828–1828 (2018).

Methods

Ethical statement

First trimester tissues were obtained from University of Tubingen and delivered to UCLA within 48 h after the procedure. The Ethics Committee at the Medical Faculty of the Eberhard Karls University Tübingen and at the University Hospital in Tübingen approved the use of human embryo tissues from elective terminations for HSC research (290/2016BO1). Second-trimester fetal liver tissues from elective terminations performed at Family Planning Associates were provided to the UCLA CFAR Cell and Gene Therapy core for distribution to UCLA investigators. All human fetal tissue samples used were discarded material from elective terminations that were obtained following informed consent. The donated human fetal tissues were anonymized and did not carry any personal identifiers. In all cases, the decision to terminate the pregnancy occurred before the decision to donate tissue. No payments were made to donors and the donors knowingly and willingly consented to provide research materials without restrictions for research and for use without identifiers. The UCLA IRB determined that the provision of anonymized fetal material for research does not constitute human subject research according to the US Federal regulations because the tissues are anonymized, that is, provided without any direct or indirect identifiers that could be linked back to a living individual. As a result, investigators using such material are not engaged in research that is subject to IRB oversight. All donors gave informed consent in compliance with the US Public Health Service Act, Sections 498A and 498B for the use of fetal material in research. All human tissue materials were treated as biosafety level 2 and approved by the UCLA Institutional Biosafety Committee (IBC) (BUA-2016-142-001, BUA-2019-186-001). Studies using human iPS cells were approved by The Royal Children's Hospital Human Research Ethics Committee (33001A).

Isolation of human developmental tissues

Sample age was denoted as developmental age, that is, two weeks less than gestational age, as determined by ultrasound or estimated from last menstrual period. Embryos until developmental week 7 were assigned Carnegie stages (CS) according to the morphological criteria reported in https://embryology.med.unsw.edu.au. After procurement, the tissue samples were washed in sterile Dulbecco's phosphate buffered saline (DPBS, Invitrogen), placed in sterile DPBS that was supplemented with 5% FBS (Thermo Fisher Scientific), 1% penicillin–streptomycin (Gibco) and 2.5 μg ml^{-1} amphotericin B (Sigma-Aldrich), and processed for flow cytometry sorting or analyses within 48 h. The first-trimester tissues included the AGM region surrounding the abdominal aorta, the liver, the placenta, the yolk sac, vitelline and umbilical vessels (distal parts), the head and the heart. Tissues were digested in 2.5 U dispase (Gibco), 90 mg collagenase A (Worthington) and 0.075 mg DNase I (Sigma-Aldrich) per ml in PBS containing 10% FBS, for 20–45 min at 37 °C. Cells were disaggregated by pipetting and filtered through a 70 μm cell strainer.

Second-trimester fetal livers were collected into PBS 5% FBS (Hyclone) and mechanically dissociated using scalpels and syringes, followed by the enzymatic dissociation described above. Liver and cord blood were enriched for mononuclear cells on a Lymphoprep layer according to the manufacturer's protocol (Stem Cell Technologies) and filtered through a 70 μm mesh. Placenta, second-trimester liver and cord blood were magnetically enriched for $CD34^+$ cells using the human CD34 MicroBead Kit UltraPure (Miltenyi) before FACS sorting. A complete inventory of the tissues originally published in this manuscript with descriptions of biological and technical attributes is provided in Supplementary Table 1.

Flow cytometry and cell sorting

$CD31^+$ and/or $CD34^+$ haemato-vascular cells were enriched for scRNA-seq by FACS sorting using permissive gating that resulted in 10–20-fold enrichment of these cells. The cells were stained with the following antibodies: anti-human-CD45-APC-H7 (2D1, 368516, BioLegend; 1:100), CD34-BV605 (581, 745105, BD; 1:100), CD90-APC (5E10, 559869, BD, 1:50), CD38-PE-Cy7 (HIT2, 560677, BD; 1:100), CD43-FITC (1G10, 555475, BD; 1:20), CD235a-PE (GA-R2, 340947, BD; 1:100). $CD34^+$ and/or $CD31^+$ cells were sorted into PBS 0.04% BSA using a BD FACS Aria cell sorter using FACSDiva v.8.0. Representative sorting plots are shown in Supplementary Table 1.

Flow cytometry analysis for evaluating HSC maturation was performed using anti-human-CD45-BV785 (HI30, 304048; BioLegend, 1:100), CD34-APC (581, 555824; BD, 1:20), CD90-BV421 (5E10, 562556, BD, 1:100), CD38-BV711 (HIT2, 303528, BioLegend 1:100), CD43-APCcy7 (1G10, 655430, BioLegend 1:20), GPI80-PE (3H9, D087-5, MBL 1:100), PROM1-FITC (AC133, 130-113-111, Miltenyi, 1:100), HLA-DR-PEcy7 (LN3, 25-9956-41, eBioscience 1:100) and FlowJo (Tree Star). Representative samples of the gating for maturation analysis are shown in Extended Data Fig. 3m.

Flow cytometry analysis for evaluating human PS cell differentiation was performed at day 14 using anti-human-CD45-FITC (HI30, 304054, BioLegend, 1:50), CD34-PEcy7 (581, 343516, BioLegend, 1:100), CD38-APC (HIT2, 303510, BioLegend, 1:50), CD44-APC (BJ18, 338806, BioLegend, 1:50), CD73-BV421 (AD2, 344008, BioLegend, 1:50), CD90-APC (5E10, 559869, BD, 1:50), CD90-BV421 (5E10, 328122, BioLegend, 1:50), KIT (CD117) -BV421 (10D2, 313216, BioLegend, 1:10). Representative gating strategies are shown in Supplementary Table 1.

scRNA-seq analysis

For the generation of single-cell gel beads in emulsion, cells were loaded onto a Chromium single-cell instrument (10x Genomics) with an average estimated targeted cell recovery of ~6,000 cells. Single-cell suspensions of cells in 0.4% BSA-PBS were added to each channel on the 10x chip. Cells were partitioned with Gel Beads into emulsion in the Chromium instrument where cell lysis and barcoded reverse transcription of RNA occurred after amplification. scRNA-seq libraries were prepared by using the Chromium single-cell 3′ library and gel bead kit v2 (10x Genomics). Sequencing was performed on the Illumina NovaSeq 6000 system. For the embryoid body samples, scRNA-seq libraries (v3) were prepared. For each supernatant and adherent fraction sample, 4,000–9,000 cells were captured and sequenced.

Single-cell data analysis

After sequencing, fastq files were generated using CellRanger mkfastq (v.2.1.1). The raw reads were mapped to the human reference genome (refdata-cellranger-GRCh38-1.2.0) using CellRanger count. Samples from ref. [16] were processed from the fastq file stage and aligned to the same human reference genome (refdata-cellranger-GRCh38-1.2.0). Digital expression matrix was extracted from the 'filtered_gene_bc_matrices' folder outputted by the CellRanger count pipeline. To identify different cell types and find signature genes for each cell type, the R package Seurat (v.3.1.2) was used to analyse the digital expression matrix. Cells with less than 500 unique molecular identifiers (UMIs) or greater than 5% mitochondrial expression were removed from further analysis. The Seurat function NormalizeData was used to normalize the raw counts. Variable genes were identified using the FindVariableGenes function. The ScaleData function was used to scale and centre expression values in the dataset, the number of UMIs was regressed against each gene. Principal component analysis (PCA), *t*-SNE and UMAP were used to reduce the dimensions of the data, and the first two dimensions were used in the plots. The FindClusters function was used to cluster the cells. Marker genes were found using the FindAllMarkers function for each cluster, which uses the Wilcoxon rank-sum test to determine the significance and the Benjamini–Hochberg Procedure to correct for multiple comparisons. Cell types were annotated based on the marker genes and their match to canonical markers. For differential expression analyses between clusters or cell selections, the FindMarkers

function was used. The adjusted *P* value and log-transformed fold change cut-offs used are reported in the respective Supplementary Tables for each analysis. The DotPlot function was used to illustrate the expression pattern of chosen genes in selected cells. Using dot plots, scorecards were created to highlight specific cell types or developmental processes.

To minimize the dropout effect for the scRNA-seq data, the MAGIC function from the R package Rmagic[51] was used to perform imputation with the default parameters after the NormalizeData step and before the FindVariableGenes step. The other steps remain the same as for the other analyses. The module scores were calculated using the Seurat function AddModuleScore with the default parameters, which measures the average expression levels of a set of genes, subtracted by the average expression of randomly selected control genes. Where HSPCs were selected on the basis of an HSC module score threshold, HSCs from the AGM were used as a reference to establish the minimum value for the HSC signature module score, which was then applied to the whole sample to extract the HSC module-defined cells from all tissues.

R package monocle (v.2.10.1) was applied for pseudotime analysis. The functions estimateSizeFactors and estimateDispersions were used to normalize the total expression depth across cells and estimate the dispersion of the genes. The detectGenes function was used to filter genes; genes expressed in more than 5 cells or with an average expression larger than 0.5 were considered to be expressed genes. The differentialGeneTest function was used to identify differentially expressed genes among the expressed genes. The reduceDimension function was used to reduce the dimensions of the dataset using the DDRTree method. The orderCells function was used to learn a trajectory describing the biological process that the cells are going through and calculate where each cell falls within that trajectory. The plot_cell_trajectory function was used to generate the trajectory plot. For the maturation analysis, the correlation between the gene and the pseudotime was calculated using the R function cor.test, which outputs the correlation coefficients and the *P* values using the parametric correlation test. The *P* values were adjusted using the Benjamini–Hochberg procedure, and genes with adjusted *P* values of smaller than 0.05 were used for further analysis. Genes with a correlation coefficient of greater than 0.4 or less than 0.4 were considered to be positively or negatively correlated with the pseudotime, respectively. Gene Ontology analysis (using Enrichr[52]) was then applied to the positively or negatively correlated gene set to identify enriched pathways. Dot plots were used to visualize the maturation process using selected genes. As there was an unexpected imbalance between the sexes at different stages (Supplementary Table 1), to ensure that the observed HSC maturation patterns reflect the age rather than the sex of the tissues, the results were also validated in other datasets outside this study. For EHT analysis, the plot_pseudotime_heatmap function was used to generate the pseudotime heat map. The number of gene sets was established using the num_clusters argument within the monocle function plot_pseudotime_heatmap. After testing different number of clusters, 10 clusters were chosen as it enabled us to optimally separate the different modules of genes that co-vary across pseudotime. Gene Ontology analysis (using Enrichr[52]) was applied to each gene set to identify enriched pathways.

Spatial transcriptome analysis

The $CS15_d$ embryo in toto was frozen in OCT medium and stored at −80 °C until sectioning. Optimization of tissue permeabilization was performed on 10-µm-thick sections using Visium Spatial Tissue Optimization Reagents Kit (10x Genomics), which established an optimal permeabilization time of 30 min. The samples were mounted onto a Gene Expression slide (10x Genomics), fixed in ice-cold methanol, stained with H&E and scanned using the Leica Aperio Versa 200 scanner (Leica Biosystems). Tissue permeabilization was performed to release the poly-A mRNA for capture by the poly(dT) primers precoated on the slide, which include an Illumina TruSeq Read, spatial barcode and UMI. The Visium Spatial Gene Expression Reagent Kit (10x Genomics) was used for reverse transcription to produce spatially barcoded full-length cDNA and for second-strand synthesis followed by denaturation to allow the transfer of the cDNA from the slide into a tube for amplification and library construction. The Visium Spatial Single Cell 3′ Gene Expression libraries consisting of P5, P7, i7 and i5 sample indexes and TruSeq Read 2 were generated by end repair, A-tailing, adaptor ligation and sample index PCR. The Dual Index Kit TT Set A (10x Genomics) was used to add unique i7 and i5 sample indexes and sequencing was performed using the Illumina NovaSeq 6000 system.

After sequencing, reads were aligned to the human genome (hg38), and the expression matrix was extracted using the spaceranger pipeline. The Loupe files generated by the spaceranger pipeline were used to visualize and plot the expression of the genes in individual sections. In Loupe browser, the total UMI counts for each spot-associated barcode are normalized towards the grand median UMI counts per spot by a scaling factor (computed as median_UMI_counts_per_barcode/UMI_counts_per_barcode). The matrix is next log-transformed, then mean-centred and scaled per gene such that the mean is 0 and the s.d. is 1. To identify genes enriched in specific areas of the sections, Seurat was used to analyse the expression matrix. All spots and genes that were detected in at least one spot were used in the analysis. Specifically, the SCTransform function was used to scale the data and find variable genes using the default parameters. PCA was performed using the RunPCA function, and the first 30 principal components were used. UMAP was generated using the RunUMAP function. Spot clusters were obtained using the FindNeighbors and FindClusters function with the resolution set to 1.2. The cluster markers were obtained using the FindAllMarkers function (Supplementary Table 7). Cell-type deconvolution analysis for the Visium sections was not appropriate due to lack of representation of all the cell types in the scRNA-seq dataset. The tissues represented in individual clusters were inferred using the Enrichr database[52].

Immunofluorescence

Tissues were washed and fixed using 4% paraformaldehyde and embedded in paraffin using a Shandon Citadel 1000 or STP120 (Thermo Fisher Scientific). All paraffin-embedded tissues were sectioned (3 µm sections) using a microtome HM340E (Thermo Fisher Scientific). To identify anatomical regions, H&E staining was performed every 20 slides. Before staining, antigen retrieval was performed consecutively in Tris-EDTA (pH 9.0) and citrate buffer (pH 6.0) in a steam cooker. For the intracellular antigens, the sections were treated with 1% Triton X-100. A serum block solution (serum type depending on secondary antibodies, either goat serum or donkey serum) was used to block unspecific binding sites. The primary antibodies rabbit anti-human PECAM-1/CD31 (Novus Biologicals, NB100-2284, 1:200, 2), mouse anti-human ALDH1A1 (SantaCruz Biotechnology, sc-374149, 1:200, L1719), mouse anti-human IL-33 (Santa Cruz Biotechnology, sc-517600, 1:50, H2720), CXCR4 (Novus Biologicals, NB100-715, 1:100, MCX4-1019, MCX4-0920), mouse anti-human KCNK17 (SantaCruz Biotechnology, sc-390435, 1:100, D0113), rabbit anti-human SPINK2 (Sigma-Aldrich, HPA026813, 1:100, B118804), mouse anti-human CD45 (Vector Laboratories, VP-V354CE, 1:50, 6009341) were diluted in antibody dilution buffer (PBS containing 1% BSA, 0.1% Triton X-100, 0.1% cold water fish skin gelatin, 0.05% Tween-20) and the samples were incubated overnight at 4 °C. After several washes with washing buffer (DPBS containing 0.05% Tween-20), the secondary antibody goat anti-rabbit IgG (Life Technologies, 1:250, A11034, AF488, 2069632; 1:250, A11037, AF594, 2079421; or 1:200, A21245, AF647, 2098544), goat anti-mouse IgG2a (Life Technologies, 1:250, A21131, AF488, 1964395; 1:250, A21135, AF594, 1163392; or 1:200, A21241, AF647, 2056280), goat anti-mouse IgG1 (Life Technologies, 1:250, A21121, AF488, 1964382; 1:250, A21125, AF594, 2306794; or 1:200, A21240, AF647, 2012512), donkey anti-goat IgG (Life Technologies, 1:250, A11055, AF488, 830720) was applied to the samples and incubated for 30 min at room temperature and, after several washes, the sections were incubated with directly labelled antibodies for 1 h at

room temperature. After several washes with DPBS, the sections were incubated with TrueVIEW autofluorescence quencher (Vector Laboratories) for 5 min at room temperature. After two washes, the sections were exposed to a DAPI solution (2 μg ml^{-1} in DPBS, Roche) and then mounted (Prolong Gold Antifade Mounting Medium, Thermo Fisher Scientific). Fluorescence images were acquired using a confocal laser scanning microscope (LSM 880 with Airyscan, Carl Zeiss Microscopy) and a fluorescence microscope (Axio Observer Z1, Carl Zeiss Microscopy). ZEISS ZEN v.2.3 (black) ZEN v.3.1 (blue) were used for acquisition and processing of bright-field and immunofluorescence images.

PS cell differentiation

RM3.5-induced human iPS cells, constitutively expressing a *tdTOMATO* transgene from the *GAPDH* locus, were derived from human foreskin fibroblasts purchased from ATCC and reprogrammed using the hSTEMCCAloxP four-factor lentiviral vector, which was excised after reprogramming[53]. Human PS cell lines were maintained on Essential 8 medium (Thermo Fisher Scientific). Haematopoietic differentiation was performed using the swirler embryoid body method as described previously[49]. In brief, cells were dissociated using Accutase cell dissociation reagent (Thermo Fisher Scientific) and resuspended in STAPEL differentiation medium with minor modifications[54]. The cells were transferred to non-tissue-culture-treated 6-cm dishes in 5 ml of medium per dish. The dishes were then placed onto a digital orbital shaker (Heathrow Scientific) rotating at 60 rpm in an incubator under 5% CO_2 at 37 °C. Mesoderm was induced on day 1 of differentiation by a combination of 4 μM CHIR99021 (Tocris Biosciences), 0 or 3 ng ml^{-1} recombinant human bone morphogenetic protein 4 (BMP4, R&D Systems), 5 or 30 ng ml^{-1} recombinant human ACTIVIN A (R&D Systems) and 20 ng ml^{-1} recombinant human fibroblast growth factor FGF2 (PeproTech) and patterned with additional 3 μM CHIR99021and 3 μM SB431542 (Cayman Chemicals), 25 ng ml^{-1} recombinant human vascular endothelial growth factor (VEGF, PeproTech), 25 ng ml^{-1} recombinant human stem cell factor (SCF, PeproTech) and 20 ng ml^{-1} recombinant human FGF2 on days 2 and 3, as described previously[33]. After 3 days, the medium was supplemented with 50 ng ml^{-1} recombinant human VEGF, 20 ng ml^{-1} BMP4, 10 ng ml^{-1} FGF2, 50 ng ml^{-1} recombinant human SCF and 10 ng ml^{-1}l recombinant human insulin-like growth factor 2 (IGF2, PeproTech). After 7–8 days of differentiation, growth factors were modified to include 50 ng ml^{-1} recombinant human VEGF, 50 ng ml^{-1} recombinant human SCF, 50 ng ml^{-1} recombinant human thrombopoietin (TPO, PeproTech), 10 ng ml^{-1} recombinant human FLT3 receptor ligand (FLT3L, PeproTech), 10 ng ml^{-1} IL3 (PeproTech), 10 ng ml^{-1} APELIN peptide (Sigma-Aldrich), 10 ng ml^{-1} FGF2 and 20 nM SR1 (Stemregenin1) (Selleck Chemical). Medium was refreshed every 2 days during the differentiation. Blood cells were shed into the medium after 10–12 days of differentiation. After 14 days, cultures were collected. Cells shed into the medium (denoted suspension haematopoietic cells) were analysed separately from cells dissociated from the swirler embryoid bodies (denoted adherent embryoid body cells). Swirler embryoid bodies were disaggregated by 45 min incubation with collagenase I (Worthington) at 37 °C.

iPS cell type prediction using ACTINN

The cells for these haemato-vascular cell types were selected as the reference training data for ACTINN[55]: AGM_CS10_EC, AGM_CS10_HE, AGM_CS10_HSPC, AGM_CS14-15_VE, AGM_CS14-15_AE/Other, AGM_CS14-15_AE/preHE, AGM_CS14-15_HE, AGM_CS14-15_HSC, YS_CS11_EC, YS_CS11_HE, YS_CS11_HSPC, YS_CS14_EC, PL_CS14_HSC, YS_CS14_HSC, Liv_CS14_SPINK2_HPC, Liv_6wk_HSC, Liv_8wk_HSC, Liv_11wk_HSC, Liv_15wk_HSC, Cord blood_HSC, EryPri, EryProg, EryMegMastProg, Gran, MonoMac, LymphB and LymphT. ACTINN prediction was run using the default parameters to obtain the prediction cell type and the prediction scores for each cell. The prediction scores ranged from 0 to 1 and were visualized in UMAP. The expression matrix for the reference training data and the cell type annotation of the cells are accessible in GitHub (https://github.com/mikkolalab/Human-HSC-Ontogeny).

Reporting summary

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

Data availability

Sequencing data supporting the findings of this study have been deposited at the Gene Expression Omnibus (GEO) under accession code GSE162950. Data from published references are available at the GEO under accession code GSE135202. An interface for data browsing and links to data are also available online (http://singlecell.mcdb.ucla.edu/Human-HSC-Ontogeny). There is no restriction in data availability.

Code availability

Custom code, R objects and metadata of these R objects are available at GitHub (https://github.com/mikkolalab/Human-HSC-Ontogeny).

51. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729 (2018).
52. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
53. Kao, T. et al. GAPTrap: a simple expression system for pluripotent stem cells and their derivatives. *Stem Cell Rep.* **7**, 518–526 (2016).
54. Nafria, M., Bonifer, C., Stanley, E. G., Ng, E. S. & Elefanty, A. G. Protocol for the generation of definitive hematopoietic progenitors from human pluripotent stem cells. *STAR Protoc.* **1**, 100130 (2020).
55. Ma, F. & Pellegrini, M. ACTINN: automated identification of cell types in single cell RNA sequencing. *Bioinformatics* **36**, 533–538 (2020).

Acknowledgements We thank the staff at the BSCRC FACS and sequencing cores, and TPCL, TCGB and CFAR cores (NIH AI028697-21) at UCLA; A. Pyle and H. Xi for assistance with tissue procurement and staging; F. Lay for the introduction to single-cell analysis and S. Morrison for reading the manuscript; J. Rodgers and M. Giannoni for assistance with the website generation. This work was supported by Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA Interim Research Award and Innovation awards and a Jonsson Cancer Center Foundation and UCLA David Geffen School of Medicine Regenerative Medicine Theme Award (to H.K.A.M.); NIH 1RO1DK125097-01, 1R01DK121557-01 and 5R01DK100959-07 (to H.K.A.M.), NIH RO1HL148714 (to R.A.) and NIH R35HL140014 (to M.L.I.A); a Swedish Research Council International Postdoc grant IPD2 2018-06635 (to S.C.-G.); the Swiss National Science Foundation (P2ZHP3_178113 and EMBO ALTF 433-2019, to J.A.-G.); BSCRC post-doctoral fellowships (to J.A.-G. and I.F.), and T32 HL-086345-13 Developmental Hematology fellowship (to I.F. and B.N.), BSCRC Rose Hills Foundation Graduate Training Program and Ruth L. Kirschstein National Research Service Award T32HL069766 (to A.V.) and the Deutsche Forschungsgemeinschaft Cluster of Excellence iFIT (EXC 2180-390900677) (to K.S.-L.). A.G.E., E.G.S. and E.S.N. were supported by NHMRC (Australia) fellowships (GNT1117596, to A.G.E.; GNT1079004 to E.G.S.) and project grants (GNT1068866 and GNT1129861 to A.G.E. and E.G.S.; GNT1164577 to E.S.N.), by the ARC (Stem Cells Australia), and by the Stafford Fox Medical Research Foundation. Infrastructure funding was provided by NHMRC and Victorian government Infrastructure Support Programs.

Author contributions V.C., S.C.-G. and H.K.A.M. designed experiments and interpreted data. F.M. led the bioinformatics analysis, which was also performed by V.C. and S.C.-G. and assisted by I.F., S.E., B.N. and J.Y.L.; V.C. and S.C.-G. performed and/or supervised wet laboratory experiments and related data analysis, assisted by I.F. and J.A.-G.; S.L. coordinated the tissue collection and procurement and performed the immunofluorescence experiments. A.V. assisted with immunofluorescence data analysis. E.S.N., E.G.S., J.Y.L. and A.G.E. generated the hPS cell in vitro differentiation data. D.L. established the web interface for data mining. R.A., M.L.I.-A., L.K.L., M.P. and K.S.-L. assisted with data interpretation and contextualization. V.C., S.C.-G. and H.K.A.M. wrote the manuscript, which all of the authors edited and approved.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04571-x.

Correspondence and requests for materials should be addressed to Vincenzo Calvanese or Hanna K. A. Mikkola.

Peer review information *Nature* thanks Georges Lacaud, Nathan Salomonis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.