



LUND UNIVERSITY

The fetal hematopoietic progenitor compartment and the initiation of childhood leukemia

Palo, Sara

2026

Document Version:

Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):

Palo, S. (2026). *The fetal hematopoietic progenitor compartment and the initiation of childhood leukemia*. [Doctoral Thesis (compilation), Department of Laboratory Medicine]. Lund University, Faculty of Medicine.

Total number of authors:

1

Creative Commons License:

CC BY-NC

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



The fetal hematopoietic progenitor compartment and the initiation of childhood leukemia

SARA PALO

DEPARTMENT OF LABORATORY MEDICINE | FACULTY OF MEDICINE | LUND UNIVERSITY





FACULTY OF MEDICINE

Department of Laboratory Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2026:142

ISBN 978-91-8021-941-9

ISSN 1652-8220



The fetal hematopoietic progenitor compartment and the
initiation of childhood leukemia

The fetal hematopoietic progenitor compartment and the initiation of childhood leukemia

Sara Palo



LUND
UNIVERSITY

Doctoral Dissertation

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Medicine at Lund University, to be publicly defended on November 5th, 2026, at 9:00 in Belfrage Hall, Klinikgatan 32, Lund

Faculty opponent: Dr. Vincenzo Calvanese,
Josep Carreras Leukaemia Research Institute

Organization LUND UNIVERSITY Department of Laboratory Medicine Sölvegatan 19 SE-221 84 LUND Sweden		Document name DOCTORAL DISSERTATION	
		Date of disputation 2026-11-05	
Author(s) Sara Palo		Sponsoring organization	
Title and subtitle The fetal hematopoietic progenitor compartment and the initiation of childhood leukemia			
Abstract The human hematopoietic system is established over several overlapping extra- and intraembryonic waves through the formation of transient progenitor cells and definitive hematopoietic stem cells (HSCs). Findings of B cell acute lymphoblastic leukemia (B-ALL) driver mutations at birth suggests fetal development is also a time of vulnerability to formation of certain preleukemic clones. Properties specific to fetal hematopoietic stem and progenitor cells (HSPCs), such as lineage priming or plasticity, or transient progenitor populations, may play a role in developmental susceptibility but have not been fully explored. In <i>Paper I</i>, we simultaneously profiled the single-cell transcriptome and surface marker expression of CD34⁺ HSPCs from first-trimester human fetal liver (FL) and compared them to an adult bone marrow (BM) dataset. Analyses uncovered an erythro-myeloid primed multipotent progenitor population in the FL that was nearly absent from adult BM, possibly constituting a transient progenitor population, and a fetal-specific gene expression signature that was enriched in the youngest patients with KMT2A::AFF1 ALL compared to older pediatric patients. In <i>Paper II</i>, we showed through multilineage-promoting single-cell cultures that FL HSCs were to a larger extent multilineage in output, while adult BM HSCs were mainly unilineage and myeloid-biased. Profiling the chromatin accessibility of first-trimester FL HSPCs by single-cell ATAC-seq, we found that FL HSCs did not exhibit clear indications of chromatin-level lineage priming, while adult BM HSCs showed signs of myeloid and megakaryocyte priming. We further found indications of lower activity of the B-lineage master regulator PAX5 in FL lymphoid progenitors compared to adult, evident through reduced accessibility of the PAX5 motif and putative regulatory elements. In <i>Paper III</i>, a murine model with inducible expression of KMT2A::AFF1 showed that fusion gene expression could result in a CD24⁺ preproB population with preleukemic properties. This population only emerged when expression was induced at the HSPC stage during embryonic development, indicating that developmental context is highly relevant to transformation. In <i>Paper IV</i>, we presented a model where KMT2A::AFF1 is introduced through CRISPR/Cas9 in induced pluripotent stem cell-derived HSPCs, thus recapitulating features of the human prenatal setting. Upon transplantation, edited cells resulted in aggressive B-ALL with features characteristic of infant disease. In all, these works provide new insights into the priming and plasticity of human fetal HSPCs and early events in the leukemogenic process.			
Key words Developmental hematopoiesis, hematopoietic stem cells, single-cell genomics, acute lymphoblastic leukemia, induced pluripotent stem cells			
Classification system and/or index terms (if any)			
Supplementary bibliographical information		Language English	
ISSN and key title 1652-8220		ISBN 978-91-8021-941-9	
Recipient's notes		Number of pages 85	Price
		Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources the permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature _____ Date 2026-09-25

The fetal hematopoietic progenitor compartment and the initiation of childhood leukemia

Sara Palo



LUND
UNIVERSITY

Cover illustration © Sara Palo. Embroidery on cotton fabric.

Primitive hematopoietic stem and progenitor cells, intermingled with lineage-committed progenitors and mature cells, make up the emerging hematopoietic landscape. In the case of one lymphoid progenitor cell, aberrant genetics and subsequent transcriptional and epigenetic dysregulation causes it to proliferate uncontrollably in a malignant fashion.

Pages i-65 © 2026 Sara Palo

Paper I © 2023 American Society of Hematology. Published by Elsevier Inc (CC BY-NC-ND 4.0).

Paper II © 2026 American Society of Hematology. Published by Elsevier Inc (CC BY-NC-ND 4.0).

Paper III © 2025 The authors. (CC BY 4.0).

Paper IV © 2026 The authors. (Unpublished manuscript).

Faculty of Medicine

Department of Laboratory Medicine

ISBN: 978-91-8021-941-9

ISSN: 1652-8220

Lund University, Faculty of Medicine Doctoral Dissertation Series 2026:142

Printed in Sweden by Media-Tryck, Lund University, Lund 2026



Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN 

*To my grandmother, in case I never win that Nobel prize I
promised to dedicate to her.*

*It's hard to succinctly describe how ggplot2 works because it embodies a deep
philosophy of visualisation.
- The ggplot2 website*

Table of Contents

List of publications	x
Abstract	xi
Popular Science summary	xii
Populärvetenskaplig sammanfattning	xv
List of abbreviations	xviii
Background	1
The hematopoietic hierarchy	1
Development of the hematopoietic system	3
Regulation of lineage choice	4
Lineage priming and output before birth and in adulthood	8
B cell development	9
Pediatric acute lymphoblastic leukemia	10
KMT2A rearrangements	11
Models of KMT2A::AFF1 leukemia	13
Methodology	15
Assessing the output of HSPCs	15
Investigating molecular properties of HSPCs	18
Processing and visualizing single cell omics data	21
Dataset annotation	24
Hematopoietic differentiation of iPSCs	26
Ethical considerations	27
Aims of the thesis	29
Author's contribution to included papers	30
Summary of included papers	31
Paper I	31
Paper II	33
Paper III	35
Paper IV	37
Discussion	39
Future perspectives	44
Acknowledgements	45
References	48

List of publications

Papers included in the thesis:

- I **Single-cell multiomics of human fetal hematopoiesis define a developmental-specific population and a fetal signature**
Sommarin MNE, Olofzon R, **Palo S**, Dhapola P, Soneji S, Karlsson G, Böiers C
Blood Advances 2023; 7(18): 5325–5340.
- II **Chromatin Accessibility Shapes Developmental-Specific Lineage Plasticity in Hematopoiesis**
Palo S, Nagaharu K, Sommarin MNE, Olofzon R, Turati V, Soneji S, Karlsson G, Böiers C
Blood Advances 2026; in press.
- III **Ontogeny-specific induction of the KMT2A::AFF1-fusion drives development of a distinct CD24 positive pre-leukemic state**
Calderón AS, Ghazanfari R, Masoumi Z, Kharazi S, **Palo S**, Lang S, Žemaitis K, Eldeeb M, Subramaniam A, Soneji S, Stam RW, Bryder D, Böiers C
Leukemia 2025; 39: 2099–2111.
- IV **A human iPSC model of KMT2A::AFF1 captures the developmental trajectory of pediatric acute B-cell leukemia**
Nagaharu K*, **Palo S***, Calderón AS, González Antón S, Ghazanfari R, Geir-
onson Ulfsson L, Bourguine P, Böiers C
Unpublished manuscript. *Equal contribution.

Abstract

The human hematopoietic system is established over several overlapping extra- and intraembryonic waves through the formation of transient progenitor cells and definitive hematopoietic stem cells (HSCs). Findings of B cell acute lymphoblastic leukemia (B-ALL) driver mutations at birth suggests fetal development is also a time of vulnerability to formation of certain preleukemic clones. Properties specific to fetal hematopoietic stem and progenitor cells (HSPCs), such as lineage priming or plasticity, or transient progenitor populations, may play a role in developmental susceptibility but have not been fully explored. In *Paper I*, we simultaneously profiled the single-cell transcriptome and surface marker expression of CD34⁺ HSPCs from first-trimester human fetal liver (FL) and compared them to an adult bone marrow (BM) dataset. Analyses uncovered an erythro-myeloid primed multipotent progenitor population in the FL that was nearly absent from adult BM, possibly constituting a transient progenitor population, and a fetal-specific gene expression signature that was enriched in the youngest patients with KMT2A::AFF1 ALL compared to older pediatric patients. In *Paper II*, we showed through multilineage-promoting single-cell cultures that FL HSCs were to a larger extent multilineage in output, while adult BM HSCs were mainly unilineage and myeloid-biased. Profiling the chromatin accessibility of first-trimester FL HSPCs by single-cell ATAC-seq, we found that FL HSCs did not exhibit clear indications of chromatin-level lineage priming, while adult BM HSCs showed signs of myeloid and megakaryocyte priming. We further found indications of lower activity of the B-lineage master regulator PAX5 in FL lymphoid progenitors compared to adult, evident through reduced accessibility of the PAX5 motif and putative regulatory elements. In *Paper III*, a murine model with inducible expression of KMT2A::AFF1 showed that fusion gene expression could result in a CD24⁺ preproB population with preleukemic properties. This population only emerged when expression was induced at the HSPC stage during embryonic development, indicating that developmental context is highly relevant to transformation. In *Paper IV*, we presented a model where KMT2A::AFF1 is introduced through CRISPR/Cas9 in induced pluripotent stem cell-derived HSPCs, thus recapitulating features of the human prenatal setting. Upon transplantation, edited cells resulted in aggressive B-ALL with features characteristic of infant disease. In all, these works provide new insights into the priming and plasticity of human fetal HSPCs and early events in the leukemogenic process.

Popular Science summary

Our blood system is made up of oxygen-transporting red blood cells, blood-clotting platelets, and the immune cells, which protect against infection. These include the innate immune cells, primarily myeloid cells, which provide protection through more general mechanisms, and the adaptive B and T lymphoid cells, which learn to recognize pathogens and infected cells in order to better combat them in the future. The blood system is replenished with billions of new cells every day by the hematopoietic stem cells (HSCs) in the bone marrow. HSCs possess the properties of self-renewal, meaning that they can divide to give rise to new HSCs, and multipotency, meaning that they are able to produce all types of mature, functional blood cells. They do this by giving rise to progenitor cells, which in turn give rise to new progenitor cells, in a process referred to as differentiation. The progenitors produced become more and more directed, or "primed" towards a certain cell type, or lineage, and eventually lose the ability to give rise to other lineages, referred to as lineage restriction. This process is regulated by a complex network of signaling molecules, proteins called transcription factors which regulate the gene expression in the cell, and changes in the structure of the tightly packed DNA, making certain genes more or less available. Despite multipotency being a defining HSC characteristic, HSCs are not all equal, but can exhibit bias towards a certain lineage.

The HSCs in our body all stem from a pool of HSCs produced before birth. However, the development of the blood system does not begin with HSCs: in humans, primitive blood progenitor cells start to emerge from the yolk sac around 2.5 weeks after conception. HSCs begin to form in a blood vessel inside the embryo around 4 weeks post-conception and migrate to the fetal liver, which is the main blood forming organ during most of the first trimester. At the start of the second trimester, HSCs begin to move to the bone marrow, although the FL remains an active site of blood formation. Previous research has shown that fetal blood differs from that in the adult in terms of proportions of different cell types and gene expression, but a lot is still unknown, particularly about blood formation during the first trimester.

Fetal development is also a time period that appears vulnerable to certain mutations that can give rise to acute lymphoblastic leukemia (ALL), one of the most common forms of cancer in children. In ALL, lymphoid progenitor cells, most often of the B cell lineage, have acquired mutations that lead to unchecked growth of non-functioning cells, disrupting the normal functions of the bone marrow. Mutations known to drive several different types of childhood ALL have been shown to already be present in the blood of some children at birth. Although most of these mutations appear to require additional events in order to transform into a leukemia, and childhood ALL generally has a very good prognosis with over 90% survival rate, certain mutations, such as KMT2A::AFF1,

are known to cause a rare, very aggressive and poor-prognosis form of ALL in infants younger than 1 year. How exactly this disease begins is unknown, but it is highly likely that the fetal context where the mutation occurs plays a role. In the works presented in this thesis, we have investigated the properties of normal human fetal blood formation and how it differs from that in the adult, and studied the effects of the KMT2A::AFF1 mutation on blood cells in fetal and adult settings and at different stages of differentiation.

In Paper I, we profiled the gene expression of individual HSCs and progenitor cells from first-trimester human fetal liver and compared the data to corresponding cell populations in the adult bone marrow. Doing so, we found a population of progenitor cells, seemingly primed towards the red blood cell, platelet and myeloid lineages, that was nearly absent from the adult data, thus potentially representing cells from an early, pre-HSC wave of blood formation. We also extracted a list of genes that were more highly expressed in fetal blood progenitors than in adult, and showed that this fetal signature was enriched in the youngest patients with KMT2A::AFF1 ALL compared to older children.

In Paper II, we assessed differences in lineage priming and output between first-trimester fetal and adult bone marrow blood stem and progenitor cells. Through individual cultures of single HSCs, we showed that fetal HSCs are more flexible in their lineage output, more often giving rise to cells of multiple different lineages. Furthermore, while single-cell gene expression can provide a lot of information about the cell, it can be difficult to assess lineage priming due to the sometimes low expression of lineage-defining transcription factors. However, as a cell differentiates, it also changes the accessibility of certain regions of its genome. We therefore investigated the accessibility of the genome in fetal liver cells, finding that fetal HSCs did not appear to be strongly primed towards any specific lineage, while adult HSCs showed signs of priming towards myeloid cells and platelets. We also showed that fetal lymphoid progenitor cells appear less restricted to the B cell lineage, an intriguing flexibility that we proposed may play a role in the initiation of childhood ALL.

In addition to studying normal blood formation during development, we also studied the effect of the KMT2A::AFF1 mutation. In Paper III, we presented a mouse model where expression of KMT2A::AFF1 could be activated either in blood stem cells and very immature progenitors, or in progenitors at different steps of B cell commitment. We showed that the gene has different effects depending on the cell type it first becomes active in and when activated in early development compared to in adulthood. Most importantly, activating the gene specifically in stem and progenitor cells during fetal development lead to the emergence of a B cell progenitor with abnormal features, such as HSC-like self-renewal ability and the ability to give rise to other lineages than B cells. We proposed that these cells represent a pre-leukemic stage in leukemia development.

Genes expressed more highly in these cells were also found to be expressed in certain patients with KMT2A::AFF1 leukemia, showing that the model can provide insights that are applicable to the disease in humans.

Finally, in Paper IV, we developed a model system to study the mutation in a human setting. Induced pluripotent stem cells are cells from, e.g., skin, which have been re-programmed to pluripotent stem cells similar to those in the very early stages of the embryo, able to give rise to all different tissues in the body. By culturing these cells in conditions that differentiate them to blood, cells similar to those found in fetal blood development are produced. Using the "genetic scissor" CRISPR/Cas9 to give the cells the KMT2A::AFF1 mutation, we showed that this model could produce B cell progenitors with expression of genes that are also expressed in patients. Furthermore, when these mutated cells were transplanted into mice, they resulted in an aggressive B cell ALL similar to the one seen in infant patients. These results established our model system as a promising platform for future studies of the disease in a setting that emulates the one in which the mutation arises in humans.

Populärvetenskaplig sammanfattning

Blodsystemet består av syretransporterande röda blodkroppar, blodplättar som stoppar blödningar och immuncellerna, som skyddar mot infektion. Till dessa räknas det medfödda immunsystemet, som främst består av myeloida celler och skyddar genom mer generella mekanismer, och de adaptiva, lymfoida B och T-cellerna, som lär sig att känna igen patogener och infekterade celler för att bekämpa dem mer effektivt i framtiden. Blodet fylls på med miljarder nya celler varje dag tack vare de hematopoietiska stamcellerna (HSCs), som finns i benmärgen. HSCs har egenskaperna självförnyelse, d.v.s. att de kan dela sig för att ge upphov till nya HSCs, och multipotens, d.v.s. förmågan att bilda alla typer av mogna blodceller. De gör detta genom att ge upphov till progenitorceller, som i sin tur ger upphov till nya blodceller, en process som kallas differentiering. Progenitorerna blir allt mer inriktade på en identitet som en specifik celltyp, och förlorar så småningom förmågan att bilda andra typer av celler, och är därmed låsta till en specifik identitet. Denna process regleras av ett komplext nätverk av signalmolekyler, proteiner som kallas för transkriptionsfaktorer som reglerar genuttrycket i cellen, samt förändringar i det tätt packade DNAs struktur som gör olika gener mer eller mindre tillgängliga för cellen. Trots att multipotens är ett av de utmärkande dragen för en HSC är inte alla HSCs lika, utan de kan uppvisa en preferens i vilka typer av celler de bildar.

Alla HSCs i kroppen har sitt ursprung i en mängd HSCs som bildades före födseln. Utvecklingen av blodsystemet börjar dock inte där: hos människor bildas de första primitiva blodprogenitorcellerna i gulesäcken ca 2.5 veckor efter befruktningen. HSCs börjar bildas i ett blodkärl i embryot ca 4 veckor efter befruktning och migrerar sedan till levern, som är det huvudsakliga blodbildande organet under större delen av första trimestern. I början av andra trimestern migrerar HSCs slutligen till benmärgen, även om levern fortsätter att vara aktiv i blodbildningen. Tidigare forskning har visat att fetalt blod skiljer sig från det hos vuxna sett till proportioner av olika celltyper och genuttryck, men mycket är fortfarande okänt, särskilt när det gäller första trimestern.

Fosterutvecklingen är också en period som verkar vara mottaglig för vissa mutationer som kan ge upphov till akut lymfatisk leukemi (ALL), som är en av de vanligaste formerna av cancer hos barn. I ALL förökar sig lymfoida icke-fungerande progenitorceller, oftast av B-cellstyp, okontrollerat och förhindrar benmärgens normala funktioner. Flera mutationer som kan driva ALL hos barn har visat sig redan finnas i blodet hos vissa barn vid födseln. Även om de flesta av dessa verkar kräva ytterligare mutationer för att omvandlas till en leukemi, och ALL hos barn generellt har en god prognos med >90% överlevnad, kan vissa mutationer, så som KMT2A::AFF1 orsaka en ovanlig, väldigt aggressiv form av ALL med dålig prognos hos spädbarn yngre än 1 år. Exakt hur sjukdomen uppstår är okänt, men det är sannolikt att den fetala kontexten där mutationen uppstår spelar en roll.

I arbetena som presenteras i denna avhandling har vi undersökt egenskaperna hos normal human blodbildning och hur den skiljer sig från hos en vuxen, och studerat effekterna av KMT2A::AFF1 på blodceller i en fetal respektive vuxen kontext samt i olika stadier av celldifferentiering.

I artikel I profilerade vi genuttrycket hos individuella HSCs och progenitorceller från fetal lever från första trimestern och jämförde vår data med motsvarande cellpopulationer från vuxen benmärg. Genom detta fann vi en population progenitorceller som verkade inriktade på att bilda röda blodceller, blodplättar och myeloida celler, som nästan helt saknades i datan från vuxna, och därmed kan representera celler från en tidig våg av blodbildning som föregår bildandet av HSCs. Vi tog också fram en lista på gener som var mer uttryckta i fetala celler än vuxna, och visade att denna fetala gensignatur var anrikad hos de yngsta patienterna med KMT2A::AFF1-ALL jämfört med äldre barn med sjukdomen.

I artikel II fokuserade vi på skillnader i cellers inriktning på specifika identiteter samt produktionen av olika celltyper mellan första trimestern och vuxen benmärg. Genom att odla HSCs individuellt kunde vi visa att fetala HSCs var mer flexibla i sina val, då de oftare producerade flera olika typer av celler från en enda HSC. Vidare, medan genuttryck i individuella celler kan ge mycket information kan det vara svårt att studera cellers inriktning på specifika identiteter då viktiga transkriptionsfaktorer kan vara uttryckta till en låg grad. Men när celler differentierar ändras också tillgängligheten av vissa delar av genomet. Vi studerade därför tillgängligheten av genomet i fetala blodceller och fann att fetala HSCs inte visade tecken på att ha valt någon specifik celltyp att bilda, medan vuxna HSCs visade tecken på att bilda myeloida celler och blodplättar. Vi visade även att fetala lymfoida progenitorer visade tecken på att de inte var lika låsta till att bilda B-celler, en flexibilitet som vi föreslår kan spela en roll i uppkomsten av ALL hos barn.

Utöver att studera normal blodbildning under fosterlivet studerade vi också effekten av KMT2A::AFF1. I artikel III använde vi en musmodell där vi kunde aktivera uttrycket av KMT2A::AFF1 i antingen HSCs och väldigt omogna progenitorer eller i progenitorer i olika steg av inriktning på en B-cellsidentitet. Vi visade att genen har olika effekt beroende på i vilken celltyp den aktiveras och beroende på om den aktiveras under fosterutvecklingen eller i vuxenlivet. Framförallt så ledde aktivering av genen specifikt i HSCs och omogna progenitorer under fosterutvecklingen till utvecklandet av en B-cellsprogenitor med onormala egenskaper, så som en HSC-lik förmåga till självförnyelse och förmågan att ge upphov till andra celltyper än B-celler. Vi föreslog att dessa celler motsvarar ett preleukemiskt förstadium i utvecklingen av ALL. Gener som uttrycktes mer i dessa celler var också mer uttryckta i vissa patienter med KMT2A::AFF1-ALL, vilket visade att modellen kan ge insikter som kan tillämpas på sjukdomen hos människor.

Slutligen utvecklade vi i artikel IV en sjukdomsmodell där mutationen kan studeras i en mänsklig kontext. Inducerade pluripotenta stamceller är celler från exempelvis hud, som har programmerats om till pluripotenta stamceller liknande de som finns i embryot i ett väldigt tidigt skede och som kan ge upphov till alla vävnadstyper i kroppen. Genom att odla dessa celler på ett sätt som driver dem till att differentiera till blod kan celler som liknar de som finns i blodbildningen hos foster bildas. Med hjälp av ”gensaxen” CRISPR/cas9 gav vi sedan dessa celler mutationen KMT2A::AFF1, och visade att vår modell kunde bilda B-cellsprogenitorer som uttryckte gener kopplade till KMT2A::AFF1-ALL. När dessa muterade celler transplanterades i möss gav de upphov till en aggressiv ALL som efterliknade flera karaktäristiska drag hos sjukdomen i spädbarn. Dessa resultat etablerar vår modell som en lovande plattform för framtida studier av sjukdomen i en kontext som efterliknar den i vilken mutationen uppstår hos människor.

List of abbreviations

AGM	Aorta-gonad-mesonephros
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
B-ALL	B-cell ALL
BCR	B cell receptor
BM	Bone marrow
CB	Cord blood
CITE-seq	Cellular indexing of transcriptomes and epitopes by sequencing
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CRISPR	Clustered regularly interspaced short palindromic repeats
CS	Carnegie stage
CSF	Colony-stimulating factor
DC	Dendritic cell
DE	Differentially expressed
E	Embryonic day
FACS	Fluorescence-activated cell sorting
FL	Fetal liver
GMP	Granulocyte-monocyte progenitor
hOSS	Humanized ossicle
HSC	Hematopoietic stem cell
HSPC	Hematopoietic stem and progenitor cell
Ig	Immunoglobulin
IL	Interleukin
IL7R	IL-7 receptor
iPSC	Induced pluripotent stem cell
KMT2A	Histone-lysine-methyltransferase 2A
KMT2Ar	KMT2A-rearranged
LIN	Lineage
LMPP	Lymphoid-primed multipotent progenitor
LSI	Latent semantic indexing
LSK	LIN ⁻ c-KIT ⁺ Sca-1 ⁺
Ly	Lymphoid
MEP	Megakaryocyte-erythroid progenitor
MHC	Major histocompatibility complex
MkE	Megakaryocyte/erythroid
MLP	Multilymphoid progenitor

MPP	Multipotent progenitor
NK cell	Natural killer cell
NSG	NOD-SCID-GAMMA
PCA	Principal component analysis
pcw	Post-conceptual weeks
pDC	Plasmacytoid dendritic cell
RNA-seq	RNA sequencing
SCF	Stem cell factor
ST-HSC	Short-term HSC
t-SNE	t-distributed stochastic neighbor embedding
TF	Transcription factor
TPO	Thrombopoietin
UMAP	Uniform manifold approximation and projection
YS	Yolk sac

Background

The hematopoietic hierarchy

Billions of new blood cells are produced every day to maintain the hematopoietic system, which is responsible for oxygen transport, wound healing and protection against infections. The hematopoietic cells can be divided into three main groups: the red blood cells, or erythrocytes, which transport oxygen through the body; the megakaryocytes and the platelets they produce, responsible for blood clotting; and the leukocytes, or immune cells. The leukocytes can be further subdivided into myeloid (to which the erythrocytes and megakaryocytes are often also counted) and lymphoid cells. The myeloid immune cells include monocytes, macrophages, granulocytes and dendritic cells (DCs), which fight infections in the form of innate immunity, through mechanisms such as phagocytosis and the release of signaling molecules to regulate inflammation and immune cell recruitment. The lymphoid cells include B and T cells, which constitute the adaptive immune system, as well as the innate natural killer (NK) cells.

With very few exceptions, the mature blood cells performing these functions all originate from the hematopoietic stem cells (HSCs), which in mammals reside in the bone marrow (BM). The defining features of the HSC are its ability to give rise to all mature hematopoietic cells in the body, i.e., multipotency, and its ability to give rise to new HSCs, i.e., self-renewal. In the early 1960s, transplantation experiments using irradiated mice showed that bone marrow cells could give rise to multilineage colonies in the spleen [1], that colonies could arise from a single cell [2], and that cells from the colonies could be transplanted into new mice and there give rise to new colonies [3], thus providing evidence for both. Three decades later, the isolation and transplantation of single HSCs into irradiated mice proved their ability to support long-term reconstitution of the hematopoietic system upon transplantation [4].

The classical text book model of hematopoiesis, largely based on studies in the murine system, typically shows the first step of lineage choice as a separation into two branches downstream of HSCs. After the HSC transitions into a short-term HSC (ST-HSC) or multipotent progenitor (MPP) with limited self-renewal capacity, the ability to produce B cells, T cells and NK cells resides in the common lymphoid progenitor (CLP), while all other lineages stem from the common myeloid progenitor (CMP), which further separates into a megakaryocyte-erythroid progenitor (MEP) and a granulocyte-monocyte progenitor (GMP) [5–7]. During the 2000s and early 2010s, discovery of a lymphoid-primed multipotent progenitor (LMPP) or multilymphoid progenitor (MLP), which retains myeloid capacity, showed that lymphoid and myeloid fates are more closely linked than previously thought, indicating that the first step in lineage commitment is rather segregation of the lymphoid fate from the megakaryocyte/erythroid (Figure 1) [6–8].

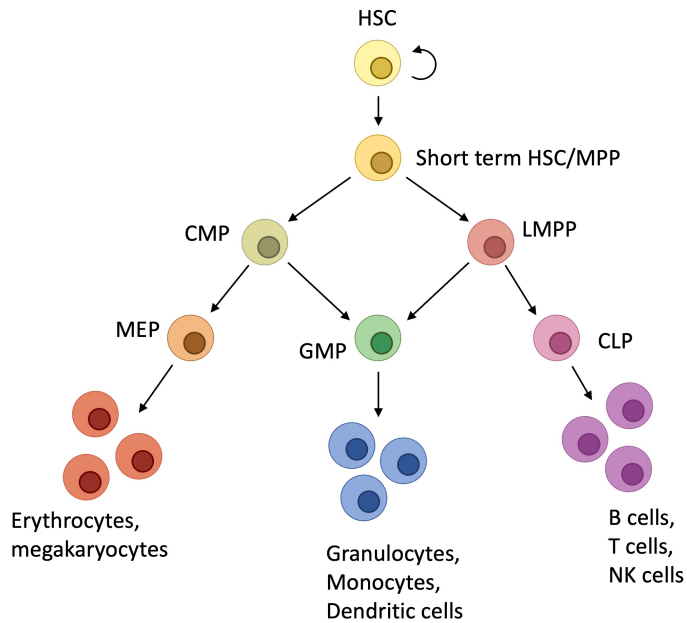


Figure 1: The hematopoietic hierarchy. Adapted from [6].

More detailed analyses have revealed heterogeneity within phenotypically defined populations [9, 10], different routes of neutrophil and eosinophil/basophil differentiation [11] and evidence of a direct branching towards the megakaryocyte fate from HSCs or MPPs, bypassing intermediate stages such as the CMP [12, 13]. Furthermore, with the advent of single-cell genomics, HSPCs have been shown to form a transcriptional continuum [14, 15], indicating that populations like the MPP or CMP do not represent discrete cell states. This has challenged the stepwise hierarchical structure and sparked discussions on whether the hematopoietic landscape should be regarded as a strict hierarchy with clear, stepwise commitment, as a continuous pool of low-primed progenitors from which unipotent progenitors emerge [14], or as something in-between, such as a punctuated continuum [16]. The heterogeneity of HSPCs extends to the HSC compartment, as phenotypical HSCs can vary substantially in both their self-renewal capacity and their bias towards generating certain hematopoietic lineages [17]. Thus, while a hierarchy like the one presented in Figure 1 is helpful, it can by no means be regarded as a definite and all-encompassing description of hematopoiesis.

Development of the hematopoietic system

The formation of the hematopoietic system takes place across a number of different tissues in the developing embryo, and later fetus, as hematopoietic cells emerge in temporally overlapping waves (Figure 2). During the first ~8 weeks of human development, which constitute the embryonic period, developmental stage is usually reported based on the 23 morphology-based Carnegie stages (CS). After the transition into the fetal period, development is usually measured in post-conceptual weeks (pcw).

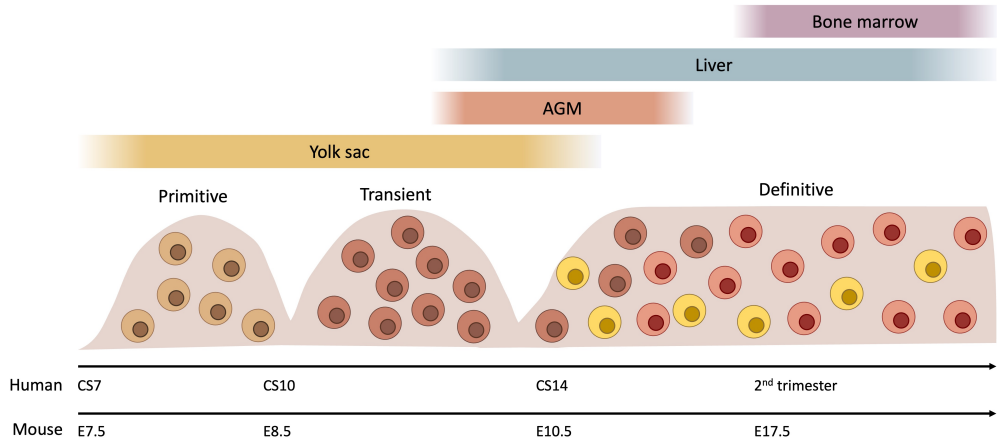


Figure 2: Timeline of hematopoiesis during human and murine development. Adapted from [18, 19].

In humans, the first hematopoietic cells appear in the yolk sac (YS) around 16-18.5 days post-conception (CS7-8) in the form of primitive erythroid, macrophage and megakaryocyte progenitor cells [20–22]. By day 21-23 (CS10), erythroblasts can be found in the cardiac cavity, suggesting that they enter circulation around the time of onset of cardiac contractions [23]. This primitive wave of hematopoiesis is followed by another HSC-independent wave of progenitor cells, which in the murine system largely consists of erythro-myeloid and lymphoid-primed progenitors which emerge from the YS and migrate to the fetal liver (FL) [24, 25]. While this wave is not as well-characterized in the human system, single-cell RNA sequencing (RNA-seq) studies have shown the presence of myeloid and innate lymphoid progenitor cells in the YS at CS10-11, which likely go on to seed the liver [26, 27].

The third, definitive wave of hematopoiesis is marked by the generation of HSCs, emerging from intra-aortic hematopoietic clusters of $CD34^+CD45^+$ cells which appear in the aorta-gonad-mesonephros (AGM) region around 27 days post-conception (CS13) and disappear by day 40 [23, 28]. HSCs, as defined by their ability to provide long-term multilineage engraftment upon transplantation into immunodeficient mice, have been detected in the AGM from CS14 and in the YS from CS16, suggesting that they

migrate there from the AGM [29]. The placenta is believed to be an additional site of HSC formation in the murine system [30]. Although phenotypic fetal HSCs, expressing CD90 and GPI-80, can be detected in the placenta at 5 weeks [31], before colonization of the FL, multilineage engraftment has not been reported before week 9 [32].

Cells with a gene expression profile matching nascent HSCs from the AGM can be detected in the FL from CS16 (~5.5 pcw) [33], but transplantable HSCs are very rare before 7-8 pcw [29, 34–36]. At this point in development, the liver is the main site of hematopoiesis, containing a mix of transient and definitive HSPCs. In the FL, HSCs mature, downregulating fetal-specific genes such as *LIN28B*, *HMG2* and *HOXB7-9*, and upregulating *PROM1* and major histocompatibility complex (MHC) class II genes [33]. In the developing bone marrow, the first hematopoietic cells observed are CD68⁺ myeloid cells, which appear prior to vascularization at around 8 weeks post-conception [37]. Transplantable HSCs are present in the BM from ~12 pcw [38], but the liver remains an important site of hematopoiesis throughout the second trimester.

Certain types of immune cells, including microglia, the tissue-resident macrophages of the brain, as well as subsets of innate-like tissue-resident T cells and B1a cells, are almost exclusively formed during fetal development and persist into adulthood [39–41]. Due to the limited possibilities to study human development, these cell types, as well as many other aspects of developmental hematopoiesis, have been more thoroughly investigated in the murine system. Murine hematopoiesis begins in the YS at embryonic day (E) 7, followed by a second wave of HSC-independent progenitors starting at E8.5. HSCs begin to emerge at E10.5 and colonize the FL by E12, before finally migrating to the BM at E17.5, only a few days before birth [18]. Hematopoietic progenitor cells formed alongside HSCs have been shown to be responsible for the majority of hematopoiesis in the FL, with only limited contribution from definitive, long-term HSCs before birth [42]. Whether this is the case in humans, who have a much longer gestation time, remains to be determined.

Regulation of lineage choice

The transition from HSC to mature hematopoietic cell occurs through a complex interplay of cell-extrinsic and cell-intrinsic signals (Figure 3). While the production of a protein requires that its corresponding gene is transcribed into RNA, the regulation of transcription is largely epigenetic.

Extrinsic signals

Cytokines are small proteins that function as important signaling molecules during maintenance and differentiation of hematopoietic cells. As different cytokines are involved in regulation of different lineages, many cytokine receptors are also used as surface markers to distinguish different cell populations.

Key factors in HSC maintenance include thrombopoietin (TPO) and stem cell factor (SCF) [43, 44]. TPO is mainly produced by hepatocytes, but both TPO and SCF are also produced in the BM niche [43, 45].

Colony-stimulating factors (CSFs), namely macrophage-CSF (M-CSF), granulocyte-CSF (G-CSF) and granulocyte-monocyte-CSF (GM-CSF), can be produced by multiple tissues in the body and are important for the maintenance, proliferation and activation of myeloid immune cells [46, 47]. Erythropoiesis is largely regulated by erythropoietin (EPO), which is produced in response to hypoxia in the fetal liver and postnatally mainly in the kidneys [48]. In addition to its role in regulation of HSCs, TPO is also an important factor in megakaryopoiesis [49].

Fms-like tyrosine kinase 3 ligand (FLT3L) and interleukin-7 (IL-7) are key factors in lymphoid differentiation [50]. Deletion of the *Flt3l* gene results in reduced numbers of B cell progenitors, although B cells are still present and appear normal in function [51], and mice lacking IL-7 are still able to produce the embryonic and early-life B1 cells [52]. However, mice deficient in both IL-7 and FLT3L are devoid of B cells [53]. IL-2 and IL-15 are both involved in NK cell regulation and development [54]. IL-7 is also involved in early steps of T cell development, which is mainly dependent on Notch signaling occurring in the thymus [55].

Transcription factors

At the molecular level, gene expression is regulated by transcription factors (TFs) which bind to gene promoters and enhancer elements to promote or suppress expression. TF binding sites, or motifs, are generally very short (6-12 bp) degenerate DNA sequences, and their binding is modulated through a number of different mechanisms [56].

Important TFs in human HSCs include HLF [57], MLLT3 [58], GATA2 [59] and MECOM (*EVII*) [60]. During fetal development, RUNX1 [61, 62] and medial *HOXA* genes such as *HOXA7* and *HOXA9* [63, 64] are critical in HSC specification. *HOXB* genes have been found to be expressed in hematopoietic progenitor cells preceding the definitive wave and in immature HSCs, but are later downregulated, while conversely, medial *HOXA* genes are absent in primitive hematopoiesis [33].

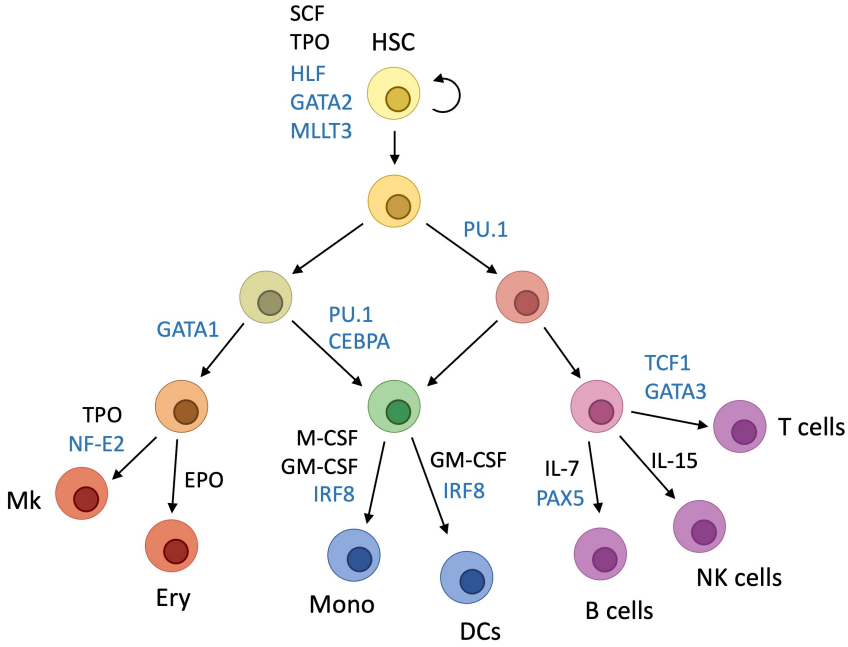


Figure 3: Examples of cytokines (black text) and TFs (blue text) involved in hematopoiesis. Ery = erythrocytes, Mk = megakaryocytes, Mono = monocytes.

GATA1 is a key player in differentiation to the erythroid and megakaryocyte lineages [5, 65], and ectopic expression is enough to convert lymphoid and myeloid progenitor cells to the megakaryocyte/erythroid (MkE) lineage [66]. Erythroid specification further involves KLF1 [67], while NF-E2 is more specific to megakaryocytes [68]. Commitment to the myeloid and lymphoid lineages generally involves upregulation of PU.1, encoded by the *SPI1* gene. GATA1 and PU.1 repress each other and upregulation of one or the other represents an early step in hematopoietic lineage commitment [5, 69, 70], although tracking of cell divisions in individual HSPCs suggests that the lineage choice is controlled by other factors upstream of changes in GATA1/PU.1 expression [71, 72].

Downstream of PU.1, C/EBP α directs progenitor cells towards the GMP state and suppresses lymphoid differentiation [73, 74], after which, among others, GATA2 is involved in the formation of basophils and mast cells and IRF8 in monocytes and dendritic cells [59, 69, 73]. B cell specification and commitment is largely orchestrated by Ikaros (*IKZF1*), E2A (*TCF3*), EBF1 and PAX5 [75, 76], while TCF1 (*TCF7*) and GATA3 are important factors in T cell development [77].

While TFs typically require the region around their motif to be open, or accessible, in order to bind, certain pioneer TFs such as PU.1 and PAX5, can promote chromatin

remodeling, opening up previously inaccessible sites and and permitting other TFs to bind [56].

Epigenetic control

In addition to TF binding, the transcription of a gene is regulated through its accessibility to the transcriptional machinery. The DNA in the cell is wound around histones to form chromatin, some of which contains few genes and is highly condensed (heterochromatin), and some which is overall more accessible for transcription (euchromatin).

Histone modifications play an important role in the regulation of gene expression and can act as either transcriptional activators or silencers. Sites of histone modification are often enhancer elements: short DNA sequences, located at a distance from the gene they regulate, that through recruitment of specific TFs can be brought into contact with the promoter region of a gene to regulate transcription [78]. While the mechanisms behind enhancer-driven regulation of gene expression are not fully understood, they typically allow for binding of multiple different TFs in a combinatorial fashion [56] and display cell type-specific activity [79]. Acetylation of histone 3 lysine 27 (H3K27ac) is a marker for active enhancers, whereas monomethylation of histone 3 lysine 4 (H3K4me1) or H3K4me2 is found at both active, primed and poised enhancers, the latter additionally marked by the repressive H3K27me3 [80].

Furthermore, the methylation of the cytosines of cytosine-guanine dinucleotides (CpGs) conveys gene silencing through preventing TFs from binding or through promoting the compaction of the chromatin [81]. CpG methylation undergoes extensive global changes throughout development and plays a role in the regulation of lymphoid versus myeloid fate in HSCs [82].

Even without investigating histone and CpG modifications, the chromatin accessibility profile alone is sufficient to assign individual cells to different cell types [83, 84]. Accessibility of lineage-affiliated enhancer elements has been shown to occur before RNA expression of the gene in question becomes evident [85, 86], and accessibility of distal, putative enhancer, elements has been suggested to be a better identifier of hematopoietic cell subtypes than mRNA expression [87].

Importantly, the amount of a given protein in the cell is also regulated post-transcriptionally through a number of other mechanisms, such as the timing and rate of synthesis and degradation of the protein, meaning that gene expression level does not always correlate well with the amount of the corresponding protein in the cell [88]. In the context of hematopoiesis, single-cell proteomic analysis of human HSPCs has

shown that protein and mRNA levels correlate better in more lineage-committed progenitors, while HSCs show poor concordance [89].

Lineage priming and output before birth and in adulthood

A number of different studies have investigated the lineage priming and functional output from human embryonic or fetal HSPCs. Findings from some of these studies are summarized below.

Functional output of single HSPCs

Although many studies have reported results of *in vitro* culture of human prenatal HSPCs, the exact developmental stages and populations investigated vary greatly. Studies utilizing single-cell cultures, which is necessary to distinguish multipotency of individual cells from a mixture of lineage-restricted cells, have often been restricted to very specific developmental stages. While providing insights into, e.g., lymphomyeloid potential of pre-definitive wave YS HPCs [27] or differences in lineage bias between disomic and trisomy 21 cells [90, 91] or between FL and fetal BM from the same donor [90, 92], these studies do not provide information about transitions in lineage potential and bias throughout ontogeny, and results between different studies cannot be directly compared due to differences in culture protocol used and technical factors.

The 2016 paper by Notta *et al* provided important insights into differences between fetal and adult hematopoiesis. Through single-cell cultures of second-trimester, neonatal CB and adult BM HSPCs, they found that the FL CD34⁺CD38⁺ progenitor compartment contained a large fraction (~30%) cells with multipotency within the Mk/Ery/My lineages, a property that was largely restricted to the more primitive CD34⁺CD38⁻ in adults [13]. Popescu *et al* performed more detailed comparisons of different stages of fetal development, finding that trilineage as well as erythroid output from single HSC/MPPs decreased between the first and second trimester, whereas bulk cultures resulted in almost no B cells from first-trimester (7-8 pcw) samples [93].

Lineage priming

Efforts have also been made to determine the lineage priming of prenatal HSPCs. As described, HSPCs are heterogeneous at the single-cell level, and studies focusing on the single-cell transcriptomic landscape will be summarized here.

The aforementioned study by Popescu *et al* profiled the single-cell transcriptomes of a large number of first and second-trimester hematopoietic cells, providing a comprehensive overview of hematopoietic and immune cells during human development, but did not extensively explore and compare transcriptional priming in the HSC/MPP compartment across developmental stages [93]. Another such study of different stages of the second trimester and neonatal CB found FL HSC/MPPs to be erythroid-biased compared to fetal BM, which was more biased toward neutrophils and B cells [90].

Two studies that covered CD34⁺ cells from the first trimester to adult or aged BM [94, 95], thus providing a more detailed survey of the HSPC compartment, reported somewhat contradictory results regarding the priming of HSCs throughout ontogeny and in adulthood. While one found indications that first-trimester FL HSC/MPPs are erythroid/megakaryocyte-primed and that both fetal and postnatal BM had more myeloid priming than FL [94], the other reported that most first-trimester and second-trimester HSCs were myeloid or lymphoid-primed, adult HSCs had a relatively balanced distribution and aged HSCs were mostly erythroid or myeloid-primed [95].

A few reports on the chromatin accessibility landscape of second-trimester HSPCs have also been published, profiling gene expression and chromatin accessibility separately or jointly in the same cell. One found no clear transcriptional priming in HSC/MPPs but reported that cells displayed chromatin priming, although they did not investigate differences in lineage affiliation between FL and fetal BM further [92]. The other focused on differences between disomic HSCs and cells with trisomy 21, finding the latter to be more erythroid-biased [91].

B cell development

The early stages of B cell development are of particular relevance to the work in this thesis, and will therefore be described in more detail below.

The first steps in B cell development occur in the BM (Figure 4). Ikaros and E2A promote early steps in lymphoid commitment, after which activation of FOXO1 and EBF1 begins to drive cells towards the B cell lineage. This is followed by expression of PAX5, another critical regulator of B cell commitment, promoting expression of B cell genes and repressing alternative fates [96, 97]. The lineage-committed CD10⁺CD19⁺ proB cell then begins undergoing rearrangement of the immunoglobulin (Ig) genes, largely orchestrated by the Recombination-activating genes *RAG1* and *RAG2*, progressing through a preB stage and losing surface CD34 expression before generation of a functional Ig results in an immature B cell expressing the B cell receptor (BCR) IgM on its surface [97–99]. Immature B cells leave the BM and migrate to the spleen for further maturation. The understanding of B lymphopoiesis is largely based on studies

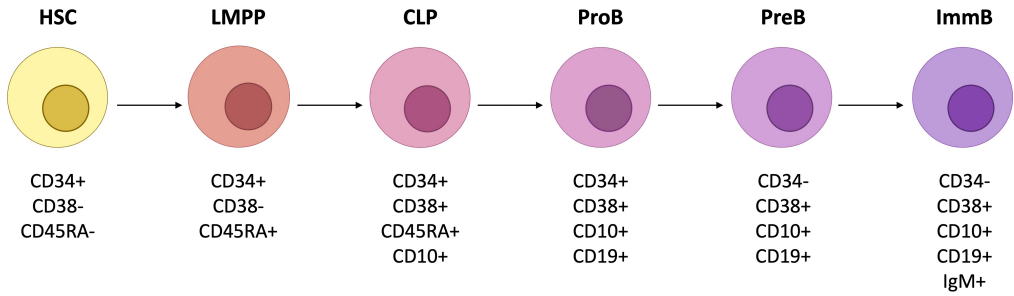


Figure 4: Surface markers at different stages of human B cell commitment. Adapted from [97].

in mice. While many regulators of murine B cell development are known to also play important roles in the human setting, certain differences have been observed, such as the fact that in humans, E2A appears to be a key regulator of BCR rearrangement [97].

During human development, B lymphopoiesis is evident by 7 pcw from the presence of both CD10⁺CD19⁺ proB cells and CD10⁻CD19⁺ preproB cells, the latter of which are rare in adult BM [100]. A fetal IL7R⁺CD10⁻CD19⁻ early lymphoid progenitor [101] has been proposed based on single-cell transcriptomics data to be a prenatal counterpart to the CLP [102].

The conventional B cells of BM origin are also sometimes referred to as B2 cells. In mice, innate-like B1 cells mainly reside in the peritoneal and pleural cavities, can secrete IgM independent of T cell activation, and are mostly produced in prenatal and early life, including a pre-definitive wave originating from the YS early in development [41]. While cells with functional characteristics of B1 cells have been identified in humans, this cell population has not been as extensively characterized in the human setting [103].

Pediatric acute lymphoblastic leukemia

In children, acute lymphoblastic leukemia (ALL) is one of the most common forms of cancer [104], with incidence peaking at around 2-5 years of age, whereas acute myeloid leukemia (AML) increases in frequency with age [105]. Pediatric ALL, which is most often of the B-cell lineage (B-ALL), differs from the disease in adults in terms of the cytogenetic abnormalities driving the disease: *BCR::ABL1*, one of the most common subtypes in adult ALL, is relatively rare in children, whereas *ETV6::RUNX1* and high hyperdiploidy are far more common (Figure 5) [106, 107]. Genetic drivers of ALL include aneuploidy, with gain or loss of multiple chromosomes; alterations of the *PAX5* gene; and chromosomal translocations which either place an oncogene under the control of a transcribed gene, as is the case when *MYC* is translocated to the *IGH* promoter, or

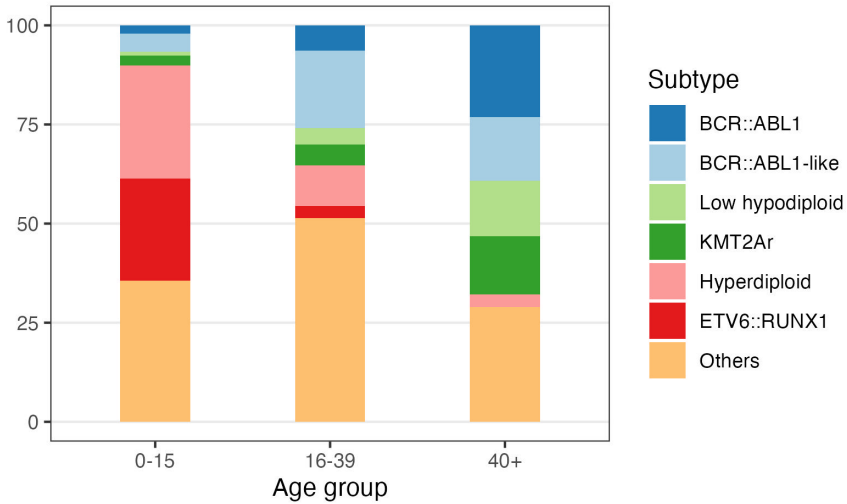


Figure 5: Distribution of common subtypes of B-ALL across age groups. Data from [107].

result in the expression of a novel fusion protein, such as ETV6::RUNX1 [107, 108].

Multiple pediatric ALL driver mutations have been detected in neonatal blood spots [109, 110]. Together with identical genomic breakpoints and immunoglobulin and T cell receptor rearrangements in concordant ALL in monozygotic twins [111, 112], this provides evidence for the *in utero* formation of a preleukemic clone. In the case of *ETV6::RUNX1*, the mutation has been estimated to occur at a 100 to 500-fold higher rate than the leukemia itself [113, 114], indicating that additional events, or secondary hits, are required for overt leukemia. Abnormal immune responses to infection have been suggested to play a role in this transformation [115].

KMT2A rearrangements

In infants below the age of 1 year, chromosomal translocations resulting in rearrangement of the histone-lysine-methyltransferase 2A (*KMT2A*) gene are present in the majority of ALL cases [106]. Whereas pediatric ALL overall has reached a survival of >90% [107], the prognosis for infants with KMT2A-rearranged (KMT2Ar) ALL remains poor [116].

KMT2A, also known as mixed-lineage leukemia 1 (MLL1), is a crucial epigenetic regulator during embryonic development and plays a role in maintenance of HSCs [117]. There are ~80 known fusion partners that generate an in-frame fusion protein, including AFF1 (AF4), MLLT3 (AF9) and MLLT1 (ENL), with the most common fusion in infant

ALL being KMT2A::AFF1 [118]. The fusion protein, which retains domains involved in chromatin binding from KMT2A, results in transcriptional dysregulation, including upregulation of *HOX* genes and the cofactor *MEIS1*, which are normally expressed in immature HSPCs and then downregulated [119]. While this may be conferred by the transcriptional elongation properties of AFF1 [119], the KMT2A::AFF1 protein has also been associated with aberrant enhancer-promoter interactions [120]. Compared to other subtypes of B-ALL, KMT2Ar ALL has a relatively clean mutational landscape [121, 122], which combined with the early onset of the disease suggests that the mutation alone may be sufficient for leukemic transformation.

Interestingly, many KMT2A rearrangements are able to give rise to both ALL and AML, although usually with different frequencies (Figure 6). Some KMT2A fusion partners also show different distributions of myeloid and lymphoid leukemias across age groups: in infants and children, KMT2A::MLLT3 can give rise to both ALL and AML, whereas it almost exclusively results in AML in adults [118]. If and how this relates to the capacity of KMT2Ar leukemia to undergo a lineage switch upon relapse, an event that may be increasing in frequency as CD19-directed immunotherapies become more common [123], is not yet known. Analysis of AML relapse in KMT2A::AFF1 ALL has shown that the lineage switch can occur at different stages of B-lineage commitment, including some cases where shared BCR rearrangement between the presentation ALL and relapse AML were found [124]. Epigenetic dysregulation [124] as well as preexisting LMPP-like and myeloid-primed blasts [125] have been implicated in post-treatment lineage switch.

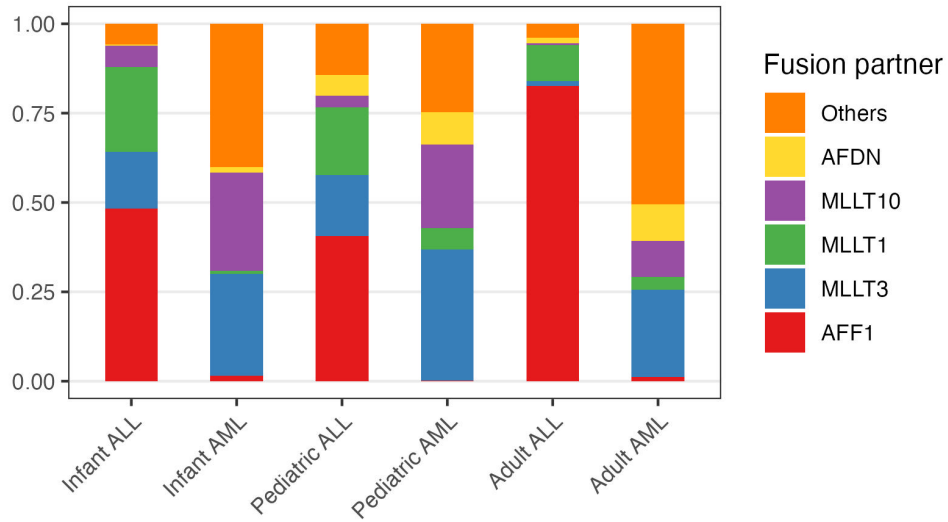


Figure 6: Distribution of KMT2A fusion partners across age groups. Data from [118].

KMT2A::AFF1 infant ALL can be divided into two subgroups based on expression of either *HOXA9* or *IRX1*, another homeobox gene [126], where the *HOXA9*⁺ or *IRX1*⁺ subtype is associated with a worse prognosis [127, 128]. *IRX1* can downregulate *HOXA9*, and instead tends to increase expression of *HOXB4* and *EGR* family members, likely promoting a stem cell program [129].

Models of KMT2A::AFF1 leukemia

A number of attempts have been made to model KMT2A::AFF1-driven B-ALL *in vivo*. Many of these have resulted in B-cell lymphoma or AML, and pro-B ALL without the introduction of cooperative mutations has proven difficult to achieve [130]. One study showed that the overexpression of a microRNA that is upregulated in KMT2A::AFF1 patient blasts in murine *Kmt2a::AFF1*-positive FL cells recapitulated some features of the disease in patients and could result in pro-B ALL upon transplantation [131], but a reliable model of the prenatal events leading to the development of a pro-B ALL has proven difficult to produce.

Different approaches have also been taken to model KMT2A::AFF1 leukemia *in vitro*. Lentiviral transduction of human embryonic stem cells with *KMT2A::AFF1* resulted in upregulation of multiple *HOX* genes and *MEIS1*, but despite an increase in hemogenic specification, the output of CD45⁺ hematopoietic cells was impaired by the fusion gene and the cells did not engraft when transplanted [132]. Another study showed that transient transfection of induced pluripotent stem cell (iPSC)-derived hematopoietic cells with *KMT2A::AFF1* conferred multilineage reconstitution capacity upon transplantation, which in many cases resulted in B-ALL, although often not in secondary transplantations [133].

The effect of KMT2A::AFF1 has also been studied in primary human cells. Retroviral transduction of umbilical cord blood (CB) HSPCs with a human-murine *KMT2A::Aff1* hybrid gene recapitulated several molecular features of the disease and led to pro-B ALL *in vivo* [134]. Besides using the murine *Aff1* to circumvent issues with producing lentiviral vectors carrying the human homologue, this approach also has the drawback that fusion gene expression level is unlikely to be similar to endogenous expression.

Multiple studies have used the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system to induce double-stranded breaks in the *KMT2A* and *AFF1* genes, allowing the chromosomes to spontaneously rearrange to form the translocation [135–138]. CRISPR-induced *KMT2A::AFF1* expression in second-trimester FL was able to recapitulate key features of infant KMT2Ar ALL, such as central nervous system infiltration [136], and the same model system has been used to study the role of the fetal-specific gene *LIN28B* in the leukemogenesis [139].

Comparative studies have shown that the same editing approach can have a different effect on cells from different stages of ontogeny. One study found that cells derived from CB, but not adult BM, could be immortalized in cultures through KMT2A::AFF1 editing, as opposed to KMT2A::MLLT3 which could immortalize both, although B-lineage output or *in vivo* engraftment was not reported in this study [140]. In line with this, a study using lentiviral transduction with KMT2A::MLLT3 could more efficiently immortalize CB HSPCs than adult BM, and only CB-derived cells gave rise acute leukemia *in vivo* [141]. Another study comparing second-trimester FL with CB found that both could be transformed by KMT2A::AFF1, but that FL cells were more sensitive to the choice of intronic breakpoint when assessed for leukemogenic potential *in vivo* [142].

Methodology

Assessing the output of HSPCs

Central to this thesis is the composition and lineage potential of the HSPC compartment. Over the past century, numerous methods have been developed to assess different aspects of this matter; the most important ones will be described below.

Flow cytometry and FACS

For decades, studies of the hematopoietic system have relied heavily on the use of flow cytometry. By staining cells with a cocktail of fluorochrome-conjugated antibodies and interrogating them with lasers, internal and surface marker proteins can be profiled at the single-cell level in millions of cells, allowing for the identification, and in the case of fluorescence-activated cell sorting (FACS) also purification, of different populations.

Many cell surface markers are cytokine receptors, including c-KIT (CD117, receptor for SCF), IL-7 receptor (CD127) and FLT3 (CD135). Example combinations of surface markers for different human HSPC populations are outlined in Table 1. Importantly, certain surface antigens have been found to differ in their capacity to distinguish a population between fetal and postnatal life. For example, GPI-80 and CD201 (EPCR) are not typically used to isolate HSCs from adult BM, but both have been shown to enrich for HSCs in human FL [31, 143]. In the murine system, primitive HSPCs are located in the LIN⁻c-KIT⁺Sca-1⁺ (LSK) compartment, and HSCs can be further defined as CD150⁺CD48⁻CD135⁻ [144].

Table 1: Surface marker combinations commonly used to define human HSPC populations. Based on [145, 146].

Population	Surface marker
HSC	LIN ⁻ CD34 ⁺ CD38 ⁻ CD90 ⁺ CD45RA ⁻
MPP	LIN ⁻ CD34 ⁺ CD38 ⁻ CD90 ⁻ CD45RA ⁻
LMPP	LIN ⁻ CD34 ⁺ CD38 ⁻ CD90 ⁻ CD45RA ⁺ CD10 ⁺
MEP	LIN ⁻ CD34 ⁺ CD38 ⁺ CD45RA ⁻ CD135 ⁻ CD10 ⁻
GMP	LIN ⁻ CD34 ⁺ CD38 ⁺ CD45RA ⁺ CD135 ⁺ CD10 ⁻

In vitro culture assays

Certain properties of HSPCs, namely their self-renewal capacity and lineage potential, can be investigated through *in vitro* culture assays. In addition to not requiring the use of animals, such assays can provide a more flexible and high-throughput approach.

Most *in vitro* HSPC culture systems rely on the addition of cytokines to maintain and/or differentiate the cells.

In colony-forming cell (CFC) assays, cells are cultured in a semi-solid medium such as methyl cellulose [147]. An advantage of this assay is that serial replating of the cells can be used to investigate their self-renewal capacity [148, 149]. As the motility of the cells is restricted due to the medium, the output of individual cells can also be assessed based on morphology. While it is possible to achieve lymphoid differentiation from murine cells in methyl cellulose culture [150], it is not done routinely, and the ability to determine the differentiation capacity of a cell is typically limited to the myeloid and megakaryocyte/erythroid (MkE) lineages, in particular from human HSPCs.

HSPCs can also be co-cultured with BM stromal cells, which can provide certain signals that would normally be provided by the BM niche. The murine OP9 cell line, which does not produce functional M-CSF due to a mutation, can facilitate B-lineage output [151]. OP9 cells modified to constitutively express the Notch ligand Delta-like-1 (OP9-DL1), lose their B-lineage promoting function, but instead allow for generation of T cells [152]. Another murine stromal cell line, MS-5, can support both myeloid and B-cell differentiation from single LIN⁻CD34⁺CD38^{low} HSPCs even without the addition of human cytokines [153].

In recent years, more advanced, three-dimensional *in vitro* culture systems have been developed. BM organoids are 3D structures recapitulating features of the BM niche and can be generated through multilineage differentiation of pluripotent stem cells to form vasculature, mesenchyme and hematopoietic cells, allowing for both the study of the normal hematopoietic system and hematological disorders [154, 155]. The fact that current pluripotent stem cell-derived BM organoids emulate the embryonic niche rather than the adult may pose an obstacle in the study of many adult-derived leukemias, but could be an advantage for leukemias of prenatal origin [154]. BM organoids present a versatile strategy, ethically preferable to animal studies and amenable to manipulation, but do as of now not fully replace the *in vivo* setting.

While *in vitro* culture assays can be very useful, they can also be sensitive to minor technical differences, including the batch of serum used [156]. The lineage output is also highly dependent on the cytokine combination used, making it difficult to translate the behavior of the cells to that in their native niche.

***In vivo* assays**

The gold standard definition of an HSC remains its ability to reconstitute the hematopoietic system upon transplantation. Prior to transplantation, the recipient mouse is usually

sublethally or lethally irradiated to make space in the BM niche and thus allow for engraftment of the transplanted cells. Output from murine donor cells can be distinguished from the cells of the recipient mouse through using strains with different isoforms of the pan-hematopoietic surface marker CD45 (CD45.1 and CD45.2), or through detection of Y-chromosome DNA from male donor cells in female hosts [150]. The long-term repopulating capacity of an HSC can be assessed through secondary transplantations into new hosts.

Xenotransplantation of human HSCs was compromised by host rejection of the foreign cells. Establishment of the severe combined immune-deficient (*Scid*) mouse [157], was an important step in transplantation of human hematopoietic cells. The lack of adaptive immunity allowed for the reconstitution of the human B and T lineages, although the lack of myelopoiesis, likely due to poor cross-reactivity with murine myeloid cytokines, made it difficult to ascertain the engraftment of human HSCs [8]. The *Scid* mouse was later crossed with the nonobese diabetic (NOD) mouse, which instead shows defect in innate immunity, resulted in the NOD-*scid* mouse strain [158]. NOD-*scid* mice, which still retain functional NK cells, were then crossed with mice with truncation or deletion of the IL-2 receptor gamma chain, the latter referred to as the NOD-*scid*-gamma (NSG) mouse, which resulted in improved multilineage reconstitution by CD34⁺ human HSPCs [159, 160]. As NSG mice still showed poor myelopoiesis and long-term engraftment of human HSCs, mice engineered to express human versions of cytokines such as TPO, IL-3 and GM-CSF have also been developed to better support these aspects [161].

While long-term engraftment of human HSCs can be achieved in the mouse, this setting cannot perfectly recapitulate the human BM. Therefore, strategies to better emulate the human BM niche have been developed. In humanized ossicles (hOss), human mesenchymal stromal cells are seeded on a scaffold to form 3D structures that when implanted into immunodeficient mice result in the formation of bone and bone marrow that can support transplanted healthy and leukemic human hematopoietic cells [162].

Nevertheless, due to the stress caused by *ex vivo* manipulation of donor cells and irradiation of the recipient, the transplantation setting is not equivalent to native hematopoiesis. Efforts have been made to trace hematopoietic output without removing the cells from their niche. Mouse strains have been developed where induced expression of e.g. a recombinase or nuclease results in a random modification that works as a barcode to track subsequent cell divisions, and single-cell readouts of mitochondrial or genomic DNA mutations or CpG island methylation patterns have been used to infer clonal relationships in humans [163]. These approaches provide complementary information to transplantation assays.

Investigating molecular properties of HSPCs

Over the past few decades, rapid technological advances have resulted in assays that allow for dissection of the genetic, transcriptomic and epigenetic landscape of cells at an unprecedented scale. The technologies of particular relevance to the work in this thesis are described below.

RNA sequencing

The advent of high throughput sequencing meant that large-scale interrogation of the transcriptome was no longer dependent on microarray assays, opening up for untargeted studies [164]. Sequencing of the transcriptome of a single cell using high throughput sequencing was first reported in the late 2000s, where single mouse blastomeres and oocytes were manually isolated and sequenced [165]. The technology then rapidly expanded to profiling hundreds or thousands of cells. Many single-cell RNA-seq protocols rely on partitioning of the reverse transcription reactions through either the sorting of single cells into wells on a plate [166], encapsulating the cells in droplets [167, 168] or distributing them on an array of microwells [169]. During the sequencing library generation, material from each cell is labeled with a unique sequence through which molecules from different cells can later be distinguished.

In the most common droplet-based methods, cells are passed through a microfluidic channel and partitioned into droplets together with beads coated in reverse transcription primers, where they are lysed in order for the first few steps of library generation to take place (Figure 7). The primers contain a bead-specific DNA "barcode" as well as a unique molecular identifier (UMI) sequence on each primer, ensuring that each reverse-transcribed cDNA molecule can be assigned to a specific cell and that unique mRNA copies can be quantified through the UMIs [167, 168]. The cDNA is then further amplified and fragmented to produce the final sequencing library. Droplet-based methods allow for higher throughput than plate-based ones, with the downside that they typically only allow for either the 3' or 5' end of the mRNA to be captured.

While single-cell RNA-seq has become highly used in studies of the hematopoietic landscape, it still suffers from dropout events. The absence of a certain transcript in a cell at the time of capture can, but does not necessarily, mean that the gene is not expressed; it could also be due to technical factors resulting in inefficient conversion of the mRNA to cDNA, or due to the sampling of molecules for sequencing [170, 171]. Additionally, the fact that gene transcription often occurs in bursts rather than continuously provides an intrinsic, technology-independent, source of variance [171]. Therefore, bulk RNA-seq still serves a purpose.

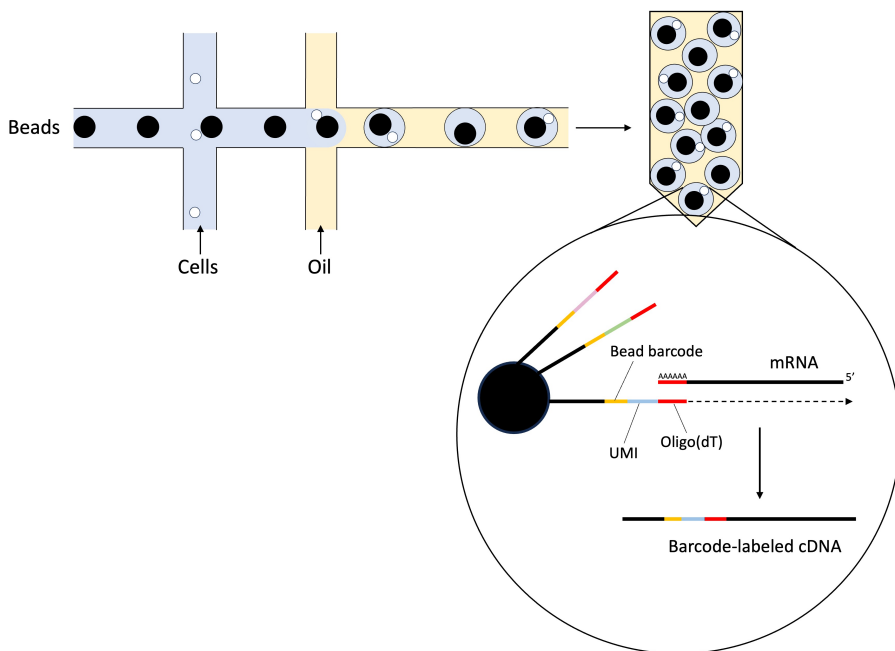


Figure 7: Schematic of the first steps of droplet-based single-cell RNA-seq library generation.

ATAC-seq

First published in 2013, the assay for transposase-accessible chromatin using sequencing (ATAC-seq) has become a widely used technique for the profiling of accessible chromatin in a relatively unbiased manner (Figure 8). Using isolated nuclei as the starting material, ATAC-seq takes advantage of the ability of a hyperactive Tn5 transposase to introduce breaks in the DNA at accessible sites in the chromatin and insert DNA fragments, which are used as PCR handles in subsequent library preparation steps [172]. Compared to the other commonly used methods DNase-seq and MNase-seq, the ATAC-seq protocol is faster and requires fewer cells [173].

Early reports of single-cell ATAC-seq profiled a few hundred up to ~15,000 cells utilizing a programmable microfluidics platform [83] or split-pool combinatorial barcoding [174]. With the adaptation of droplet-based single-cell technologies, data generated in individual studies reached the hundreds of thousands [175, 176].

Just as single-cell RNA-seq suffers from dropout events, single-cell ATAC-seq has its limitations: as a mammalian nucleus typically only contains two copies of each chromosome, single-cell ATAC-seq data is by its nature sparse and nearly binary, meaning that individual genomic sites are very sensitive to noise [83].

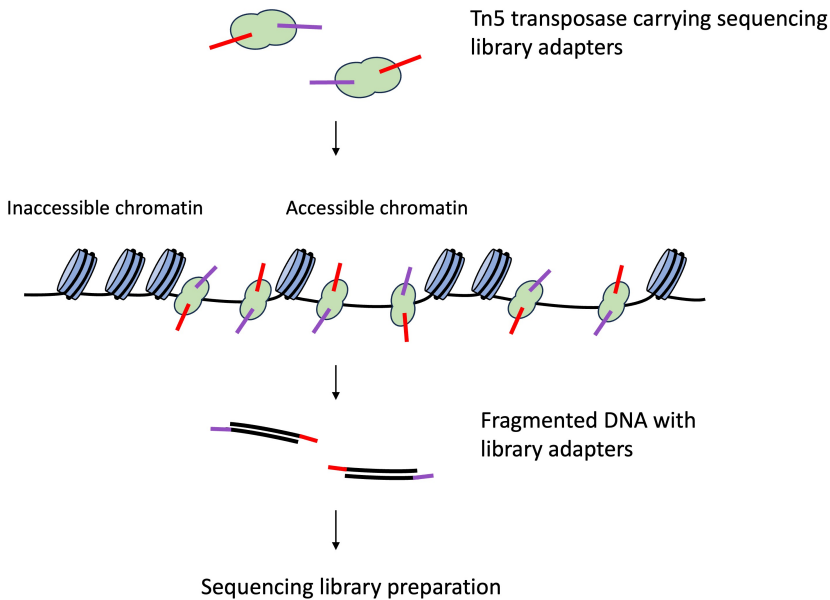


Figure 8: Schematic of ATAC-seq.

Multiomic assays

With the establishment of single-cell RNA-seq and ATAC-seq techniques, a logical next step is to combine the two. This was first accomplished through split-pooling techniques where cells are labeled through combinatorial barcoding [177, 178], but droplet-based technology soon followed [179].

Methods such as REAP-seq and cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) stain cells with surface marker antibodies labeled with polyadenylated DNA barcodes before processing for single-cell RNA-seq to simultaneously capture transcriptome and surface markers [180, 181]. The antibody-specific barcode is then used to assign sequenced fragments to their corresponding surface marker.

Over the years, numerous other techniques for joint profiling of different transcriptomic, epigenomic and proteomic modalities have emerged [182], but they are beyond the scope of this thesis.

Processing and visualizing single cell omics data

Initial processing

Once raw sequencing data has been generated, the sequencing reads are typically aligned to a reference genome and quantified with the goal of generating a matrix containing the genomic features, e.g. the genes, and the cells (or, in the case of bulk RNA-seq, the samples). Reads are assigned to specific cells, or demultiplexed, based on the cell barcodes. In the case of RNA-seq, reads are mapped to the genome using a splice-aware software such as STAR [183], and the resulting cell-gene matrix is used for further analysis.

In ATAC-seq, the chromatin regions are not usually used directly. Instead, peak calling is performed to find genomic regions that are enriched for sequenced fragments, referred to as "peaks", using a software such as MACS [184]. Due to the near-binary nature of single-cell ATAC-seq data, peak calling cannot be performed on individual cells. Instead, peak calling is performed on the aggregated data from the whole dataset or on a per-population basis (Figure 9). The latter can preserve rare cell type-specific peaks that would be lost in peak calling on the whole dataset [185], but requires a priori knowledge of cell types or an iterative approach where peaks called on the whole dataset are first used to inform cell clustering, and then peak calling is performed on individual clusters. Alternatively, the genome can be divided into bins and fragments quantified within each bin [186]. The cell-peak or cell-bin matrix is then used for further analysis.

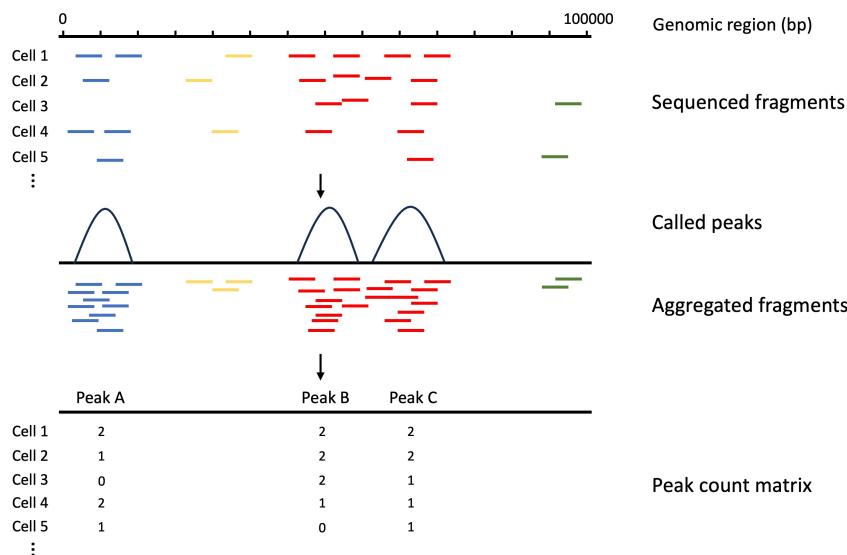


Figure 9: Schematic of peak calling in single-cell ATAC-seq.

Quality control

Once the feature-cell matrix has been generated, downstream data analysis usually begins with filtering out low-quality cells, e.g. dead or dying cells, and potential doublets where two cells have received the same cell barcode. These steps, and subsequent analyses, are often largely performed using tools specifically designed for single-cell data analysis. A popular choice for the R environment, and the one primary used in the work presented in this thesis, is the Seurat package [187, 188] for single-cell RNA-seq and Signac [185], which provides Seurat-compatible tools for single-cell ATAC-seq data analysis.

Cells with a low number of unique features, e.g. gene transcripts or peaks, or with a high percentage of reads from mitochondrial genes may represent dying cells with leaking cell membranes, and cells with unusually high feature or total read counts are potential doublets [189]. Specialized software for the detection of doublets [190, 191] and the removal of ambient RNA (cell-free RNA encapsulated in the droplet together with a cell) [192, 193], can also be employed. For single-cell ATAC-seq data, cells can additionally be filtered based on enrichment of sequenced fragments at transcription start sites (TSS enrichment) and the ratio of nucleosome-length to non-nucleosome-length fragments (nucleosome signal) [189].

Dimensionality reduction, visualization and batch correction

Once the data has been filtered based on the aforementioned quality parameters, this is usually followed by normalization to account for differences in sequencing depth between individual cells. A common type of normalization for single-cell RNA-seq is log-normalization, where transcript counts are normalized based on the total number of transcripts detected in the cell, after which a pseudo-count of e.g. + 1 is added before log-transformation [194]. Other methods, such as the negative binomial regression-based *sctransform* of Seurat, have been proposed to better stabilize the variance across different sequencing depths [195, 196], although a study comparing different transformation methods found that the simpler log-transformation method overall performs well [194].

After normalization, highly variable genes are selected and are used as input for dimensionality reduction. Despite the thousands of genes included in most datasets, single-cell RNA-seq data can be regarded as relatively low-dimensional due to the covariance of many genes, making it highly suitable for dimensionality reduction using principal component analysis (PCA) [197]. In PCA, the dimensionality of the data is reduced by finding the linear combinations of the variables, i.e., the genes, that best capture the

variation within the dataset [198].

PCA is not usually used to visualize single-cell data directly. Instead, the principal components that capture the majority of dataset heterogeneity are used as input for non-linear dimensionality reduction methods such as t-distributed stochastic neighbor embedding (t-SNE) [199] or uniform manifold approximation and projection (UMAP) [200]. As UMAP tends to better preserve global relationships between cell types [201], it has become a popular choice for the visualization of single-cell data. However, UMAP is still a non-linear reduction method, meaning that distance or lack thereof between different populations should not be interpreted as biologically meaningful.

For single-cell ATAC-seq data, latent semantic indexing (LSI) [202] is commonly used instead of log-transformation and PCA [185, 189]. In LSI, peak counts are first normalized based on sequencing depth and then across the dataset based on the inverse of their frequency, thus giving more weight to rare peaks, and this normalized matrix is then used as input for singular value decomposition (SVD) [174]. The SVD dimensions are then used as input for UMAP generation.

In single-cell RNA-seq combined with surface marker analysis, as in CITE-seq, the readout of the antibody barcodes, or antibody-derived tags (ADTs) can be analyzed separately, but are often analyzed jointly with the transcriptome data. The ADT counts matrix is normalized separately using, e.g., centred log-ratio transformation, and the different modalities measured in the same cell can be used to create joint dimensionality reduction embeddings for visualization and further analysis [189].

When combining multiple datasets, technical variation can result in considerable batch effects on the clustering and visualization, and a number of different methods have been developed to integrate different datasets to correct for this. One commonly used method is Harmony [203], which has been found to perform well on both relatively non-complex single-cell RNA-seq datasets and on single-cell ATAC-seq data when benchmarked against other methods [204]. Harmony takes PCA embeddings as input, and corrects the embeddings by iterative soft assignment of cells to clusters, penalizing clusters with cells from few different batches, and calculation of a correction factor for each dataset based on the cluster centroids [203].

Factors such as the cell cycle status can also obscure cell-type specific features of the dataset, and it may be desirable to remove these effects from the analysis [189, 205]. However, cell cycle status can also provide relevant biological information, and therefore it is sometimes preferable to not correct for cell cycle, but instead analyze the data with these effects in mind.

Grouping and ordering cells

In order to identify, analyze and compare cell populations, similar cells need to be grouped together, or clustered. Clustering based on a nearest neighbor-based graph representation, where cells are nodes connected to their most similar neighbors based on Euclidean distances in the PCA space, has been found to work well for single-cell data [189, 206]. Communities of densely connected regions in the graph are detected using algorithms such as Louvain [207] or Leiden [208]. Importantly, the number of clusters detected depends on the clustering resolution, which is set by the user. Appropriate clustering resolution depends on the dataset, and the number of clusters produced needs to be evaluated based on insight into the sample biology and the research question.

In addition to studying differences between discrete clusters of cells, it is sometimes desirable to study gradual changes in a genomic feature over a process such as hematopoietic differentiation. This can be accomplished by ordering the cells along a trajectory based on their pseudotime [209]. A benchmarking study comparing a large number of trajectory inference methods [210] found that Slingshot, which uses clustered cells as a starting point to draw a minimum spanning tree connecting the most similar clusters [211], overall performed well. Partition-based graph abstraction (PAGA) [212], which estimates the strength of connectivity between different Louvain clusters using a statistical model, performed better on more complex trajectories and can handle multiple disconnected trajectories. Monocle 3 [213] is built on the same principles as PAGA but with further smoothing and pruning of small branches.

The performance of many trajectory inference methods has been shown to vary greatly between different datasets [210]. Furthermore, there is no guarantee that the inferred trajectories will recapitulate true biological phenomena. Therefore, while pseudotime ordering is a valuable tool for calculating and visualizing gradual changes across the dataset, caution should be exercised when drawing conclusions about biology based on the inferred trajectories.

Dataset annotation

Once cells have been clustered, the clusters are usually assigned labels based on the type of cell they primarily consist of. One way to do this in single-cell RNA-seq data analysis is by assessing the expression of cell type-specific genes, or marker genes. This process often starts with finding differentially expressed genes. There are a number of different methods that can achieve this; the default method in Seurat uses the Wilcoxon rank-sum test for statistical evaluation and reports the results as log-fold difference between clusters [188, 214].

In single-cell ATAC-seq, individual peaks are typically not used as cell type markers the way individual genes are. The aggregated peak counts overlapping a gene body can be informative, but this measurement is usually quite noisy compared to mRNA expression. An alternative approach for the sparse chromatin data is to calculate the relative accessibility of TF motifs, or "motif activity", in each cell, and then evaluate the difference in motif activity between clusters. The chromVAR package [215] was developed for this purpose. chromVAR calculates how the number of fragments in peaks containing a given TF motif in each cell deviates from the expected number based on dataset average, adjusted for background such as GC content of the peak set. This results in a relative measure of TF motif activity compared to the other cells, and differences in motif activity between different clusters can then be used to assign cell type labels, similarly to differences in gene expression.

Annotating clusters using this approach requires curated lists of marker genes or TFs based on biological insight. A marker that is specific to a certain group of cells but not differential between subtypes of said cell type within the dataset might also fail to be detected. Instead of relying on curated lists for manual cell type annotation, previously published annotated data can be used as a reference. Seurat accomplishes this by identifying "anchor" cell-cell pairs across the query and reference datasets based on mutual nearest neighbors [216] and then annotating the query cells based on the anchors to the reference data in its local neighborhood, scoring the prediction based on the consistency between different anchors [217].

For single-cell ATAC-seq data, the total number of sequenced fragments overlapping the gene body can be used as a proxy for RNA expression of the gene to facilitate annotation based on a single-cell RNA-seq reference dataset [185]. It is also possible to take advantage of a separate multiome dataset as "bridge" to translate between RNA-seq and ATAC-seq modalities [188].

Reference mapping, or label transfer, has the advantage that multiple datasets can be labeled using the same reference, providing a degree of standardization [218]. However, the label transfer strategy is highly dependent on high-quality, well-annotated reference datasets. Over the past decade, efforts have been made to produce organism-wide single-cell omics datasets, such as the Human Cell Atlas [219] and the Tabula Muris [220]. In parallel, new tools, such as those leveraging machine learning algorithms, are rapidly being developed to annotate new datasets based on reference data [218]. Still, a reference dataset comprised of healthy adults may not encompass all cell types and states required to fully and accurately annotate a dataset derived from, e.g., prenatal development, aged individuals, disease or *in vitro*-generated cells. As an alternative to reference mapping, generative artificial intelligence (AI)-based annotation has been introduced as an automated method that relies on marker genes and shown to overall agree well with manual annotation [221]. However, as with other uses of generative

AI, this method has limitations when it comes to determining the source of the information used to annotate the data, reproducibility and the risk of hallucination. Results therefore still require careful manual evaluation. The best approach is therefore likely a combination of automated and manual annotation.

Hematopoietic differentiation of iPSCs

20 years ago, screening of factors involved in the maintenance of embryonic stem cells led to the discovery that enforced expression of *Oct3/4*, *Sox2*, *c-Myc* and *Klf4* was enough to reprogram somatic murine cells into pluripotent stem cells able to form all three germ layers, termed induced pluripotent stem cells (iPSCs) [222]. The generation of human iPSCs soon followed [223, 224]. As opposed to embryonic stem cells, iPSCs do not require the destruction of an embryo, and the fact that they can be generated from essentially any donor make them a promising avenue for disease modeling and therapeutics.

The *in vitro* differentiation of pluripotent stem cells towards hematopoietic cells has been shown to recapitulate properties of developmental hematopoiesis [225, 226], making them a suitable platform for the study of human hematopoietic development. Depending on the type of mesoderm patterning and the inhibition or induction of pathways such as WNT and Nodal/activin, the hematopoietic process can be modeled after extraembryonic, i.e., yolk sac, or intraembryonic hematopoiesis [226]. Hematopoietic progenitor cells derived from iPSCs can be further differentiated *in vitro* to produce different lineages, including erythrocytes [227], myeloid cells [228] and B cells [229].

Although the *in vitro* generation of human HSCs holds great promise for therapeutic HSC transplantations, cells with long-term multilineage reconstitution capacity have proven difficult to produce. Human multilineage-engrafting iPSC-derived HSCs were first reported from teratomas resulting from injection of iPSCs into immunodeficient mice [230, 231], but this method is not suitable for large-scale manufacturing for clinical applications. Another study showed that lentiviral vector-based transduction of seven TFs in human pluripotent stem cell-derived hemogenic endothelium could confer long-term multilineage engraftment upon transplantation into NSG mice [232], but genetic manipulation may still pose a risk in a clinical context. However, intense research has recently led to reports of transgene-free human iPSC-derived HSCs capable of long-term multilineage engraftment in mouse transplantation experiments, first in cultures including human plasma [233], and more recently with a more refined protocol using a completely defined medium [234].

Ethical considerations

The work in this thesis has included both primary human material and animal experiments. Each of these comes with their own ethical considerations, which will be described below.

Human primary material

The findings presented in this thesis include analysis of human material from fetal tissue, umbilical cord blood and adult bone marrow. Samples were collected anonymously, with informed consent, and handled according to the ethical principles outlined in the Declaration of Helsinki [235], which was created to ensure that research involving human participants is conducted with respect for their health and rights. Per regulations regarding collection and use of human samples for research purposes in Sweden, approval by the Swedish Ethical Review Authority is also required.

Research using human samples can only be approved when the value of the research can be considered greater than the risks it poses to the research subject and no acceptable alternative method exists. The overall purpose in this case was to better understand the initiation of pediatric ALL, which can hopefully lead to improved patient outcomes in the future. Due to the differences between murine and human biology, and the fact that *in vitro* model systems using pluripotent stem cell lines cannot yet perfectly recapitulate all aspects of human hematopoiesis, the use of primary human material is motivated.

While genomic data can be anonymized in the sense that no sample information that could lead to identification of the donor is kept, the genomic data itself is unique to each individual and may be identifiable, and could for this reason be regarded as sensitive data [236]. Therefore, raw sequencing data generated during the work was stored securely, and only processed data was published in a freely accessible manner.

Animal experiments

Experiments conducted in mice were also included in the work. Animal experiments are necessary in the cases where no sufficient alternative methods, such as *in vitro* culture systems, exist. As described in previous sections, existing *in vitro* methods modeling human prenatal hematopoiesis and the BM niche are promising, but not able to fully replace the *in vivo* setting.

Animal experiments were performed according to the 3Rs principles, meaning that they should be replaced whenever possible, reduced in number, and refined with regards to

the severity of the procedures the animals are subjected to, principles that have been widely adopted across the world [237]. Experiments were repeated enough times that the results could be considered reliable and significant, but not conducted in unnecessarily large numbers. Some experiments required the mice to develop leukemia, in which case they were carefully monitored for signs of illness and humanely sacrificed when disease was evident, in order to prevent unnecessary suffering.

Aims of the thesis

The overall aims of this thesis were to investigate the human fetal hematopoietic context and to study the initiation of KMT2A-rearranged ALL in a prenatal setting. These topics have been approached through a combination of single-cell genomic assays, *in vitro* cultures and flow cytometry readouts, and experiments introducing the human *KMT2A::AFF1* fusion gene in murine and pluripotent stem cell-based models.

The aims can be summarized as addressing the following questions:

1. What are the molecular properties of human fetal liver HSPCs, and how do they differ from those in the adult setting? (Papers I and II)
2. Can connections be drawn between features of normal prenatal hematopoiesis and pediatric ALL? (Papers I and II)
3. What are the effects of the KMT2A::AFF1 fusion gene at different stages of B-lineage commitment and at different ontogenic stages? (Paper III)
4. What are the effects of the KMT2A::AFF1 fusion gene on HSPCs in a human, prenatal-like setting? (Paper IV)

Author's contribution to included papers

Paper I

This project was ongoing when I started my PhD, and I contributed with analysis of previously published RNA-seq data. My main contributions are related to Figure 5. During revisions, after the first author had left the lab, I took on a bigger role in addressing the reviewers' questions, contributing to data analysis and interpretation as well as revising the manuscript text.

Paper II

The single-cell ATAC-seq dataset had been acquired and the raw data processed before I took over, after which I performed virtually all of the included analyses of the dataset. The project as presented in the paper was conceived and the manuscript written by me and my supervisor. My contributions constitute the majority of Figures 2-5, and I participated in the experiments presented in Figure 1. Together with my supervisor, I was also in charge of communication with the journal and addressing reviewer comments.

Paper III

My contributions to Paper III were primarily to the analysis of human bulk and single-cell RNA-seq data seen in Figure 6. Before the first submission of the paper, I added additional analyses and reviewed code written by another author. During revision of the paper, when further analysis of human data was needed, I took the lead on those analyses. I also reviewed and gave feedback on the manuscript before the first and revised submissions.

Paper IV

In Paper IV, I was in charge of establishing the protocol for the CRISPR-induced chromosomal translocation. I then continued to be involved in planning and execution of experiments and interpretation of data throughout the project. I also took the lead on planning and coordinating, as well as participated in, generation of the single-cell RNA-seq dataset which constitutes a large part of the manuscript. Once the sequencing data was obtained, I was the main person processing and analyzing it. My contributions to the paper are mainly found in Figures 1, 2 and 4, and in the writing of the manuscript.

Summary of included papers

Paper I

Single-cell multiomics of human fetal hematopoiesis define a developmental-specific population and a fetal signature

Better understanding of the development of the human blood system may aid in the efforts to produce transplantable HSCs *in vitro* and in the understanding of the pre-natal initiation of pediatric leukemia. The heterogeneity of the hematopoietic system, especially as the FL at this stage contains a mixture of transient and definitive-wave HSPCs [225], motivates the use of single-cell technologies. In this study, we set out to molecularly profile first-trimester HSPCs and compare them to adult BM counterparts. By sorting LIN⁻CD45⁺CD34⁺ cells from CS22 (~8 pcw) FL and subjecting them to CITE-seq, we simultaneously obtained single-cell transcriptomes and surface marker expression data from around 6,700 cells from two donors.

Using the combination of transcriptome and surface markers, we were able to select cell populations corresponding to traditional surface marker-based phenotypes, such as CLP and GMP, and then assess the transcriptional profile of these populations. We found that certain populations, such as HSCs and LMPPs, corresponded well to transcriptome-based definitions of the same populations. However, other phenotypically defined populations, such as CMPs and GMPs, showed high transcriptional heterogeneity, containing a mix of transcriptionally defined primitive and lineage-primed populations. We also found that IL-7 receptor (IL7R) was a better surface marker to capture a fetal transcriptional CLP-like population than CD10, which is more commonly used for postnatal cells.

In order to compare the fetal HSPCs to adult counterparts, we used the fetal dataset as a reference and projected a previously published adult BM HSPC dataset [238] onto the fetal data and labeled the adult cells based on their predicted corresponding cell type. Doing so, we found that the MPP-I population, which made up ~17% of the FL dataset, was nearly absent from adult HSPCs. Performing CITE-seq on CS16 and 9 pcw FL cells, and evaluating previously published FL, fetal BM, CB and postnatal BM data [94, 238], we could confirm that this MPP population was largely specific to the FL. Differentially expressed genes between the MPP-I, MPP-II and HSC clusters indicated that MPP-I constitutes an erythro-myeloid primed multipotent progenitor population.

To investigate transcriptional differences between FL and adult BM HSPCs, we pseudobulked each cluster, using the predicted FL cell types from the label transfer for the adult BM dataset, and calculated differentially expressed genes on a per-cluster basis. Selecting genes that were differentially expressed in all primitive, lymphoid or

erythromyeloid populations, we extracted fetal and adult signature gene sets which we validated using external datasets [94, 239]. Using a publicly available RNA-seq dataset containing samples from a large number of B-ALL patients [240], we could show that in the case of KMT2A::AFF1, fetal gene signature was more highly expressed in the youngest patients, aged 0-2 years, whereas the adult gene set was more highly expressed in the other age groups.

Paper II

Chromatin Accessibility Shapes Developmental-Specific Lineage Plasticity in Hematopoiesis

In this study, we further investigated the HSPC compartment of early human FL, evaluating differences in functional lineage output and chromatin-level priming.

While previous studies have assessed the *in vitro* lineage output of human fetal HSCs [13, 90, 92, 93, 95], multilineage-promoting single-cell cultures that facilitate B lymphopoiesis and direct, parallel comparisons to adult HSCs, particularly the combination of the two, have been few or nonexistent. We addressed this knowledge gap by index sorting single LIN⁻CD45⁺CD34⁺CD38^{-/low}CD90⁺CD45RA⁻ HSCs from first-trimester FL, neonatal CB and adult BM and culturing them in an MS-5 co-culture system allowing for broad lineage output. We found that the colonies emerging from adult HSCs were to a large extent (~70 %) unilineage, usually myeloid, whereas both FL and CB cultures resulted in around 40% multilineage clones. Additionally, FL and CB HSCs readily produced B cells, even in the absence of human IL-7, while the same conditions rarely resulted in B-cell positive colonies in adult BM HSC cultures.

To profile the chromatin landscape of early FL HSPCs, we performed single-cell ATAC-seq on LIN⁻CD45⁺CD34⁺ cells from CS22 FL. Taking advantage of previously published single-cell ATAC-seq data from adult BM HSPCs [238], we projected our FL dataset onto the adult BM UMAP and annotated them by their predicted identities. We then compared the HSCs of the two datasets by finding differentially accessible peaks and searching for enriched TF motifs within these peaks. In the FL HSC-upregulated peaks, the relatively few enriched motifs included motifs for HOXB7 and HOXB8, but little indication of hematopoietic lineage priming. In the adult HSCs, on the other hand, myeloid-associated TFs such as JUN and megakaryocyte-associated NF-E2 [241, 242] were enriched in upregulated peaks and also appeared occupied by their corresponding TF as per TF footprinting analysis.

Incorporating the CS22 FL and adult BM CITE-seq datasets analyzed in Paper I, we could also show that at the transcriptome level, FL HSCs scored higher for a gene set associated with multilineage HSCs while adult HSCs scored higher for a platelet bias gene set [243]. Integrating the ATAC-seq and RNA-seq data to create pseudo-multiome data, we used SCENIC+ [244] to infer enhancer-driven gene regulatory networks. Certain factors, such as the HSC-associated HLF [57], were found to be active in both FL and adult BM HSCs, but known fetal-associated factors such as LIN28B and HMGA2 [245] were found to be more active in the FL HSCs, whereas adult BM HSCs scored higher for, e.g., gene or chromatin region sets associated with FOS, which forms complexes with JUN and acts in myelopoiesis [241, 246] and the erythroid TF NFI-A [247].

Overall, the analysis of the chromatin landscape and transcriptome indicated that early fetal HSCs exist in a more permissive, less lineage-restricted state, whereas adult HSCs are primed towards the myeloid and M_kE lineages.

We then investigated the lymphoid progenitor compartment. Calculating TF motif activity using chromVAR [215] and selecting differentially active motifs between HSCs and the lymphoid (Ly)-II and GMP-II clusters, we found that motifs upregulated in GMP-II had a clearly myeloid gene expression profile in both the FL and adult BM datasets. However, in the FL dataset, the Ly-II cluster displayed a lympho-myeloid profile that stood in contrast to the more distinctly lymphoid profile in the adult counterparts, suggesting that early fetal Ly progenitors retain some degree of myeloid plasticity. Additionally, the motif for PAX5, a master regulator of the B-cell lineage [248], was upregulated in adult, but not FL, Ly-II cells.

We further investigated the regulation of the B-lineage by using the Cicero package [249] to find distal peaks co-accessible with peaks in promoter regions and intersecting these with peaks differentially accessible between FL and adult Ly-II cells, we identified 3 peaks which were co-accessible with the PAX5 promoter and upregulated in adult Ly-II cells. These peaks overlapped histone enhancer marks and binding sites for PAX5 and EBF1, suggesting they are regulatory elements of the B-lineage. Finally, we used PCA to more generally profile the lymphoid compartment. Selecting variable motifs between the Ly-I, Ly-II and GMP-II clusters for PCA, we found that the FL Ly-II cells clustered closely with the more immature Ly-I cells, while the two populations separated in adult BM. Performing the same analysis on cells from a 12 pcw FL dataset [91], we found that they exhibited an intermediate behavior. In all, these results suggest that at the chromatin level, early FL lymphoid progenitors are less restricted to the B-cell lineage than adult counterparts.

Paper III

Ontogeny-specific induction of the *KMT2A::AFF1*-fusion drives development of a distinct CD24 positive pre-leukemic state

To accurately model the initiation of *KMT2A::AFF1*-driven infant ALL, the oncogene should be introduced in the prenatal setting where it occurs in human patients. In this project, we used an inducible murine model of *KMT2A::AFF1* leukemia and evaluated the effect of oncogene induction in different hematopoietic populations and at different stages of development.

Through crossing of different mouse strains, we generated a model where the human *KMT2A::AFF1* fusion gene could be induced in Vav-Cre-expressing HSPCs and their downstream progeny. Inducing expression of *KMT2A::AFF1* in HSPCs from early stages of embryonic development resulted by E18.5 in increased *Hoxa9* expression and a reduction in LSKs and LMPPs, with severely impaired *in vitro* cloning efficiency in the HSC compartment and increased myeloid bias in both HSCs and LMPPs. In the B220⁺CD43⁺ B-cell progenitor compartment, *KMT2A::AFF1* expression led to an increase in preproB and proB cells, as well as expansion of a CD24⁺CD19⁻ preproB population that was barely detectable in controls. These expanded *KMT2A::AFF1*⁺ CD24⁺ preproBs also showed increased cloning efficiency and aberrant myeloid output.

Inducing *KMT2A::AFF1* in lympho-myeloid or B cell progenitors using Rag1-Cre or Mb1-Cre strains did not result in significant changes in the early B cell progenitor compartment and only minor increases in *Hoxa9* expression. When induction of *KMT2A::AFF1* in HSPCs was delayed until post-birth, HSCs still exhibited reduced cloning efficiency and the LMPPs higher myeloid bias, but there was no difference in preproBs or proBs. In combination, these results show that prenatal *KMT2A::AFF1* induction at the HSPC stage is needed for the B220⁺CD24⁺ preproB phenotype to emerge, indicating that both developmental stage and target cell play a role in the leukemogenic properties of *KMT2A::AFF1*.

Upon prenatal HSPC induction, the *KMT2A::AFF1*⁺ CD24⁺ preproB population was further shown to possess self-renewal capacity in a replating assay, making it a likely leukemia-initiating cell candidate. LSKs and CD24⁺ preproBs were transplanted into adult sublethally irradiated NSG mice. While control CD24⁺ preproBs did not result in long-term reconstitution, *KMT2A::AFF1*⁺ counterparts showed long-term multilineage engraftment, producing T and myeloid cells despite their initial preproB phenotype. Several of the engrafted *KMT2A::AFF1*⁺ CD24⁺ preproB mice showed enlarged spleens and an immature Lin⁻CD19⁻IL7R⁺ lymphoid population, properties that were not observed when transplanting LSKs.

Having observed *in vitro* and *in vivo* effects of *KMT2A::AFF1* induction on proliferation, self-renewal and lineage bias, we next profiled the transcriptome of *KMT2A::AFF1*⁺ and control cells using single-cell RNA-seq. In the *KMT2A::AFF1*⁺ preproBs, *KMT2A::AFF1* target genes such as *Hoxa9* and *Meis1* and stemness-associated genes such as *Hlf* and *Mecom* were upregulated compared to controls, whereas downregulated genes overall showed a proB profile. We found a subpopulation of *KMT2A::AFF1*⁺ cells that clustered together with the more primitive cells and also expressed *Cd24a*. By selecting genes upregulated in this population compared to both the primitive, HSC-like cells they clustered with and to the control preproBs, as well as genes downregulated compared to the same populations, we extracted a preleukemic gene signature.

Using publicly available bulk RNA-seq data from B-ALL patients [240], we could show that the preleukemic gene set was enriched in *KMT2A::AFF1* patients compared to other subtypes. Analysis of single-cell RNA-seq data from infants with *KMT2A::AFF1* ALL [125, 250] further showed that cells positive for the preleukemic signature were present in patient samples, but not in a healthy BM reference. Thus, the murine model provided insights into the early stages of leukemogenesis, uncovering transcriptional changes that were also applicable to human patients.

Paper IV

A human iPSC model of KMT2A::AFF1 captures the developmental trajectory of pediatric acute B-cell leukemia

Studies of the earliest steps of leukemic transformation in pediatric ALL of *in utero* origin, such as infant KMT2A::AFF1 ALL, are hindered by the lack of model systems recapitulating human prenatal hematopoiesis. While it has been shown that introduction of the fusion gene in human fetal HSPCs can generate an aggressive B-lineage leukemia upon transplantation [136], human fetal tissue is scarce and comes with ethical concerns. In this study, we present an iPSC-based model of KMT2A::AFF1-driven ALL.

We first evaluated our culture system's ability to recapitulate prenatal hematopoiesis. We differentiated iPSCs to HSPCs and then co-cultured them with MS-5 stromal cells to allow for B-lineage differentiation. We sampled the cultures at day 0, meaning the point of harvest of CD34⁺ cells, and after 7 days of MS-5 co-culture. Mapping our data onto a reference compiled from pre- and postnatal human hematopoietic datasets [33, 94], we found that the majority of cells at day 0 mapped to YS, AGM and early FL populations. Furthermore, the dataset contained cells predicted to be HSPCs, MPPs and lymphoid progenitor cells, suggesting that populations of relevance to the prenatal initiation of B-lineage ALL could be generated using this culture system.

To study the effect of the chromosomal translocation on the cells, we used CRISPR/Cas9 to simultaneously induce double-stranded breaks in the *KMT2A* and *AFF1* loci. Maintaining the cells in MS-5 co-culture, we observed an increase over time in output of CD19⁺ cells compared to control cultures and an increase in proportion of fusion-carrying cells, indicating that mutated cells gained a proliferative advantage. Edited cultures also showed higher expression of *HOXA9*, *IRX1* and *HMGA2*, genes previously implicated in KMT2Ar leukemia. *IRX1* in particular is associated with a more aggressive subtype of infant KMT2A::AFF1 ALL [126].

Investigating different populations of the edited cultures, we found that the CD133⁺ (*PROM1*⁺) fraction was enriched in the CD34⁺ compartment. CD133 has been previously shown to be expressed in KMT2Ar leukemia [127] and *KMT2A::AFF1* expression was highest in this population in our cultures. Purifying CD133⁺ cells for further culture showed that cells from edited cultures produced higher proportions of CD133⁺ cells compared to controls. These results indicated that the fusion gene conferred increased self-renewal capacity to this population.

We next performed single-cell RNA-seq on cells from day 7 and day 28 of MS-5 co-culture. When projecting the data onto a healthy BM reference [125], no major differ-

ences in cell type distribution were seen at day 7, but by day 28 an increased proportion of HSPC-like and B cell progenitor-like cells were found in the KMT2A::AFF1 cultures. Differential gene expression analysis also revealed upregulation of the stem cell-associated *HMG2*, as well as *HOXB* genes, in CD133⁺ cells, overall indicating an immature population with fetal-like properties that was maintained over several weeks of culture.

Finally, we transplanted cells from day 7 of MS-5 co-culture into NSG mice implanted with humanized ossicles, providing a humanized niche for the cells [251]. While cas9 control cells did not show engraftment, mice transplanted with KMT2A::AFF1-edited cells succumbed to leukemia within ~5 months. In addition to enlarged spleens, histological examination showed that blasts had infiltrated the brain, a trait characteristic of infant KMT2Ar ALL [252]. In secondary intravenous transplantations, all four leukemias caused recipients to develop leukemia with shorter latency compared to the primary transplantations. Single-cell RNA-seq of leukemic cells from the BM of primary transplant recipients showed that they had very low expression of *HOXA9*. Instead, 3/4 leukemias expressed *IRX1*, indicating a subtype associated with particularly poor prognosis [127]. In agreement with the expression of *IRX1*, when mapped onto a reference dataset comprised of KMT2Ar infant ALL samples [125, 250], the iPSC-derived leukemias mainly projected to patients of the *IRX1* subtype.

In all, the results in this study establish the iPSC-based KMT2A::AFF1-driven ALL model as a platform for future investigation of leukemogenesis in a setting recapitulating the otherwise largely inaccessible early stages of human hematopoietic development.

Discussion

Properties of the fetal hematopoietic landscape have been implicated in the *in utero* origin of pediatric leukemia [100], but many aspects of human prenatal hematopoiesis and how it differs from that in adulthood are not fully explored. Through the work described in this thesis, we have addressed outstanding questions regarding both normal developmental hematopoiesis and leukemic transformation in the prenatal setting. Findings from these studies will be discussed below.

Using the combined transcriptome and surface marker data in Paper I, we showed that immunophenotypically defined cell populations are transcriptionally heterogeneous. In particular, CMP and GMP populations showed high heterogeneity. This is in line with other studies showing that human phenotypic CD38⁺ progenitor populations are mostly comprised of cells with different unilineage priming and potential [14]. We also found that surface expression of CD10 did not delineate the CLP-like Ly-II cluster in first-trimester FL, whereas IL7R was highly specific to this population. In agreement with this, a recent publication proposed that the fetal-specific CD10⁻ early lymphoid progenitor population is transcriptionally equivalent to the postnatal CLP [102].

Using a reference mapping approach, we found that when using adult BM data as the reference, most of the cells defined as HSCs in the FL were predicted to be MPPs instead of HSCs. This could be due to transcriptional differences between HSCs at different developmental stages, such as the fact that early FL HSCs have higher expression of genes associated with proliferation [33, 94] in comparison to the more quiescent adult BM [94]. When instead using the CS22 FL as the reference, we identified an MPP population that was present in FL from multiple stages of development, but nearly absent in adult BM. While it is possible that transcriptional differences between functionally equivalent populations also played a role here, an alternative explanation is that this cluster represents a fetal-specific, possibly pre-definitive wave progenitors, which would be in line with the finding that HSC-independent progenitor cells contribute substantially to FL hematopoiesis during murine development [42]. If a suitable surface marker combination could be identified, such a population could be of interest in the context of prenatal leukemia initiation.

By population-wise comparison of the CS22 and adult datasets, we were able to extract a list of genes that were upregulated in fetal or adult cell types, respectively. Applying this signature to a dataset covering both fetal and postnatal HSPCs [94], we found that the fetal signature decreased in the second trimester and was also slightly lower in fetal BM than FL from the same donor, while the adult-derived signature showed the opposite trend. The difference between the two niches at the same developmental stage suggests that the microenvironment may play a role in regulating the expression of these genes.

In Paper II, we further investigated the differences between fetal and adult HSPCs. Parallel single-HSC cultures showed that first-trimester FL HSCs had a higher percentage of multilineage clones compared to those from adult BM, which were mainly of unilineage myeloid output. We also found that the fetal HSC clones often contained B cells, in contrast to a previous study where very few B cells were generated by HSC/MPPs from 7-8 pcw FL [93], possibly due to differences in culture conditions.

At the chromatin level, we found adult BM HSCs to be enriched for myeloid and megakaryocyte-associated motifs, some of which also appeared to be differentially occupied by their corresponding TF. In line with this, a postnatal switch to increased chromatin-level megakaryocyte priming in HSCs has been previously reported in the murine system, and attributed to downregulation of the fetal-specific RNA-binding protein LIN28B [253]. Although we did not observe higher output of megakaryocytes from adult HSCs *in vitro*, human HSCs have been shown to become myeloid and megakaryocyte-biased with age [243, 254], and we also found that adult BM HSCs scored higher than those from FL for platelet bias at the transcriptional level. One possible explanation for this is that chromatin and transcriptional priming towards the megakaryocyte lineage emerges relatively early in life, but is not realized in the form of functional bias until much later.

Jacobsen and Nerlov previously pointed out the lack of transcriptional lymphoid priming in adult HSCs, proposing that future studies of the single-cell chromatin landscape might provide information about regulatory networks driving cell fate [7]. However, despite abundant B cell output *in vitro*, our analysis did not result in clear evidence of lymphoid chromatin priming in FL HSCs. One possible explanation is that lymphoid priming is in fact very low at the HSC stage, both transcriptionally and epigenetically, and that lymphopoiesis is dependent on a low degree of priming towards other lineages rather than a high degree of priming towards the lymphoid lineage. Another explanation is that HSC lymphoid priming is manifested through molecular mechanisms that are not yet fully elucidated. Further analyses of enhancer elements may provide insights here.

Seemingly contradictory to the high frequency of B-lineage positive HSC-derived colonies, we found the fetal lymphoid progenitor compartment to exhibit signs of reduced PAX5 activity, including the inaccessibility of putative enhancer elements that were accessible in adult counterparts. Likewise, while the FL contains a lot of erythroid cells [93], we did not see clear indications of erythroid priming in the HSCs, although these cells produced more erythroid colonies than adult HSCs *in vitro*. These discrepancies may be explained by the fact that at this stage of development, the liver contains a mixture of definitive HSCs and pre-definitive wave progenitor cells [19], and the progenitors profiled in our datasets may stem from HSC-independent waves.

Alongside the reduced PAX5 activity, we found indications of a retained myeloid profile

in the FL lymphoid progenitor populations. While we did not assess if these cells have higher functional myeloid output than adult counterparts, IL7R⁺ progenitor populations with B cell and myeloid potential have been reported early in both murine and human fetal development [25, 239]. The Ly-II cells in our CS22 FL ATAC-seq dataset are most likely IL7R⁺, based on mapping onto the CITE-seq dataset from Paper I (data not shown). In the mouse, this progenitor was shown to emerge from the YS prior to definitive HSCs [25]. Although the origin is more difficult to trace in the human setting, the fact that CD34⁺IL7R⁺ cells are present in the FL already at CS17 [239], very soon after the first HSCs have been proposed to migrate there [33], makes an HSC-independent origin likely. It is possible, however, that at CS22 this population also contains HSC-derived progenitors.

Prenatal development constitutes a period during which the whole hematopoietic system is established, and at the same time, an apparent vulnerability to certain malignancies. Despite therapeutic advances over the last few decades, KMT2A::AFF1-driven infant ALL remains a disease with poor prognosis [116]. Knowledge of the type of cell the disease originates from and what regulatory programs are active at the time could help in the development of therapeutic or preventive measures. A fetal B cell differentiation trajectory where myeloid potential is retained for longer could be of interest in the context of B-ALL of *in utero* origin, especially considering the lineage plasticity that KMT2A::AFF1 infant ALL demonstrates in its ability to relapse as a lineage-switched AML.

We showed in Paper I that when applied to publicly available patient data, our fetal gene signature was enriched in the youngest KMT2A::AFF1 patients. The hypothesis that a fetal gene expression program is the result of a fetal cell of origin has been presented before, as discussed in [255], although postnatal cells can also acquire fetal-like gene expression through oncofetal reprogramming upon mutation [256]. The fact that gene expression in overt leukemia cannot be assumed to directly correspond to the state of the cell of origin necessitates the use of model systems where the effect of the introduction of a leukemic driver mutation can be studied.

It has proven difficult to model KMT2A::AFF1 ALL in murine models, with most attempts failing to generate pro-B ALL. In the model we presented in Paper III, mice also did not develop overt leukemia, but a preleukemic preproB state displaying enhanced self-renewal and aberrant lineage plasticity, emerged when fusion gene expression was induced prenatally. This stage was absent when inducing expression postnatally, suggesting that a fetal HSPC state is required for transformation. In line with this, human CB and FL HSPCs have been shown to be more susceptible to transformation by KMT2A::AFF1 and NUP98::NSD1 than adult BM cells [140, 257], although neither of these studies investigated B-ALL. Additionally, niche age has been previously shown to play a role in other KMT2Ar acute leukemias, as a study showed that transplantation

into neonatal mice resulted in mixed myeloid and B-lineage leukemia, whereas the adult niche primarily promoted myeloid leukemia [258]. Together, these findings, along with our findings in Paper III, provide compelling arguments for studying the fusion gene in the developmental context where it occurs in infants with this disease.

Moreover, induction of the fusion gene at more lineage-committed progenitor stages did not result in a preleukemic phenotype, suggesting that even if the cell ultimately transforming into overt leukemia is a lymphoid progenitor, upstream events driven by the fusion protein may be required for this to occur. One reason for this could be that the *KMT2A::AFF1* requires access to certain target genes that are more active in primitive HSPCs, such as *Hox* genes and *Hmga2* [245, 259] to initiate a leukemogenic program.

Finally, the iPSC-derived model of leukemia initiation presented in Paper IV has several advantages over other methods. Compared to animal models, the model provides a promising avenue for investigation of the cell of origin, as populations of interest can be purified before or after genomic editing to assess the effect of the fusion gene on specific cell types. As opposed to CD34⁺ human FL cells, which have also successfully been edited using CRISPR/Cas9 to induce *KMT2A::AFF1*-driven B-ALL [136], iPSCs can be maintained in culture and expanded near-indefinitely and do not bring the same ethical challenges. Importantly, we demonstrated that most HSPCs derived from our culture system were transcriptionally similar to cells from early in the first trimester, where primary cells are nearly inaccessible to these types of perturbations due to their scarcity. On the other hand, this system does not recapitulate the prenatal niche that the murine model from Paper III provides.

Compared to, e.g. a retroviral transduction, the introduction of the fusion gene through simultaneous double-stranded breaks in both loci followed by spontaneous rearrangement has the advantage of better replicating the occurrence in the patient, resulting in the loss of one copy of the wild-type *KMT2A* gene and placing the fusion gene under the control of the endogenous *KMT2A* promoter. However, this also means that there is no marker through which successfully edited cells can be enriched. While we showed in Paper IV that the CD133⁺ compartment is enriched for the fusion gene, this population also exists in non-edited cultures, and CD133 can therefore not be considered a marker specific to fusion-positive cells. Therefore, the model relies on cells carrying the translocation out-competing the non-edited cells in culture.

We found that CD133⁺ cells from edited cultures showed increased expression of stem- and fetal-associated genes compared to primitive control cells, even after several weeks of MS-5 co-culture. This indicates that aspects of an early leukemic program are evident *in vitro*, which is important, as the early stages of transformation are difficult to study in an *in vivo* setting. In future studies, this model system may therefore provide new insights into early events during leukemogenesis.

All four mice transplanted with cells from KMT2A::AFF1 edited cultures developed CD19⁺CD10⁻ ALL, but showed variable surface expression of CD34, CD133 and IL7R. Furthermore, not all mice had clear expression of *IRX1*. The cells transplanted came from separate cuvettes of nucleofected cells, but all cells were from the same pool of CD34⁺ iPSC-derived HSPCs. This suggests some degree of stochasticity in the transformation event, where highly similar target cells can result in leukemias with a different molecular profile. Further investigation may determine the details of this.

In all, these works have led to new insights regarding the composition, priming and plasticity of the progenitor compartment in human hematopoietic development, as well as leukemogenic transformation. The new datasets and model systems presented here have the potential to contribute to future investigations of both normal and malignant hematopoiesis.

Future perspectives

At the moment, our understanding of and ability to manipulate hematopoiesis are developing rapidly through advances in single-cell omics, lineage tracing and *in vitro* generation of human HSCs.

The single-cell datasets generated from human fetal liver in Papers I and II allowed for a detailed dissection of the molecular properties of the early HSPC compartment, but the chromatin accessibility data was nevertheless produced from different cells, and different donors, than the transcriptome and surface marker data. As technologies advance towards lower cost and higher quality data, single-cell multiomic assays covering both transcriptome and chromatin accessibility will likely become standard procedure, allowing for more accurate inference of, e.g., gene regulatory networks through combined analysis of gene expression and accessibility of transcription factor binding sites in the same cell [244]. Already, assays such as DOGMA-seq, which allows for simultaneous profiling of RNA expression, chromatin accessibility and surface marker proteins in single cells, have been presented [260].

Though we found indications that certain cell populations in our datasets were derived from pre-definitive hematopoietic waves, we were not able to prove that. In the future, joint profiling of transcriptome or chromatin accessibility and, e.g., mitochondrial DNA [261] or epigenetic mutations [262] may be able to resolve relationships between and within different populations in the early fetal liver.

Finally, advances in *in vitro* culture of induced pluripotent stem cells have led to systems modeling different waves of extra- and intraembryonic human hematopoiesis [226], while bone marrow organoids strive to emulate the bone marrow niche [154]. These hold potential for dissection of the leukemogenic process with unprecedented levels of insight and control, which will likely lead to more accurate and precise modeling of disease initiation and development of therapeutic interventions.

In all, recent findings provide many promising avenues to pursue and, hopefully, eventually, a brighter outlook for the patients and their families suffering from a vicious disease.

Acknowledgements

Turns out, doing a PhD is not so easy and can actually be quite a lot of work, and some days you might even wonder if everything is actually going to work out. Therefore, I am very grateful for the people who have helped make this possible, either by directly contributing to the work presented in this thesis or by supporting me in other ways.

First and foremost, I would like to thank my supervisor **Charlotta**. Thank you for giving me the opportunity to learn from you and for your support when things were anything but easy, I couldn't have asked for a better start to my scientific career. Your scientific rigor, your passion for teaching and dedication to your lab will always be an inspiration to me. **Shamit**, you have gone above and beyond the role of co-supervisor you have sometimes jokingly (?) told me you wish you didn't accept. Thank you for everything you've taught me about bioinformatics, for giving me a second workspace and for all the conversations about important and, perhaps most importantly, unimportant things. Thank you also to my two other co-supervisors **Göran** and **Virginia**, for all your valuable feedback throughout the years, nice conversations, and for your contributions to the papers.

Ariana, I'm happy I got to share my PhD journey with you. Whether it's venting lab frustrations, painting ceramics or climbing the stairs of a Milanese cathedral during a heatwave, it's been great to have you by my side. **Roshanak**, what would I have done without you? Thank you for your patience, kindness and everything you taught me. **Keiki**, thank you for all your insightful comments, kind words and your hard work, especially on the ATAC-seq and CRISPR projects. **Mikael**, working on your CITE-seq paper was a great introduction to my PhD work, thank you for your help when everything from developmental hematopoiesis to R programming was new to me.

The FACS core members, **Anna F**, **Zhi**, **Helena** and **Isabel**, thank you for your invaluable work in keeping the instruments running and for everything you taught me about flow cytometry and sorting. **David**, thank you for all the manuscript feedback, your contributions to paper III, and all your other insightful comments over the years. **Linda** and **Giorgia**, thank you for your kindness and for all your assistance with the single cell omics experiment. **Stefan**, thank you for your insights regarding bioinformatics analysis and for your work on paper III. Thank you also to all the other collaborators who have contributed to the papers included in this thesis.

Thank you also to all other current and former members of **BMC B12** who have helped me, shared the lunch table with me, given me feedback, and in general provided a supportive, interactive and fun environment. I have had a great time at this department during my PhD, and you have all contributed to this!

Elena, thank you for being a great gym group co-organizer, hotel roommate, movie co-producer, advice-giver, and friend. These last few years would have been kind of awful without you. **Johanna**, thank you for letting me be a little bit insane when needed, and for all the times you've taken on the burden of being the responsible one. **Rasmus**, thank you for all the hangouts, conversations about anything nerdy and more or less horrible puns (and for your work on Papers I and II). **Oscar**, thank you for all the post-lunch games, and all the google maps and wiki deep dives. They surely made me a better researcher. **Sandro**, thank you for being a driving force in the social interactions at B12 during the better part of my PhD, and for the honor of letting me co-host the now-mythological annual mario kart tournament. **Anna R**, thank you for being a generally great person to be around, whether in the gym or exploring Japan. **Julie**, thank you for always bringing the mood up and for never being afraid to speak your mind. **Jack**, you very quickly found your place in the group, thank you for helping make sure fun things kept happening during the late stages of my PhD. **Chun-Chi**, thank you for being such a cheerful, kind person, having you around always brightens the mood.

To everyone who has been part of my personal BMC B12 extended universe; Elena, Johanna, Rasmus, Oscar, Sandro, Julie, Anna R, Jack, Chun-Chi, Ariana, **Niklas**, **Mo**, **Qinyu**, **Ania**, **Anna P**, **Simone**, **Roger**, **Joaquin**, **Martino** and all the others, thank you for all the workouts, saunas, concerts, whiskey thursdays, barbecues, karaoke, mario kart tournaments, bar hangouts and miscellaneous adventures. Thanks to you I not only survived these last five years, but also had a lot of fun.

Jana, thank you for all the weeknight dinners, reality TV marathons, climbing parks and other excursions, and some truly truly awful movies watched. Sometimes we have very different ideas about what counts as recovering (for example, one of us may want to go for a run when she gets home from a hike, and the other one, no names mentioned, might prefer to collapse on the couch), but hanging out with you, and lately also **Maeve**, has always been a great respite from the stress of work for me. **Louise**, twelve years after we met during my first week living in Lund, thus both completing the mission of making a friend in our new class, I'm happy to still call you my friend and that we never run out of things to talk about. **Johan**, I can't believe you're not here anymore, you live on in many memories and some of my best anecdotes. **Susanna**, we haven't lived in the same place since we were 20, but we have managed to remain in each other's lives, and whenever we do see each other it's as if barely any time has passed. Thank you for all the laughs and support through various life crises. **Lena**, I always feel like I could tell you anything. Thank you for being such an honest, supportive friend. **Elin and Joel**, thank you for adopting me into the friend group, together with **Oscar I**, **Nico** and **Eric**, and for all the dinners and board game nights, and for always providing such a warm, welcoming home. **Albert**, thank you for all the Ariman Fridays and subsequent falafels, it's always something I look forward to after a long week. **Per** and **Johannes**, together

with Rasmus and Oscar L, thank you for providing an escape into the table-top RPG world. One day we will manage to schedule the sessions close enough in time that we actually remember what happened the time before!

Till min familj: mina föräldrar, **Anita** och **Leif**, och till **Linn**, **Magnus**, **Svante**, **Hedda**, **Katja**, **Vincent** och **Vilda**, tack för allt ert stöd. Tack för att ni alltid har trott på mig, för att ni varit förstående när jag kanske inte alltid varit den mest närvarande dottern, systemen eller mostern de senaste åren, och för att ni alltid varit stolta över mig. Och till min mormor **Eva**: ifall jag aldrig vinner det där Nobelpriset och den här avhandlingen blir min största akademiska bedrift, så har jag för säkerhets skull dedikerat den till dig.

And finally, **Jaime**, I'm so glad we didn't write our theses at the same time. Thank you for always being my biggest fan, for always showing interest in and asking such great questions about my work even though it's so far from your own, and for how supportive you've been during these last, very intense, months. I look forward to going back to spending time together that doesn't involve me staring at my laptop screen. Thank you for everything.

References

1. Till, J. E. & McCulloch, E. A. A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. *Radiation Research* **14**, 213–222. doi:10.2307/3570892 (1961).
2. Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological Demonstration of the Clonal Nature of Spleen Colonies Derived from Transplanted Mouse Marrow Cells. *Nature* **197**, 452–454. doi:10.1038/197452a0 (1963).
3. Siminovitch, L., McCulloch, E. A. & Till, J. E. The distribution of colony-forming cells among spleen colonies. *Journal of Cellular and Comparative Physiology* **62**, 327–336. doi:10.1002/jcp.1030620313 (1963).
4. Osawa, M., Hanada, K.-i., Hamada, H. & Nakauchi, H. Long-Term Lymphohematopoietic Reconstitution by a Single CD34-Low/Negative Hematopoietic Stem Cell. *Science* **273**, 242–245. doi:10.1126/science.273.5272.242 (1996).
5. Orkin, S. H. & Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **132**, 631–644. doi:10.1016/j.cell.2008.01.025 (2008).
6. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426. doi:10.1038/nature25022 (2018).
7. Jacobsen, S. E. W. & Nerlov, C. Haematopoiesis in the era of advanced single-cell technologies. *Nature Cell Biology* **21**, 2–8. doi:10.1038/s41556-018-0227-8 (2019).
8. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: A Human Perspective. *Cell Stem Cell* **10**, 120–136. doi:10.1016/j.stem.2012.01.006 (2012).
9. Pietras, E. M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35–46. doi:10.1016/j.stem.2015.05.003 (2015).
10. Pronk, C. J. H. *et al.* Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy. *Cell Stem Cell* **1**, 428–442. doi:10.1016/j.stem.2007.07.005 (2007).
11. Görgens, A. *et al.* Revision of the Human Hematopoietic Tree: Granulocyte Subtypes Derive from Distinct Hematopoietic Lineages. *Cell Reports* **3**, 1539–1552. doi:10.1016/j.celrep.2013.04.025 (2013).
12. Rodriguez-Fraticelli, A. E. *et al.* Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216. doi:10.1038/nature25168 (2018).
13. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116. doi:10.1126/science.aab2116 (2016).
14. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology* **19**, 271–281. doi:10.1038/ncb3493 (2017).
15. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31. doi:10.1182/blood-2016-05-716480 (2016).

16. Liggett, L. A. & Sankaran, V. G. Unraveling Hematopoiesis through the Lens of Genomics. *Cell* **182**, 1384–1400. doi:10.1016/j.cell.2020.08.030 (2020).
17. Haas, S., Trumpp, A. & Milsom, M. D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell* **22**, 627–638. doi:10.1016/j.stem.2018.04.003 (2018).
18. Karlsson, G., Sommarin, M. N. E. & Böiers, C. Defining the Emerging Blood System During Development at Single-Cell Resolution. *Frontiers in Cell and Developmental Biology* **9**. doi:10.3389/fcell.2021.660350 (2021).
19. Calvanese, V. & Mikkola, H. K. A. The genesis of human hematopoietic stem cells. *Blood* **142**, 519–532. doi:10.1182/blood.2022017934 (2023).
20. Bloom, W. & Bartelmez, G. W. Hematopoiesis in young human embryos. *American Journal of Anatomy* **67**, 21–53. doi:10.1002/aja.1000670103 (1940).
21. Luckett, W. P. Origin and differentiation of the yolk sac and extraembryonic mesoderm in presomite human and rhesus monkey embryos. *American Journal of Anatomy* **152**, 59–97. doi:10.1002/aja.1001520106 (1978).
22. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289. doi:10.1038/s41586-021-04158-y (2021).
23. Tavian, M., Hallais, M.-F. & Péault, B. Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development* **126**, 793–803. doi:10.1242/dev.126.4.793 (1999).
24. McGrath, K. E. *et al.* Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Reports* **11**, 1892–1904. doi:10.1016/j.celrep.2015.05.036 (2015).
25. Böiers, C. *et al.* Lymphomyeloid Contribution of an Immune-Restricted Progenitor Emerging Prior to Definitive Hematopoietic Stem Cells. *Cell Stem Cell* **13**, 535–548. doi:10.1016/j.stem.2013.08.012 (2013).
26. Bian, Z. *et al.* Deciphering human macrophage development at single-cell resolution. *Nature* **582**, 571–576. doi:10.1038/s41586-020-2316-7 (2020).
27. Ni, Y. *et al.* Human yolk sac-derived innate lymphoid-biased multipotent progenitors emerge prior to hematopoietic stem cell formation. *Developmental Cell* **59**, 2626–2642.e6. doi:10.1016/j.devcel.2024.06.010 (2024).
28. Tavian, M. *et al.* Aorta-associated CD34⁺ hematopoietic cells in the early human embryo. *Blood* **87**, 67–72. doi:10.1182/blood.V87.1.67.67 (1996).
29. Ivanovs, A. *et al.* Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *Journal of Experimental Medicine* **208**, 2417–2427. doi:10.1084/jem.20111688 (2011).
30. Gekas, C. *et al.* Hematopoietic stem cell development in the placenta. *The International Journal of Developmental Biology* **54**, 1089–1098. doi:10.1387/ijdb.103070cg (2010).
31. Prashad, S. L. *et al.* GPI-80 Defines Self-Renewal Ability in Hematopoietic Stem Cells during Human Development. *Cell Stem Cell* **16**, 80–87. doi:10.1016/j.stem.2014.10.020 (2015).

32. 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. doi:10.1016/j.stem.2009.08.020 (2009).
33. Calvanese, V. *et al.* Mapping human haematopoietic stem cells from haemogenic endothelium to birth. *Nature* **604**, 534–540. doi:10.1038/s41586-022-04571-x (2022).
34. 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 Reports* **15**, 811–816. doi:10.1016/j.stemcr.2020.08.008 (2020).
35. Oberlin, E. *et al.* VE-cadherin expression allows identification of a new class of hematopoietic stem cells within human embryonic liver. *Blood* **116**, 4444–4455. doi:10.1182/blood-2010-03-272625 (2010).
36. Zhang, Y. *et al.* VE-Cadherin and ACE Co-Expression Marks Highly Proliferative Hematopoietic Stem Cells in Human Embryonic Liver. *Stem Cells and Development* **28**, 165–185. doi:10.1089/scd.2018.0154 (2019).
37. Charbord, P., Tavian, M., Humeau, L. & Peault, B. Early Ontogeny of the Human Marrow From Long Bones: An Immunohistochemical Study of Hematopoiesis and Its Microenvironment. *Blood* **87**, 4109–4119. doi:10.1182/blood.V87.10.4109.bloodjournal187104109 (1996).
38. Zheng, Z. *et al.* Uncovering the emergence of HSCs in the human fetal bone marrow by single-cell RNA-seq analysis. *Cell Stem Cell* **29**, 1562–1579.e7. doi:10.1016/j.stem.2022.10.005 (2022).
39. Ginhoux, F. *et al.* Fate Mapping Analysis Reveals That Adult Microglia Derive from Primitive Macrophages. *Science* **330**, 841–845. doi:10.1126/science.1194637 (2010).
40. Qu, G. *et al.* Comparing Mouse and Human Tissue-Resident $\gamma\delta$ T Cells. *Frontiers in Immunology* **13**. doi:10.3389/fimmu.2022.891687 (2022).
41. Mattos, M. S., Vandendriessche, S., Waisman, A. & Marques, P. E. The immunology of B-1 cells: from development to aging. *Immunity & Ageing* **21**, 54. doi:10.1186/s12979-024-00455-y (2024).
42. Yokomizo, T. *et al.* Independent origins of fetal liver haematopoietic stem and progenitor cells. *Nature* **609**, 779–784. doi:10.1038/s41586-022-05203-0 (2022).
43. De Graaf, C. A. & Metcalf, D. Thrombopoietin and hematopoietic stem cells. *Cell Cycle* **10**, 1582–1589. doi:10.4161/cc.10.10.15619 (2011).
44. Kent, D. *et al.* Regulation of Hematopoietic Stem Cells by the Steel Factor/KIT Signaling Pathway. *Clinical Cancer Research* **14**, 1926–1930. doi:10.1158/1078-0432.CCR-07-5134 (2008).
45. Huang, Z. *et al.* Cellular crosstalk in the bone marrow niche. *Journal of Translational Medicine* **22**, 1096. doi:10.1186/s12967-024-05900-6 (2024).
46. Metcalf, D. Hematopoietic cytokines. *Blood* **111**, 485–491. doi:10.1182/blood-2007-03-079681 (2008).

47. Hamilton, J. A. & Achuthan, A. Colony stimulating factors and myeloid cell biology in health and disease. *Trends in Immunology* **34**, 81–89. doi:10.1016/j.it.2012.08.006 (2013).
48. Tsiftoglou, A. S. Erythropoietin (EPO) as a Key Regulator of Erythropoiesis, Bone Remodeling and Endothelial Transdifferentiation of Multipotent Mesenchymal Stem Cells (MSCs): Implications in Regenerative Medicine. *Cells* **10**, 2140. doi:10.3390/cells10082140 (2021).
49. Behrens, K. & Alexander, W. S. Cytokine control of megakaryopoiesis. *Growth Factors* **36**, 89–103. doi:10.1080/08977194.2018.1498487 (2018).
50. Tsapogas, P., Mooney, C. J., Brown, G. & Rolink, A. The Cytokine Flt3-Ligand in Normal and Malignant Hematopoiesis. *International Journal of Molecular Sciences* **18**, 1115. doi:10.3390/ijms18061115 (2017).
51. McKenna, H. J. *et al.* Mice lacking flt3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells. *Blood* **95**, 3489–3497. doi:10.1182/blood.V95.11.3489 (2000).
52. Carvalho, T. L., Mota-Santos, T., Cumano, A., Demengeot, J. & Vieira, P. Arrested B Lymphopoiesis and Persistence of Activated B Cells in Adult Interleukin 7^{-/-} Mice. *Journal of Experimental Medicine* **194**, 1141–1150. doi:10.1084/jem.194.8.1141 (2001).
53. Jensen, C. T. *et al.* FLT3 ligand and not TSLP is the key regulator of IL-7-independent B-1 and B-2 B lymphopoiesis. *Blood* **112**, 2297–2304. doi:10.1182/blood-2008-04-150508 (2008).
54. Marr, B., Jo, D., Jang, M. & Lee, S.-H. Cytokines in Focus: IL-2 and IL-15 in NK Adoptive Cell Cancer Immunotherapy. *Immune Network* **25**, e17. doi:10.4110/in.2025.25.e17 (2025).
55. Hosokawa, H. & Rothenberg, E. V. Cytokines, Transcription Factors, and the Initiation of T-Cell Development. *Cold Spring Harbor Perspectives in Biology* **10**, a028621. doi:10.1101/cshperspect.a028621 (2018).
56. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics* **13**, 613–626. doi:10.1038/nrg3207 (2012).
57. Lehnertz, B. *et al.* HLF expression defines the human hematopoietic stem cell state. *Blood* **138**, 2642–2654. doi:10.1182/blood.2021010745 (2021).
58. Calvanese, V. *et al.* MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. *Nature* **576**, 281–286. doi:10.1038/s41586-019-1790-2 (2019).
59. Peters, I. J. A., de Pater, E. & Zhang, W. The role of GATA2 in adult hematopoiesis and cell fate determination. *Frontiers in Cell and Developmental Biology* **11**. doi:10.3389/fcell.2023.1250827 (2023).
60. Voit, R. A. *et al.* A genetic disorder reveals a hematopoietic stem cell regulatory network co-opted in leukemia. *Nature Immunology* **24**, 69–83. doi:10.1038/s41590-022-01370-4 (2023).

61. Ditadi, A. *et al.* Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages. *Nature Cell Biology* **17**, 580–591. doi:10.1038/ncb3161 (2015).
62. Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891. doi:10.1038/nature07619 (2009).
63. Dou, D. R. *et al.* Medial HOXA genes demarcate haematopoietic stem cell fate during human development. *Nature Cell Biology* **18**, 595–606. doi:10.1038/ncb3354 (2016).
64. Ng, E. S. *et al.* Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. *Nature Biotechnology* **34**, 1168–1179. doi:10.1038/nbt.3702 (2016).
65. Kim, S.-I. & Bresnick, E. H. Transcriptional control of erythropoiesis: emerging mechanisms and principles. *Oncogene* **26**, 6777–6794. doi:10.1038/sj.onc.1210761 (2007).
66. Iwasaki, H. *et al.* GATA-1 Converts Lymphoid and Myelomonocytic Progenitors into the Megakaryocyte/Erythrocyte Lineages. *Immunity* **19**, 451–462. doi:10.1016/S1074-7613(03)00242-5 (2003).
67. Siatecka, M. & Bieker, J. J. The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood* **118**, 2044–2054. doi:10.1182/blood-2011-03-331371 (2011).
68. Gasiorrek, J. J. & Blank, V. Regulation and function of the NFE2 transcription factor in hematopoietic and non-hematopoietic cells. *Cellular and Molecular Life Sciences* **72**, 2323–2335. doi:10.1007/s00018-015-1866-6 (2015).
69. Iwasaki, H. & Akashi, K. Myeloid Lineage Commitment from the Hematopoietic Stem Cell. *Immunity* **26**, 726–740. doi:10.1016/j.immuni.2007.06.004 (2007).
70. Burda, P., Laslo, P. & Stopka, T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia* **24**, 1249–1257. doi:10.1038/leu.2010.104 (2010).
71. Hoppe, P. S. *et al.* Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* **535**, 299–302. doi:10.1038/nature18320 (2016).
72. Strasser, M. K. *et al.* Lineage marker synchrony in hematopoietic genealogies refutes the PU.1/GATA1 toggle switch paradigm. *Nature Communications* **9**, 2697. doi:10.1038/s41467-018-05037-3 (2018).
73. Friedman, A. D. Transcriptional control of granulocyte and monocyte development. *Oncogene* **26**, 6816–6828. doi:10.1038/sj.onc.1210764 (2007).
74. Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nature Reviews Immunology* **7**, 105–117. doi:10.1038/nri2024 (2007).
75. Nutt, S. L. & Kee, B. L. The Transcriptional Regulation of B Cell Lineage Commitment. *Immunity* **26**, 715–725. doi:10.1016/j.immuni.2007.05.010 (2007).

76. Welinder, E., Åhsberg, J. & Sigvardsson, M. B-lymphocyte commitment: Identifying the point of no return. *Seminars in Immunology. Parsing Gene Networks Underpinning Immune System Development* **23**, 335–340. doi:10.1016/j.smim.2011.08.005 (2011).
77. Hosokawa, H. & Rothenberg, E. V. How transcription factors drive choice of the T cell fate. *Nature Reviews Immunology* **21**, 162–176. doi:10.1038/s41577-020-00426-6 (2021).
78. Barral, A. & Déjardin, J. The chromatin signatures of enhancers and their dynamic regulation. *Nucleus* **14**, 2160551. doi:10.1080/19491034.2022.2160551 (2023).
79. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics* **12**, 283–293. doi:10.1038/nrg2957 (2011).
80. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology* **16**, 144–154. doi:10.1038/nrm3949 (2015).
81. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology* **20**, 590–607. doi:10.1038/s41580-019-0159-6 (2019).
82. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Reviews Genetics* **14**, 204–220. doi:10.1038/nrg3354 (2013).
83. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490. doi:10.1038/nature14590 (2015).
84. Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535–1548.e16. doi:10.1016/j.cell.2018.03.074 (2018).
85. Lara-Astiaso, D. *et al.* Chromatin state dynamics during blood formation. *Science* **345**, 943–949. doi:10.1126/science.1256271 (2014).
86. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell* **183**, 1103–1116.e20. doi:10.1016/j.cell.2020.09.056 (2020).
87. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics* **48**, 1193–1203. doi:10.1038/ng.3646 (2016).
88. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550. doi:10.1016/j.cell.2016.03.014 (2016).
89. Furtwängler, B. *et al.* Mapping early human blood cell differentiation using single-cell proteomics and transcriptomics. *Science* **390**, eadr8785. doi:10.1126/science.adr8785 (2025).
90. Jardine, L. *et al.* Blood and immune development in human fetal bone marrow and Down syndrome. *Nature* **598**, 327–331. doi:10.1038/s41586-021-03929-x (2021).
91. Marderstein, A. R. *et al.* Single-cell multi-omics map of human fetal blood in Down syndrome. *Nature* **634**, 104–112. doi:10.1038/s41586-024-07946-4 (2024).

92. Ranzoni, A. M. *et al.* Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. *Cell Stem Cell* **28**, 472–487.e7. doi:10.1016/j.stem.2020.11.015 (2021).
93. Popescu, D.-M. *et al.* Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371. doi:10.1038/s41586-019-1652-y (2019).
94. Roy, A. *et al.* Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development. *Cell Reports* **36**, 109698. doi:10.1016/j.celrep.2021.109698 (2021).
95. Li, H. *et al.* The dynamics of hematopoiesis over the human lifespan. *Nature Methods* **22**, 422–434. doi:10.1038/s41592-024-02495-0 (2025).
96. Fischer, U. *et al.* Cell Fate Decisions: The Role of Transcription Factors in Early B-cell Development and Leukemia. *Blood Cancer Discovery* **1**, 224–233. doi:10.1158/2643-3230.BCD-20-0011 (2020).
97. Korzhenevich, J., Janowska, I., van der Burg, M. & Rizzi, M. Human and mouse early B cell development: So similar but so different. *Immunology Letters* **261**, 1–12. doi:10.1016/j.imlet.2023.07.004 (2023).
98. Blom, B. & Spits, H. Development of human lymphoid cells. *Annual Review of Immunology* **24**, 287–320. doi:10.1146/annurev.immunol.24.021605.090612 (2006).
99. Chi, X., Li, Y. & Qiu, X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology* **160**, 233–247. doi:10.1111/imm.13176 (2020).
100. Jackson, T. R., Ling, R. E. & Roy, A. The Origin of B-cells: Human Fetal B Cell Development and Implications for the Pathogenesis of Childhood Acute Lymphoblastic Leukemia. *Frontiers in Immunology* **12**. doi:10.3389/fimmu.2021.637975 (2021).
101. O’Byrne, S. *et al.* Discovery of a CD10-negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs. *Blood* **134**, 1059–1071. doi:10.1182/blood.2019001289 (2019).
102. Iacobucci, I. *et al.* Multipotent lineage potential in B cell acute lymphoblastic leukemia is associated with distinct cellular origins and clinical features. *Nature Cancer* **6**, 1242–1262. doi:10.1038/s43018-025-00987-2 (2025).
103. Kageyama, Y. & Katayama, N. Ontogeny of human B1 cells. *International Journal of Hematology* **111**, 628–633. doi:10.1007/s12185-019-02775-y (2020).
104. Howlader, N. *et al.* *SEER Cancer Statistics Review, 1975–2012* tech. rep. (National Cancer Institute, Bethesda, MD, https://seer.cancer.gov/archive/csr/1975_2012/, based on November 2014 SEER data submission, posted to the SEER web site, April 2015., 2015).
105. Hein, D., Borkhardt, A. & Fischer, U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. *Cancer and Metastasis Reviews* **39**, 161–171. doi:10.1007/s10555-019-09841-1 (2020).
106. Harrison, C. J. Cytogenetics of paediatric and adolescent acute lymphoblastic leukaemia. *British Journal of Haematology* **144**, 147–156. doi:10.1111/j.1365-2141.2008.07417.x (2009).

107. Iacobucci, I. & Mullighan, C. G. A genomic and epigenomic lens into the biology of acute lymphoblastic leukaemia. *Nature Reviews Cancer* **26**, 689–712. doi:10.1038/s41568-026-00951-x (2026).
108. Hunger, S. P. & Mullighan, C. G. Acute Lymphoblastic Leukemia in Children. *New England Journal of Medicine* **373**, 1541–1552. doi:10.1056/NEJMr1400972 (2015).
109. Gale, K. B. *et al.* Backtracking leukemia to birth: Identification of clonotypic gene fusion sequences in neonatal blood spots. *Proceedings of the National Academy of Sciences* **94**, 13950–13954. doi:10.1073/pnas.94.25.13950 (1997).
110. Wiemels, J. L. *et al.* Prenatal origin of acute lymphoblastic leukaemia in children. *The Lancet* **354**, 1499–1503. doi:10.1016/S0140-6736(99)09403-9 (1999).
111. Ford, A. M. *et al.* In utero rearrangements in the trithorax-related oncogene in infant leukaemias. *Nature* **363**, 358–360. doi:10.1038/363358a0 (1993).
112. Alpar, D. *et al.* Clonal origins of ETV6-RUNX1+ acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia* **29**, 839–846. doi:10.1038/leu.2014.322 (2015).
113. Mori, H. *et al.* Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Sciences* **99**, 8242–8247. doi:10.1073/pnas.112218799 (2002).
114. Schäfer, D. *et al.* Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. *Blood* **131**, 821–826. doi:10.1182/blood-2017-09-808402 (2018).
115. Greaves, M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nature Reviews Cancer* **18**, 471–484. doi:10.1038/s41568-018-0015-6 (2018).
116. Rice, S. & Roy, A. MLL-rearranged infant leukaemia: A ‘thorn in the side’ of a remarkable success story. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**, 194564. doi:10.1016/j.bbagr.2020.194564 (2020).
117. Schuettengruber, B., Martinez, A.-M., Iovino, N. & Cavalli, G. Trithorax group proteins: switching genes on and keeping them active. *Nature Reviews Molecular Cell Biology* **12**, 799–814. doi:10.1038/nrm3230 (2011).
118. Meyer, C. *et al.* The KMT2A recombinoome of acute leukemias in 2023. *Leukemia* **37**, 988–1005. doi:10.1038/s41375-023-01877-1 (2023).
119. Winters, A. C. & Bernt, K. M. MLL-Rearranged Leukemias—An Update on Science and Clinical Approaches. *Frontiers in Pediatrics* **5**. doi:10.3389/fped.2017.00004 (2017).
120. Milne, T. A. Chromatin and aberrant enhancer activity in KMT2A rearranged acute lymphoblastic leukemia. *Current Opinion in Genetics & Development* **86**, 102191. doi:10.1016/j.gde.2024.102191 (2024).
121. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758–764. doi:10.1038/nature05690 (2007).
122. Andersson, A. K. *et al.* The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nature Genetics* **47**, 330–337. doi:10.1038/ng.3230 (2015).

123. Silbert, S. K. *et al.* Project EVOLVE: an international analysis of postimmunotherapy lineage switch, an emergent form of relapse in leukemia. *Blood* **146**, 437–455. doi:10.1182/blood.2024026655 (2025).
124. Tirtakusuma, R. *et al.* Epigenetic regulator genes direct lineage switching in MLL/AF4 leukemia. *Blood* **140**, 1875–1890. doi:10.1182/blood.2021015036 (2022).
125. Chen, C. *et al.* Single-cell multiomics reveals increased plasticity, resistant populations, and stem-cell-like blasts in KMT2A-rearranged leukemia. *Blood* **139**, 2198–2211. doi:10.1182/blood.2021013442 (2022).
126. Symeonidou, V. & Ottersbach, K. HOXA9/IRX1 expression pattern defines two subgroups of infant MLL-AF4-driven acute lymphoblastic leukemia. *Experimental Hematology* **93**, 38–43.e5. doi:10.1016/j.exphem.2020.10.002 (2021).
127. Agraz-Doblas, A. *et al.* Unraveling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. *Haematologica* **104**, 1176–1188. doi:10.3324/haematol.2018.206375 (2019).
128. Isobe, T. *et al.* Multi-omics analysis defines highly refractory RAS burdened immature subgroup of infant acute lymphoblastic leukemia. *Nature Communications* **13**, 4501. doi:10.1038/s41467-022-32266-4 (2022).
129. Kühn, A., Löscher, D. & Marschalek, R. The IRX1/HOXA connection: insights into a novel t(4;11)-specific cancer mechanism. *Oncotarget* **7**, 35341–35352. doi:10.18632/oncotarget.9241 (2016).
130. Milne, T. A. Mouse models of MLL leukemia: recapitulating the human disease. *Blood* **129**, 2217–2223. doi:10.1182/blood-2016-10-691428 (2017).
131. Malouf, C. *et al.* miR-130b and miR-128a are essential lineage-specific codrivers of t(4;11) MLL-AF4 acute leukemia. *Blood* **138**, 2066–2092. doi:10.1182/blood.202006610 (2021).
132. Bueno, C. *et al.* A human ESC model for MLL-AF4 leukemic fusion gene reveals an impaired early hematopoietic-endothelial specification. *Cell Research* **22**, 986–1002. doi:10.1038/cr.2012.4 (2012).
133. Tan, Y.-T. *et al.* Respecifying human iPSC-derived blood cells into highly engraftable hematopoietic stem and progenitor cells with a single factor. *Proceedings of the National Academy of Sciences* **115**, 2180–2185. doi:10.1073/pnas.1718446115 (2018).
134. Lin, S. *et al.* Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia. *Cancer Cell* **30**, 737–749. doi:10.1016/j.cccell.2016.10.008 (2016).
135. Secker, K.-A. *et al.* Inhibition of DOT1L and PRMT5 promote synergistic anti-tumor activity in a human MLL leukemia model induced by CRISPR/Cas9. *Oncogene* **38**, 7181–7195. doi:10.1038/s41388-019-0937-9 (2019).
136. Rice, S. *et al.* A human fetal liver-derived infant MLL-AF4 acute lymphoblastic leukemia model reveals a distinct fetal gene expression program. *Nature Communications* **12**, 6905. doi:10.1038/s41467-021-27270-z (2021).

137. Pan, F. *et al.* Genome editing–induced t(4;11) chromosomal translocations model B cell precursor acute lymphoblastic leukemias with KMT2A-AFF1 fusion. *The Journal of Clinical Investigation* **134**. doi:10.1172/JCI171030 (2024).
138. Benz, T. *et al.* CRISPR/Cas9-mediated t(4;11) translocation in human hematopoietic stem/precursor cells demonstrates plasticity to differentiate into either the myeloid or lymphoid lineage. *Leukemia* **40**, 72–86. doi:10.1038/s41375-025-02791-4 (2026).
139. Ling, R. E. *et al.* The fetal specific gene LIN28B is essential for human fetal B-lymphopoiesis and initiation of KMT2A::AFF1 infant leukemia. *Blood*, blood.2025028877. doi:10.1182/blood.2025028877 (2026).
140. Secker, K.-A. *et al.* Only Hematopoietic Stem and Progenitor Cells from Cord Blood Are Susceptible to Malignant Transformation by MLL-AF4 Translocations. *Cancers* **12**, 1487. doi:10.3390/cancers12061487 (2020).
141. Horton, S. J. *et al.* MLL–AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. *Leukemia* **27**, 1116–1126. doi:10.1038/leu.2012.343 (2013).
142. Bueno, C. *et al.* A human genome editing–based *MLL::AF4* ALL model recapitulates key cellular and molecular leukemogenic features. *Blood* **142**, 1752–1756. doi:10.1182/blood.2023020858 (2023).
143. Subramaniam, A., Talkhoncheh, M. S., Magnusson, M. & Larsson, J. Endothelial protein C receptor (EPCR) expression marks human fetal liver hematopoietic stem cells. *Haematologica* **104**, e47–e50. doi:10.3324/haematol.2018.198515 (2019).
144. Purton, L. E. Adult murine hematopoietic stem cells and progenitors: an update on their identities, functions, and assays. *Experimental Hematology* **116**, 1–14. doi:10.1016/j.exphem.2022.10.005 (2022).
145. Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. *Cell Stem Cell* **1**, 635–645. doi:10.1016/j.stem.2007.10.001 (2007).
146. Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nature Immunology* **11**, 585–593. doi:10.1038/ni.1889 (2010).
147. Iscove, N. N., Sieber, F. & Winterhalter, K. H. Erythroid colony formation in cultures of mouse and human bone marrow: Analysis of the requirement for erythropoietin by gel filtration and affinity chromatography on agarose-concanavalin A†. *Journal of Cellular Physiology* **83**, 309–320. doi:10.1002/jcp.1040830218 (1974).
148. Humphries, R. K., Eaves, A. C. & Eaves, C. J. Self-renewal of hemopoietic stem cells during mixed colony formation in vitro. *Proceedings of the National Academy of Sciences* **78**, 3629–3633. doi:10.1073/pnas.78.6.3629 (1981).
149. Nakahata, T. & Ogawa, M. Identification in culture of a class of hemopoietic colony-forming units with extensive capability to self-renew and generate multipotential hemopoietic colonies. *Proceedings of the National Academy of Sciences* **79**, 3843–3847. doi:10.1073/pnas.79.12.3843 (1982).

150. Purton, L. E. & Scadden, D. T. Limiting Factors in Murine Hematopoietic Stem Cell Assays. *Cell Stem Cell* **1**, 263–270. doi:10.1016/j.stem.2007.08.016 (2007).
151. Nakano, T., Kodama, H. & Honjo, T. Generation of Lymphohematopoietic Cells from Embryonic Stem Cells in Culture. *Science* **265**, 1098–1101. doi:10.1126/science.8066449 (1994).
152. Schmitt, T. M. & Zúñiga-Pflücker, J. C. Induction of T Cell Development from Hematopoietic Progenitor Cells by Delta-like-1 In Vitro. *Immunity* **17**, 749–756. doi:10.1016/S1074-7613(02)00474-0 (2002).
153. Berardi, A. C. *et al.* Individual CD34⁺CD38^{low}CD19[–]CD10[–] Progenitor Cells From Human Cord Blood Generate B Lymphocytes and Granulocytes. *Blood* **89**, 3554–3564. doi:10.1182/blood.V89.10.3554 (1997).
154. Bourguine, P. E. Human bone marrow organoids: emerging progress but persisting challenges. *Trends in Biotechnology* **44**, 11–21. doi:10.1016/j.tibtech.2025.05.028 (2026).
155. Stolz, A. *et al.* The rise of bone marrow organoids as next-generation models for blood formation and failure. *Experimental Hematology* **159**, 105428. doi:10.1016/j.exphem.2026.105428 (2026).
156. Richardson, S. E., Ghazanfari, R., Chhetri, J., Enver, T. & Böiers, C. In vitro differentiation of human pluripotent stem cells into the B lineage using OP9-MS5 co-culture. *STAR Protocols* **2**, 100420. doi:10.1016/j.xpro.2021.100420 (2021).
157. Bosma, G. C., Custer, R. P. & Bosma, M. J. A severe combined immunodeficiency mutation in the mouse. *Nature* **301**, 527–530. doi:10.1038/301527a0 (1983).
158. Shultz, L. D. *et al.* Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. *The Journal of Immunology* **154**, 180–191. doi:10.4049/jimmunol.154.1.180 (1995).
159. Ito, M. *et al.* NOD/SCID/ γ cnul mouse: an excellent recipient mouse model for engraftment of human cells. *Blood* **100**, 3175–3182. doi:10.1182/blood-2001-12-0207 (2002).
160. Shultz, L. D. *et al.* Human Lymphoid and Myeloid Cell Development in NOD/LtSz-scid IL2R γ cnul Mice Engrafted with Mobilized Human Hemopoietic Stem Cells, *The Journal of Immunology* **174**, 6477–6489. doi:10.4049/jimmunol.174.10.6477 (2005).
161. Willinger, T., Rongvaux, A., Strowig, T., Manz, M. G. & Flavell, R. A. Improving human hemato-lymphoid-system mice by cytokine knock-in gene replacement. *Trends in Immunology* **32**, 321–327. doi:10.1016/j.it.2011.04.005 (2011).
162. Dupard, S. J., Grigoryan, A., Farhat, S., Coutu, D. L. & Bourguine, P. E. Development of Humanized Ossicles: Bridging the Hematopoietic Gap. *Trends in Molecular Medicine* **26**, 552–569. doi:10.1016/j.molmed.2020.01.016 (2020).
163. Rodriguez-Fraticelli, A. E. & Parreno, V. Charting single-cell lineages with synthetic and natural barcodes. *Nature Reviews Genetics* **27**, 442–461. doi:10.1038/s41576-026-00943-5 (2026).
164. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63. doi:10.1038/nrg2484 (2009).

165. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382. doi:10.1038/nmeth.1315 (2009).
166. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* **30**, 777–782. doi:10.1038/nbt.2282 (2012).
167. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214. doi:10.1016/j.cell.2015.05.002 (2015).
168. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187–1201. doi:10.1016/j.cell.2015.04.044 (2015).
169. Fan, H. C., Fu, G. K. & Fodor, S. P. A. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367. doi:10.1126/science.1258367 (2015).
170. Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578. doi:10.1093/biostatistics/kxx053 (2018).
171. Jiang, R., Sun, T., Song, D. & Li, J. J. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biology* **23**, 31. doi:10.1186/s13059-022-02601-5 (2022).
172. 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. *Nature Methods* **10**, 1213–1218. doi:10.1038/nmeth.2688 (2013).
173. Minnoye, L. *et al.* Chromatin accessibility profiling methods. *Nature Reviews Methods Primers* **1**, 10. doi:10.1038/s43586-020-00008-9 (2021).
174. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914. doi:10.1126/science.aab1601 (2015).
175. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nature Biotechnology* **37**, 916–924. doi:10.1038/s41587-019-0147-6 (2019).
176. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nature Biotechnology* **37**, 925–936. doi:10.1038/s41587-019-0206-z (2019).
177. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385. doi:10.1126/science.aau0730 (2018).
178. Zhu, C. *et al.* An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nature Structural & Molecular Biology* **26**, 1063–1070. doi:10.1038/s41594-019-0323-x (2019).
179. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature Biotechnology* **37**, 1452–1457. doi:10.1038/s41587-019-0290-0 (2019).

180. Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nature Biotechnology* **35**, 936–939. doi:10.1038/nbt.3973 (2017).
181. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* **14**, 865–868. doi:10.1038/nmeth.4380 (2017).
182. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology* **24**, 695–713. doi:10.1038/s41580-023-00615-w (2023).
183. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. doi:10.1093/bioinformatics/bts635 (2013).
184. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137. doi:10.1186/gb-2008-9-9-r137 (2008).
185. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nature Methods* **18**, 1333–1341. doi:10.1038/s41592-021-01282-5 (2021).
186. Fang, R. *et al.* Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nature Communications* **12**, 1337. doi:10.1038/s41467-021-21583-9 (2021).
187. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology* **33**, 495–502. doi:10.1038/nbt.3192 (2015).
188. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology* **42**, 293–304. doi:10.1038/s41587-023-01767-y (2024).
189. Heumos, L. *et al.* Best practices for single-cell analysis across modalities. *Nature Reviews Genetics* **24**, 550–572. doi:10.1038/s41576-023-00586-w (2023).
190. Germain, P.-L., Lun, A., Meixide, C. G., Macnair, W. & Robinson, M. D. *Doublet identification in single-cell sequencing data using scDblFinder* 2022. doi:10.12688/f1000research.73600.2.
191. Thibodeau, A. *et al.* AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biology* **22**, 252. doi:10.1186/s13059-021-02469-x (2021).
192. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* **9**, giaa151. doi:10.1093/gigascience/giaa151 (2020).
193. Yang, S. *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biology* **21**, 57. doi:10.1186/s13059-020-1950-6 (2020).
194. Ahlmann-Eltze, C. & Huber, W. Comparison of transformations for single-cell RNA-seq data. *Nature Methods* **20**, 665–672. doi:10.1038/s41592-023-01814-1 (2023).
195. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* **20**, 296. doi:10.1186/s13059-019-1874-1 (2019).

196. Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biology* **23**, 27. doi:10.1186/s13059-021-02584-9 (2022).
197. Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing. *Cell Systems* **2**, 239–250. doi:10.1016/j.cels.2016.04.001 (2016).
198. Ringnér, M. What is principal component analysis? *Nature Biotechnology* **26**, 303–304. doi:10.1038/nbt0308-303 (2008).
199. Van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).
200. McInnes, L., Healy, J. & Melville, J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* 2020. doi:10.48550/arXiv.1802.03426.
201. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **37**, 38–44. doi:10.1038/nbt.4314 (2019).
202. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* **41**, 391–407. doi:10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9 (1990).
203. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289–1296. doi:10.1038/s41592-019-0619-0 (2019).
204. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods* **19**, 41–50. doi:10.1038/s41592-021-01336-8 (2022).
205. Chervov, A. & Zinovyev, A. *Computational challenges of cell cycle analysis using single cell transcriptomics* 2022. doi:10.48550/arXiv.2208.05229.
206. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology* **15**, MSB188746. doi:10.15252/msb.20188746 (2019).
207. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008. doi:10.1088/1742-5468/2008/10/P10008 (2008).
208. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* **9**, 5233. doi:10.1038/s41598-019-41695-z (2019).
209. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32**, 381–386. doi:10.1038/nbt.2859 (2014).
210. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nature Biotechnology* **37**, 547–554. doi:10.1038/s41587-019-0071-9 (2019).
211. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477. doi:10.1186/s12864-018-4772-0 (2018).

212. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology* **20**, 59. doi:10.1186/s13059-019-1663-x (2019).
213. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502. doi:10.1038/s41586-019-0969-x (2019).
214. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29. doi:10.1016/j.cell.2021.04.048 (2021).
215. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods* **14**, 975–978. doi:10.1038/nmeth.4401 (2017).
216. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* **36**, 421–427. doi:10.1038/nbt.4091 (2018).
217. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21. doi:10.1016/j.cell.2019.05.031 (2019).
218. Lotfollahi, M., Hao, Y., Theis, F. J. & Satija, R. The future of rapid and automated single-cell data analysis using reference mapping. *Cell* **187**, 2343–2358. doi:10.1016/j.cell.2024.03.009 (2024).
219. Regev, A. *et al.* The Human Cell Atlas. *eLife* **6** (ed Gingeras, T. R.) e27041. doi:10.7554/eLife.27041 (2017).
220. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372. doi:10.1038/s41586-018-0590-4 (2018).
221. Hou, W. & Ji, Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *Nature Methods* **21**, 1462–1465. doi:10.1038/s41592-024-02235-4 (2024).
222. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676. doi:10.1016/j.cell.2006.07.024 (2006).
223. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872. doi:10.1016/j.cell.2007.11.019 (2007).
224. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318**, 1917–1920. doi:10.1126/science.1151526 (2007).
225. Ivanovs, A. *et al.* Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337. doi:10.1242/dev.134866 (2017).
226. Castiglioni, C., Piussi, R. & Ditadi, A. Charting human developmental hematopoiesis in vitro to open new routes for regenerative medicine. *Blood* **147**, 1680–1692. doi:10.1182/blood.2024028063 (2026).
227. Lapillonne, H. *et al.* Red blood cell generation from human induced pluripotent stem cells: perspectives for transfusion medicine. *Haematologica* **95**, 1651–1659. doi:10.3324/haematol.2010.023556 (2010).

228. Choi, K.-D., Vodyanik, M. A. & Slukvin, I. I. Generation of mature human myelomonocytic cells through expansion and differentiation of pluripotent stem cell–derived lin[−]CD34⁺CD43⁺CD45⁺ progenitors. *The Journal of Clinical Investigation* **119**, 2818–2829. doi:10.1172/JCI38591 (2009).
229. Carpenter, L. *et al.* Human induced pluripotent stem cells are capable of B-cell lymphopoiesis. *Blood* **117**, 4008–4011. doi:10.1182/blood-2010-08-299941 (2011).
230. Suzuki, N. *et al.* Generation of Engraftable Hematopoietic Stem Cells From Induced Pluripotent Stem Cells by Way of Teratoma Formation. *Molecular Therapy* **21**, 1424–1431. doi:10.1038/mt.2013.71 (2013).
231. Amabile, G. *et al.* In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. *Blood* **121**, 1255–1264. doi:10.1182/blood-2012-06-434407 (2013).
232. Sugimura, R. *et al.* Haematopoietic stem and progenitor cells from human pluripotent stem cells. *Nature* **545**, 432–438. doi:10.1038/nature22370 (2017).
233. Piau, O. *et al.* Generation of transgene-free hematopoietic stem cells from human induced pluripotent stem cells. *Cell Stem Cell* **30**, 1610–1623.e7. doi:10.1016/j.stem.2023.11.002 (2023).
234. Ng, E. S. *et al.* Long-term engrafting multilineage hematopoietic cells differentiated from human induced pluripotent stem cells. *Nature Biotechnology* **43**, 1274–1287. doi:10.1038/s41587-024-02360-7 (2025).
235. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA* **333**, 71–74. doi:10.1001/jama.2024.21972 (2025).
236. Shabani, M. & Marelli, L. Re-identifiability of genomic data and the GDPR. *EMBO Reports* **20**, EMBR201948316. doi:10.15252/embr.201948316 (2019).
237. Lauwereyns, J. *et al.* Toward a common interpretation of the 3Rs principles in animal research. *Lab Animal* **53**, 347–350. doi:10.1038/s41684-024-01476-2 (2024).
238. Sommarin, M. N. E. *et al.* Single-Cell Multiomics Reveals Distinct Cell States at the Top of the Human Hematopoietic Hierarchy 2021. doi:10.1101/2021.04.01.437998.
239. Böiers, C. *et al.* A Human IPS Model Implicates Embryonic B-Myeloid Fate Restriction as Developmental Susceptibility to B Acute Lymphoblastic Leukemia-Associated ETV6-RUNX1. *Developmental Cell* **44**, 362–377.e7. doi:10.1016/j.devcel.2017.12.005 (2018).
240. Gu, Z. *et al.* PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nature Genetics* **51**, 296–307. doi:10.1038/s41588-018-0315-5 (2019).
241. Lord, K. A., Abdollahi, A., Hoffman-Liebermann, B. & Liebermann, D. A. Proto-Oncogenes of the fos/jun Family of Transcription Factors are Positive Regulators of Myeloid Differentiation. *Molecular and Cellular Biology* **13**, 841–851. doi:10.1128/mcb.13.2.841-851.1993 (1993).
242. Shivdasani, R. A. *et al.* Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* **81**, 695–704. doi:10.1016/0092-8674(95)90531-6 (1995).

243. Aksöz, M. *et al.* Hematopoietic stem cell heterogeneity and age-associated platelet bias are evolutionarily conserved. *Science Immunology* **9**, eadk3469. doi:10.1126/sciimmunol.adk3469 (2024).
244. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nature Methods* **20**, 1355–1367. doi:10.1038/s41592-023-01938-4 (2023).
245. Copley, M. R. *et al.* The Lin28b–let-7–Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nature Cell Biology* **15**, 916–925. doi:10.1038/ncb2783 (2013).
246. Ryseck, R. P. & Bravo, R. c-JUN, JUN B, and JUN D differ in their binding affinities to AP-1 and CRE consensus sequences: effect of FOS proteins. *Oncogene* **6**, 533–542 (1991).
247. Starnes, L. M. *et al.* NFI-A directs the fate of hematopoietic progenitors to the erythroid or granulocytic lineage and controls β -globin and G-CSF receptor expression. *Blood* **114**, 1753–1763. doi:10.1182/blood-2008-12-196196 (2009).
248. Somasundaram, R., Prasad, M. A. J., Ungerback, J. & Sigvardsson, M. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood* **126**, 144–152. doi:10.1182/blood-2014-12-575688 (2015).
249. Pliner, H. A. *et al.* Cicero Predicts *cis*-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Molecular Cell* **71**, 858–871.e8. doi:10.1016/j.molcel.2018.06.044 (2018).
250. Candelli, T. *et al.* Identification and characterization of relapse-initiating cells in MLL-rearranged infant ALL by single-cell transcriptomics. *Leukemia* **36**, 58–67. doi:10.1038/s41375-021-01341-y (2022).
251. Grigoryan, A. *et al.* Engineering human mini-bones for the standardized modeling of healthy hematopoiesis, leukemia, and solid tumor metastasis. *Science Translational Medicine* **14**, eabm6391. doi:10.1126/scitranslmed.abm6391 (2022).
252. Stutterheim, J., van der Sluis, I. M., Vrenken, K. S., Pieters, R. & Stam, R. W. KMT2A-rearranged acute lymphoblastic leukemia in infants: current progress and challenges. *Haematologica* **110**, 1951–1961. doi:10.3324/haematol.2024.285642 (2025).
253. Kristiansen, T. A. *et al.* Developmental cues license megakaryocyte priming in murine hematopoietic stem cells. *Blood Advances* **6**, 6228–6241. doi:10.1182/bloodadvances.2021006861 (2022).
254. Pang, W. W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proceedings of the National Academy of Sciences* **108**, 20012–20017. doi:10.1073/pnas.1116110108 (2011).
255. Ling, R. E., Cross, J. W. & Roy, A. Aberrant stem cell and developmental programs in pediatric leukemia. *Frontiers in Cell and Developmental Biology* **12**. doi:10.3389/fcell.2024.1372899 (2024).
256. Hartmann, M. *et al.* Molecular Plasticity Results in Oncofetal Reprogramming and Therapeutic Vulnerabilities in Juvenile Myelomonocytic Leukemia. *Blood Cancer Discovery* **7**, 306–330. doi:10.1158/2643-3230.BCD-25-0246 (2026).

257. Wang, K. *et al.* Ontogeny Dictates Oncogenic Potential, Lineage Hierarchy, and Therapy Response in Pediatric Leukemia. *Cancer Discovery* **16**, 541–570. doi:10.1158/2159-8290.CD-25-0556 (2026).
258. Rowe, R. G. *et al.* The developmental stage of the hematopoietic niche regulates lineage in MLL-rearranged leukemia. *Journal of Experimental Medicine* **216**, 527–538. doi:10.1084/jem.20181765 (2019).
259. Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* **27**, 1000–1008. doi:10.1038/leu.2012.356 (2013).
260. Mimitou, E. P. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology* **39**, 1246–1258. doi:10.1038/s41587-021-00927-2 (2021).
261. Ludwig, L. S. *et al.* Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325–1339.e22. doi:10.1016/j.cell.2019.01.022 (2019).
262. Scherer, M. *et al.* Clonal tracing with somatic epimutations reveals dynamics of blood ageing. *Nature* **643**, 478–487. doi:10.1038/s41586-025-09041-8 (2025).

