LUND UNIVERSITY

Resolving the Cellular Heterogeneity of Human Hematopoietic Stem and Progenitor Cells. The influence of cellular heterogeneity on ontogeny, ageing and leukemia.

Sommarin, Mikael

2021

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Sommarin, M. (2021). *Resolving the Cellular Heterogeneity of Human Hematopoietic Stem and Progenitor Cells. The influence of cellular heterogeneity on ontogeny, ageing and 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/

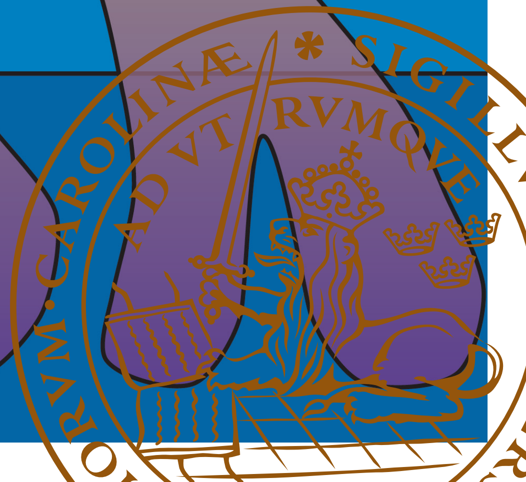
Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

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

Download date: 24. Jun. 2024

MIKAEL N.E. SOMMARIN
DEPARTMENT OF LABORATORY MEDICIN | FACULTY OF MEDICINE | LUND UNIVERSITY



Resolving the Cellular Heterogeneity of Human Hematopoietic Stem and Progenitor Cells

Resolving the Cellular Heterogeneity of Human Hematopoietic Stem and Progenitor Cells

The influence of cellular heterogeneity on ontogeny, ageing and leukemia

Mikael N.E. Sommarin

LUND UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at LUX Aula. Date 16th of December 2021 at 9.00.

Faculty opponent
David Kent, PhD

University of York

York, UK

<table>
<tr><td colspan="2">Organization
LUND UNIVERSITY

Author(s) Mikael N.E. Sommarin</td><td colspan="2">Document name
Doctorial Dissertation
Date of issue 2021-12-16
Sponsoring organization</td></tr>
<tr><td colspan="4">Title and subtitle: Resolving the Cellular Heterogeneity of Human Hematopoietic Stem and Progenitor Cells - The influence of cellular heterogeneity on ontogeny, ageing and leukemia</td></tr>
<tr><td colspan="4">Abstract
The hematopoietic system is a diverse and expansive system representing over 70% of all cells within the human body. This immense number of specialized cells all arise from one common precursor, the hematopoietic stem cell (HSC). The HSC is defined by its ability to self-renew and differentiate, imbuing the HSC with the ability to generate a myriad of specialized cell types while still replenishing its own pool of cells. This was long viewed to occur through a highly hierarchical process in which HSCs differentiated into increasingly lineage restricted subpopulation while increasing the number of cells in each step. However, with the introduction of single cell technologies extensive heterogeneity was observed within the hematopoietic stem and progenitor (HSPC) population. Thus, hematopoietic differentiation came to be viewed in a new light. The current paradigm of hematopoietic differentiation is considered to occur through continuous acquisition of lineage restriction and loss of self-renewal potential without any discrete intermediate steps. While the new paradigm of differentiation has deepened our knowledge of how the hematopoietic system is maintained, it also opens for further questions related to how cellular heterogeneity effects different aspects of hematopoiesis. Therefore, the work within this thesis attempts to resolve some of these questions, focusing on defining the epigenetic factors controlling HSC identity. The changes in heterogeneity occurring during ontogeny and ageing. Finally, the heterogeneity of primitive leukemic cells and how cellular heterogeneity contribute to relapse following therapy. Within the first paper we performed immunophenotypic, transcriptional and epigenetic interrogation of human HSPCs through integrative analysis of immunophenotypic screens, CITE-seq and scATAC-seq. This allowed us to define both the epigenetic drivers of HSC identity and the changes in heterogeneity caused by ageing. By performing this multi-omic analysis we could define a set of 650 enhancers linked with HSC identity. Additionally, this enhancer profile was shown to be enriched within CD35+ cells enabling in-vivo and in-vitro functional validation of stem cell capacity. Moreover, by performing CITE-seq on HSPCs from different age groups, CB, yBM and aBM we could observe extensive changes in heterogeneity with age. HSCs were shown to increase with age while multipotent and lymphoid progenitors decreased. Considering these observations in adult HSPCs we, in the second paper, investigated the heterogeneity of first trimester fetal liver HSPCs. While the general heterogeneity of FL HSPCs resembled their adult counterparts, the comparison of FL and BM cells again showed an increase of HSCs and a decrease of lymphoid progenitors with age. Additionally, by using differentially expressed genes between FL and adult BM we could derive a fetal gene signature. By applying this signature on Acute Lymphoid Leukemia (ALL) samples, it was shown that the signature could separate the samples based on age. This implies that certain subtypes of infant ALL retain a fetal derived gene signature, which could assist in defining the cell of origin in childhood ALL.Finally, we interrogated the heterogeneity of chronic myeloid leukemia (CML), a leukemic subtype known to be derived from a leukemic stem cell (LSC). These LSCs are thought to be therapy resistant and responsible for relapse following treatment. By combining immunophenotypic screens with multiplexed single-cell qPCR we could define a primitive therapy insensitive CML specific subpopulation. This subpopulation could be prospectively isolated using Lin-CD34+CD38-/lowCD117-CD45RA-CD26+, enabling for further molecular and functional characterization of the therapy insensitive CML cells. To conclude, this thesis emphasizes the power of single-cell molecular technologies to interrogate the heterogeneity of human HSPCs, leading to novel insights of human hematopoiesis in health and disease. It also stresses the need to interrogate cellular heterogeneity using several different features when defining cell states.</td></tr>
<tr><td colspan="4">Key words: Hematopoieis, Single-Cell, HSC, Leukemia, LSC, HSPC, sc-RNAseq, sc-ATAC-seq, Human</td></tr>
<tr><td colspan="4">Classification system and/or index terms (if any)</td></tr>
<tr><td colspan="3">Supplementary bibliographical information</td><td>Language English</td></tr>
<tr><td colspan="3">ISSN and key title 1652-8220</td><td>ISBN 978-91-8021-158-1</td></tr>
<tr><td rowspan="2">Recipient’s notes</td><td colspan="2">Number of pages 82</td><td>Price</td></tr>
<tr><td colspan="3">Security classification</td></tr>
</table>

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

Signature Date 2021-11-10

Resolving the Cellular Heterogeneity of Human Hematopoietic Stem and Progenitor Cells

The influence of cellular heterogeneity on ontogeny, ageing and leukemia

Mikael N.E. Sommarin

LUND
UNIVERSITY

Coverphoto by Mikael N.E. Sommarin

Copyright Mikael N.E. Sommarin

Paper 1 © by the Authors (Manuscript unpublished)

Paper 2 © by the Authors (Manuscript unpublished)

Paper 3 © Publisher The American Society of Hematology

Paper 4 © Publisher Journal of Visualized Experiments

Faculty of Medicine Doctoral Dissertation Series 2021:151
Department Laboratory Medicine

ISBN 978-91-8021-158-1
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2021

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

“No man ever steps in the same river twice,
for it's not the same river and he's not the same man.”
-Heraclitus

Table of Contents

Original articles and manuscripts **11**

Abstract **13**

List of abbreviations **15**

Introduction **17**

Hematopoiesis 17
The hematopoietic stem cells 17
Sources of HSCs 18
Defining the hematopoietic stem cell 19
Prospective isolation of HSCs and hematopoietic progenitors 20
The hematopoietic hierarchy; an ever-changing paradigm 22
Development and ageing of the hematopoietic system 24
Leukemia 27

Cell identity, cell states and cell types 30
Epigenomics, Transcriptomics and proteomics 32

Methods commonly used to study hematopoiesis 35
Flow cytometry and fluorescent activated cell sorting (FACS) 35
In-vivo analysis of human HSPC 35
In-vitro assays to evaluate cell potential 37
Single cell molecular technologies 38

Ethical considerations 45

Aims of the thesis **47**

Summary of articles included in the thesis **49**

Paper I 49

Paper II 52

Paper III 54

Paper IV 56

General discussion and future perspectives **57**

Populärvetenskaplig sammanfattning **61**

Acknowledgement **65**

References 69

Original articles and manuscripts

Paper I

Single-Cell Multiomics Reveals Distinct Cell States at the Top of the Human Hematopoietic Hierarchy

Mikael N.E. Sommarin, Parashar Dhapola, Fatemeh Safi, Rebecca Warfvinge, Linda Geironson Ulfsson, Eva Erlandsson, Anna Konturek-Ciesla, Ram Krishna Thakur, Charlotta Böiers, David Bryder and Göran Karlsson. Manuscript in preparation.

Paper II

A Combined Immunophenotypic and Transcriptional Map of First Trimester Human Fetal Liver Hematopoiesis

Mikael N.E. Sommarin, Rasmus Olofzon, Sara Palo, Parashar Dhapola, Göran Karlsson and Charlotta Böiers. Manuscript in preparation.

Paper III

Single-cell molecular analysis defines therapy response and immunophenotype of stem cell subpopulations in CML

Rebecca Warfvinge*, Linda Geironson*, **Mikael N. E. Sommarin***, Stefan Lang, Christine Karlsson, Teona Roschupkina, Leif Stenke, Jesper Stentoft, Ulla Olsson-Strömberg, Henrik Hjorth-Hansen, Satu Mustjoki, Shamit Soneji, Johan Richter, and Göran Karlsson, *Shared First Name, Blood, 25/01/2017, Vol 129, Nr17, 2384-2394, doi:10.1182/blood-2016-07-728873.

Paper IV

A Combinatorial Single-cell Approach to Characterize the Molecular and Immunophenotypic Heterogeneity of Human Stem and Progenitor Populations.

Mikael N.E. Sommarin, Rebecca Warfvinge, Fatemeh Safi, Göran Karlsson, J. Vis. Exp, 10/25/2018, e57831, doi:10.3791/57831

Abstract

The hematopoietic system is a diverse and expansive system representing over 70% of all cells within the human body. This immense number of specialized cells all arise from one common precursor, the hematopoietic stem cell (HSC). The HSC is defined by its ability to self-renew and differentiate, imbuing the HSC with the ability to generate a myriad of specialized cell types while still replenishing its own pool of cells. This was long viewed to occur through a highly hierarchical process in which HSCs differentiated into increasingly lineage restricted subpopulation while increasing the number of cells in each step. However, with the introduction of single cell technologies extensive heterogeneity was observed within the hematopoietic stem and progenitor (HSPC) population. Thus, hematopoietic differentiation came to be viewed in a new light. The current paradigm of hematopoietic differentiation is considered to occur through continuous acquisition of lineage restriction and loss of self-renewal potential without any discrete intermediate steps.

While the new paradigm of differentiation has deepened our knowledge of how the hematopoietic system is maintained, it also opens for further questions related to how cellular heterogeneity effects different aspects of hematopoiesis. Therefore, the work within this thesis attempts to resolve some of these questions, focusing on defining the epigenetic factors controlling HSC identity. The changes in heterogeneity occurring during ontogeny and ageing. Finally, the heterogeneity of primitive leukemic cells and how cellular heterogeneity contribute to relapse following therapy.

Within the first paper we performed immunophenotypic, transcriptional and epigenetic interrogation of human HSPCs through integrative analysis of immunophenotypic screens, CITE-seq and scATAC-seq. This allowed us to define both the epigenetic drivers of HSC identity and the changes in heterogeneity caused by ageing. By performing this multi-omic analysis we could define a set of 650 enhancers linked with HSC identity. Additionally, this enhancer profile was shown to be enriched within CD35+ cells enabling in-vivo and in-vitro functional validation of stem cell capacity. Moreover, by performing CITE-seq on HSPCs from different age groups, CB, yBM and aBM we could observe extensive changes in heterogeneity with age. HSCs were shown to increase with age while multipotent and lymphoid progenitors decreased.

Considering these observations in adult HSPCs we, in the second paper, investigated the heterogeneity of first trimester fetal liver HSPCs. While the general heterogeneity of FL HSPCs resembled their adult counterparts, the comparison of FL and BM cells again showed an increase of HSCs and a decrease of lymphoid progenitors with age. Additionally, by using differentially expressed genes between FL and adult BM we could derive a fetal gene signature. By applying this signature on Acute Lymphoid Leukemia (ALL) samples, it was shown that the signature could separate the samples based on age. This implies that certain subtypes of infant ALL retain a fetal derived gene signature, which could assist in defining the cell of origin in childhood ALL.

Finally, we interrogated the heterogeneity of chronic myeloid leukemia (CML), a leukemic subtype known to be derived from a leukemic stem cell (LSC). These LSCs are thought to be therapy resistant and responsible for relapse following treatment. By combining immunophenotypic screens with multiplexed single-cell qPCR we could define a primitive therapy insensitive CML specific subpopulation. This subpopulation could be prospectively isolated using Lin-CD34+CD38-/lowCD117-CD45RA-CD26+, enabling for further molecular and functional characterization of the therapy insensitive CML cells.

To conclude, this thesis emphasizes the power of single-cell molecular technologies to interrogate the heterogeneity of human HSPCs, leading to novel insights of human hematopoiesis in health and disease. It also stresses the need to interrogate cellular heterogeneity using several different features when defining cell states.

List of abbreviations

ALL	Acute Lymphoid Leukemia
AML	Acute Myeloid Leukemia
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
BM	Bone Marrow
CB	Cord Blood
CITE-seq	Cellular Indexing of Transcriptomes and Epitopes by Sequencing
CLL	Chronic Lymphoid Leukemia
CLOUD	Continuum of LOw primed Un Differentiated
CLP	Common Lymphoid Progenitors
CML	Chronic Myeloid Leukemia
CS	Carnegie Stage
dHSC	definitive HSC
EMP	erythro-myeloid progenitors
FACS	Fluorescence-Activated Cell sorting
FL	Fetal Liver
GMP	Granulocyte and Macrophage Progenitors
HSC	Hematopoietic Stem Cell
HSPC	Hematopoietic Stem and Progenitor Cell
HVGs	Highly Variable Genes
iPS	induced pluripotent stem
Lin-	Lineage negative
LMPP	Lympho-Myeloid Progenitor Populations
LSC	Leukemic stem cell
MEP	Megakaryocyte and Erythroid progenitors

MLL	Mixed Lineage Leukemia
MLP	Multi-Lymphoid Progenitors
MNC	Mononuclear Cells
MPP	Multipotent Progenitor Populations
NSG	NOD/LtSz-scid/Il2rg-/-
PCA	Principal Component Analysis
PCW	Post Conception Week
PTM	post-translationally modified
sc	single-cell
sc-qPCR	single-cell quantitative real-time Polymerase Chain Reaction
scRNA-seq	single-cell RNA sequencing
TF	transcription factor
TFBS	Transcription Factor Binding Sites
t-SNE	t-distributed stochastic neighbor embedding
TSS	transcription start sites
UMAP	Uniform Approximation and Projection
UMIs	Unique Molecular Identifiers

Introduction

Hematopoiesis

The hematopoietic system is a complex and diverse system interacting with all part of the body, producing more than 300 million cells each day. The cells produced have varying functions, from transporting oxygen to defending the body from invading pathogens. These varying functions are derived from several different specialized cell types and more than ten distinct cell types have been described, among them are lymphoid (B- and T-lymphocytes, dendritic cells (DC) and natural killer cells (NK)), myeloid (monocytes/macrophage and granulocytes), megakaryocyte/platelets, mast cells, and red blood cells (erythrocytes) [1]. All of which are derived from the same cell of origin, the hematopoietic stem cell (HSC). The high variation of effector cells is created by differentiation through a series of progenitor populations with increasing specialization. This have led to the hematopoietic system being described as a hierarchical structure with discrete steps and with the hematopoietic stem cell at its apex [2].

The hematopoietic stem cells

The stem cell, one of the most versatile cell-types in biology, with an almost limitless potential, due to its ability to self-renew and differentiate, has been extensively studied since its existence was suggested in the nineteenth century. The first use of the word stem cell or "stammzelle" was by the German scientist Ernst Haeckel, originally to describe the unicellular organisms from which life evolved [3]. However, he later expanded this concept to include the fertilized egg *i.e.*, the embryonic stem cell. At the turn of the century, one of the big questions in the field of hematopoiesis were if myeloid and lymphoid blood cells originated from the same progenitor cell or if they came from two independent progenitors [3]. Proponents of the unitarian model of hematopoiesis were inspired by the studies of embryology and started to refer to the common hematopoietic progenitor cell as a "stem cell". This definition has led to a dichotomy within the stem cell field where research is divided between the study of the embryonic stem cell, which can make all cells within the body (totipotent), and the adult or tissue resident stem cell, which can "only" make the cells of its own lineage (pluripotent) and is responsible for homeostasis.

The existence of the hematopoietic stem cell and the validation of the unitarian model was not proven until approximately 60 years later through the seminal experiments by Till and McCulloch [4-7]. Later, other stem cell populations within the adult organism have been shown to exist as well, including but not limited to, epidermal, hair follicle, mesenchymal and brain [8].

As previously stated, the stem cells have the unique ability to self-renew and differentiate [1, 2]. Self-renewal refers to a property where the cell can make a copy of itself without losing "stemness", thus expanding the stem cell pool indefinitely. Differentiation, on the other hand, refers to the property where the stem cell can produce daughter cells which become increasingly committed to a specific lineage and subsequently an effector cell which exerts a specific function. This means that a stem cell can rebuild an entire organism or a specific cell tissue for example, blood, liver, or hair., making the stem cell a highly regenerative cell type.

Due to its ability to regenerate tissues, the stem cell has huge potential in regenerative medicine, this led researchers to imagine using stem cells to rebuilding lost or damaged tissues and transplanting them into patients. It sounds almost like science fiction; however, stem cell transplants have been routinely used in the clinic for the last 40 years. The stem cell therapy referred to here is the bone marrow transplantation that was first performed in 1957 and is currently used in a myriad of deceases including, acute myeloid leukemia (AML), chronic myeloid leukemia (CML), aplastic anemia and thalassemia [9-11]. Bone marrow (BM) transplantations are often used as a last resort when all other treatments fail, since it is an invasive procedure and requires a donor, and matching donors are scarce.

Sources of HSCs

The main site of hematopoiesis in the adult is the BM, where the HSCs reside within a specific microenvironment called a niche [2, 9]. This promotes and regulates function and maintenance. The niche is composed of several specific cell types, including mesenchymal, hematopoietic, endothelial and vasculature cell types. These cell types collaborate to provide the Hematopoietic Stem and Progenitor Cells (HSPCs) with the factors needed to maintain their properties and to differentiate. These cell types provide the HSPCs with a myriad of factors, including cytokines to promote stem cell maintenance and differentiation. Importantly, the cytokines Stem Cell Factor (SCF) and thrombopoietin (TPO) both maintain stem cell properties [9].

Although the BM is the main site of HSC production and maintenance, HSCs can be found in other sites as well, including circulating in the blood vessels. During early life HSC can be found in the fetal liver (FL) and within the umbilical cord blood (CB) [12, 13]. HSCs migrate through peripheral blood vessel to other BM niches to colonize them. This can be induced by the cytokine Granulocyte Colony

stimulating Factor (G-CSF) [14] and is currently a method is commonly used to collect HSCs for transplants into patients following myeloablation.

The two other sources of HSCs, FL and CB, are developmental specific sites where the HSCs migrate to the FL during development and undergo extensive expansion before migrating into the BM before birth [15, 16]. The CB derived HSCs are the most accessible source of HSCs, due to the umbilical cord blood being a by-product of birth. The ease of accessibility and non-invasive nature of collection has made it the premiere source of HSCs in research and much of the knowledge of human HSCs and their progenitor populations comes from studies conducted on CB HSCs. However, due to the difference in niche and developmental stage between HSPCs derived from CB and BM, it is unclear if the results from studies on CB HSCs can be generalized to HSCs residing in the BM.

Defining the hematopoietic stem cell

All stem cells are defined by their ability to self-renew and differentiate, and thus the gold standard for defining the HSC is through transplantations [2]. In brief, a population of cells potentially containing HSCs are transplanted into a lethally irradiated host, and after some time the potential of the donor cells to recreate the hematopoietic system is analyzed. By measuring the duration of lineage output from the transplanted cells one can ascribe a function to the population, *i.e.*, if they are HSCs, restricted progenitors, multi-potent progenitors or more mature cells with low multi-lineage reconstitution potential.

These experiments, initially performed by Till and MacCulloch in mice, assessed lineage output in the spleens of the transplanted animals where they identified colonies of hematopoietic cells [4, 6, 7]. Even through that these seminal experiments were performed over 60 years ago, the way we define an HSC has not changed much since but merely been refined. These refinements have led to a standardized way of assessing HSC potential in mice, were a bona-fide mouse HSC should be able to produce all hematopoietic lineages, as measured by flow cytometry in blood and BM, for up to 16 weeks. This is followed by a re-transplantation into a secondary lethally irradiated recipient (referred to as a secondary transplant) and 16-week follow-up, whereafter HSCs and hematopoietic repopulation should be detected. This assay has made transplantations the gold standard of assaying HSCs and their function [1, 2].

However, the investigation of human HSCs was long hampered by the inability to transplant human HSCs into irradiated recipients since the mouse immune system would reject the cells. Therefore, mouse models with reduced immune response have been produced to allow for human xenografts, most notably the NOD/LtSz-scid/Il2rg-/- (NSG) mouse [17]. This model can sustain engraftment of human HSCs for up to several months. However, these mice have caveats since they do not

allow for the production of all human lineages and the cell lineage proportions are different compared to the human steady state hematopoiesis [18]. This have led to the development of new transgenic mouse models produces human cytokines to give balanced human engraftment, and even models with implanted human bone marrow niches [19-21].

By transplanting limiting dilutions of cell numbers, referred to as a limiting dilution assay (LDA), the frequency of HSCs has been assessed. This showed that within the BM, HSCs are quite scarce, with a frequency of only 1 cell in 1 000 000 [12].

Even though the transplantation assay has its flaws, it has still given us the ability to measure the functional potential of HSCs. This has led to hematopoiesis being the most extensively studied stem cell system and have laid the conceptional foundation for how we view stem cells in general.

Prospective isolation of HSCs and hematopoietic progenitors

To circumvent the issues of low HSC frequencies in BM researchers have utilized Fluorescence-Activated Cell sorting (FACS) to enrich HSC and other hematopoietic progenitor populations. The initial enrichment strategies relied on removing cells with known functions i.e., highly differentiated cell types like B-cells, monocytes, platelets etc., using cell surface proteins. This set of cell markers are referred to as lineage markers, and the resulting population as lineage negative (Lin-) [22].

The increasing ability to produce monoclonal antibodies together with the development of multi-color FACS allowed for inclusion of more and more surface markers which in turned has led to a refinement in HSPC isolation [1]. These technological advances have facilitated extensive characterization of the mouse hematopoietic system including intermediate progenitor populations, as well as an impressive ability for HSC purification, where murine HSCs can be prospectively isolated to a purity of approximately one in two to three cells [23-25]. However, characterization of the human HSPC compartment have been lagging behind its murine counterpart, where the most advanced isolation strategies only achieve a purity of one transplantable HSC for every ten cells [26]. Even so, this represents an impressive 100.000-fold enrichment compared to unfractionated BM. It also implies that 90% of the cells within the human immunophenotypic HSC populations are not HSCs, a fact which confounds molecular and functional characterization of these cells.

The same pattern as with the HSCs can also be seen within the progenitor populations, where murine progenitors have been classified in more than nine different sub-populations, whereas the human counterpart has only been divided into approximately seven populations (Figure 1A-B) [2]. As with the HSCs the murine HSPCs has served as a blueprint when defining the different progenitor populations in humans. Thus, many of the progenitors share name conventions and

immunophenotypic markers between mice and humans. Roughly, the progenitors can be divided as follows: Common Myeloid Progenitors (CMPs), Granulocyte and Macrophage Progenitors (GMPs), Megakaryocyte and Erythroid progenitors (MEPs), Common Lymphoid Progenitors (CLPs), Lympho-Myeloid Progenitor Populations/Multi-Lymphoid Progenitors (LMPPs/MLPs) and Multipotent Progenitor Populations (MPPs) [2, 26-28].

In the human system all HSPCs express the surface marker CD34, representing approximately 1% of all mononuclear cells (MNCs) [29]. Despite this, it has been reported that some HSCs lack CD34 expression, and these CD34- HSCs has even been suggested to represent a more primitive population compared to the CD34+ HSCs [30-32]. This is especially interesting since murine HSCs lack the expression of CD34, and this population could represent a bridge between human and mouse. However, the human CD34- HSCs are very infrequent and has only been found CB, thus it has been largely ignored by the research community at large.

Next, the HSPCs can be subdivided on their expression of CD38, where the lineage affiliated progenitors; GMP, CMP, MEP and CLP are defined by high expression of CD38, while the primitive progenitors; HSC, MPP and LMPP have negative/low expression [33]. CLPs are then separated from GMPs, CMPs, and MEPs by its expression of CD45RA and CD10 [27, 34, 35]. Furthermore, GMPs, CMPs, and MEPs are subdivided based on the expression of CD45RA and CD123 (IL2R) or 135 (FLT3) [27, 28], with GMPs being positive for both CD45RA and CD123 or CD135. While MEPs are negative for both (CD45RA-CD123-/CD135-), and the rest is defined as CMPs (CD45RA-CD123+/CD135+).

The CD38 negative/low fraction of HSPCs can be further fractionated into HSCs, MPPs and LMPPs based on the expression of CD90 and CD45RA [36, 37]. Where the LMPPs are negative for CD90 and positive for CD45RA (CD90-CD45RA+), the MPPs are negative/low for both CD90 as well as CD45RA. This leaves the HSCs, which are defined by positive expression of CD90 and negative for CD45RA (CD90+CD45RA-). The HSC population can be further fractionated using CD49F, and the resulting CD49F+ HSC population archives the highest purity of human HSCs to date where 1 in 10 cells being a transplantable HSC with multi-lineage potential [26].

Extensive immunophenotypic characterization of both human and mouse HSPCs have allowed for a relatively detailed description of the hematopoietic hierarchy. However, immunophenotypic characterization of hematopoietic cells is an active field of research and novel subpopulations are continuously being described.

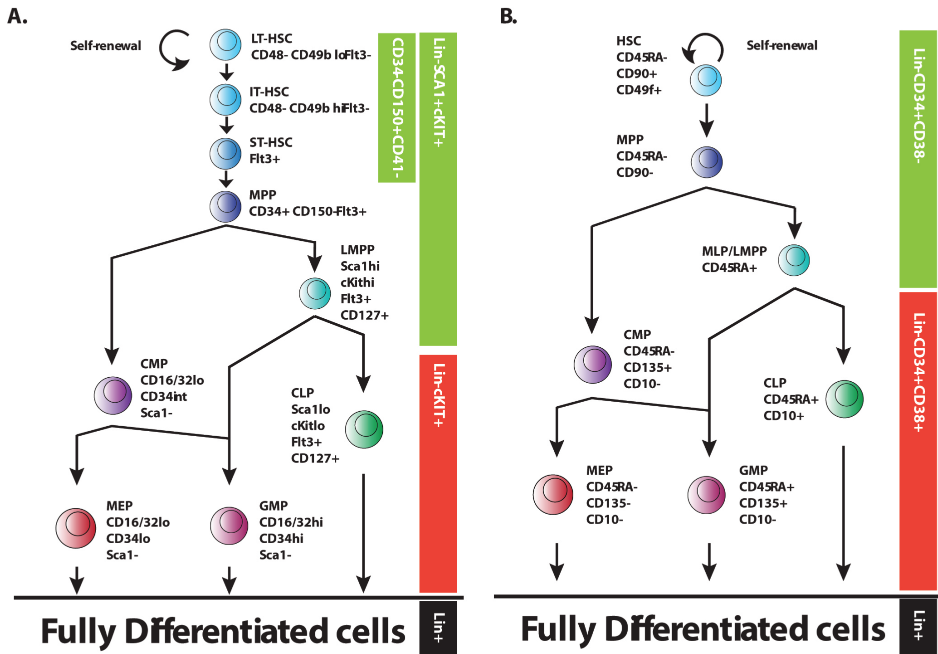


Figure 1. The hematopoietic hierarchy of human and mouse. A) The murine hematopoietic hierarchy including the cell surface makers to purify stem and progenitor populations. B) The human hematopoietic hierarchy including the cell surface makers to purify stem and progenitor populations. Adapted from [2].

The hematopoietic hierarchy; an ever-changing paradigm

The hematopoietic system has become the conceptual paradigm of which all stem cell systems are compared to, much because of our ability to prospectively isolate and functionally test individual stem and progenitor populations. Through this iterative process a hematopoietic hierarchy has been postulated (Figure 2A) [2, 38].

Early attempts utilized microscopy to describe the relationships between morphologically different blood cells, mainly separating myeloid and lymphoid cells. However, as technologies have evolved, researchers turned to flow cytometry coupled to functional in-vivo and in-vitro experiments [1, 2]. When modern molecular methods arrived, the genetic and transcriptional drivers of differentiation could begin to be delineated [38, 39]. These factors contributed to hematopoietic differentiation being described as a hierarchical stepwise process were populations of cells sequentially gains lineage identity while losing multipotency, until they finally become fully differentiated effector cells (Figure 2A).

This paradigm remained relatively unchanged for several decades, with only minor additions of intermediate populations and differentiation paths. However, it was noted that these populations were not functionally pure, something which was often

attributed to insufficiently characterized populations [40, 41]. This was drastically changed with the advent of the single-cell molecular profiling, where the molecular characteristics of each cell can be measured. Early studies using single-cell quantitative real-time polymerase chain reaction (sc-qPCR) observed molecular heterogeneity in immunophenotypically defined populations, explaining the functional heterogeneity previously observed [42, 43].

However, sc-qPCR is limited, both in number of cells as well in transcripts that it can measure and can therefore not satisfyingly interrogate the entire heterogeneity of the HSPC population. Therefore, it was not until the advent of high through-put single-cell RNA sequencing (scRNA-seq) that the entire cellular heterogeneity of HSPCs could be investigated. In the now seminal study by Velten *et.al.* investigators showed that the human HSPCs differentiate along a continuum without discrete steps [44]. This discovery formed the basis for the current paradigm of the hematopoietic differentiation hierarchy, referred to as the Continuum of LOw primed Un Differentiated (CLOUD) HSPCs (Figure 2B). This study together with several other studies displayed the same results in the murine system that has further been confirmed with alternative methods such as single-cell Assay for Transpose-Accessible Chromatin using sequencing (scATAC-seq) [45, 46].

The CLOUD-HSPC model suggests, that the changes observed during differentiation, occur through minor transcriptional/epigenetic alterations, where the differentiated cell being highly similar to the cells which preceded it, more akin to marbles rolling down a hill. Interestingly, this model of differentiation was suggested as early as 1957, by Conrad Waddington, and is referred to as the Waddington's landscape [47]. In this metaphor the cells are in a constant flux and "rolls" down from the top of the hill. Lineage commitment is represented by the cells getting stuck in different valleys until they finally reach the end of the hill and become fully differentiated lineage committed effector cells (Figure 2C).

The Waddington landscape/CLOUD-HSPC model represents the current paradigm of lineage commitment in the hematopoietic hierarchy and is a fascinating example on how technological development can prove an almost 60-year-old theory. The history of the hematopoietic hierarchy also shows that the hematopoietic hierarchy is an ever-changing paradigm and will most likely change again as novel methods are developed.

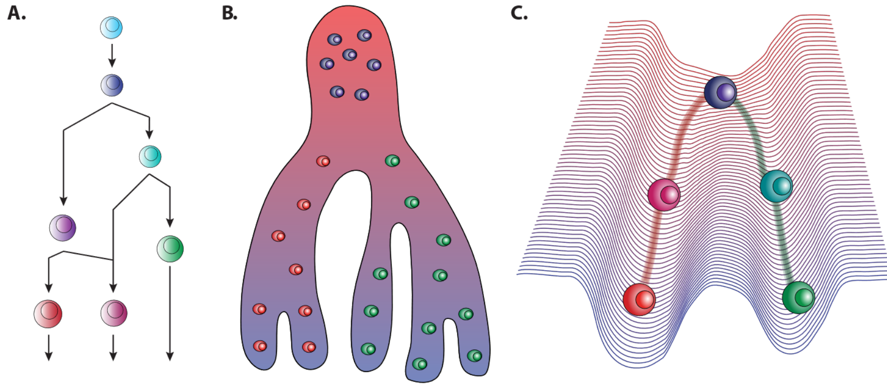


Figure 2. Conceptual hierarchies of hematopoietic differentiation. A) Step wise lineage commitment though defined cell states. B) Continuum model of differentiation (CLOUD-HSPC model). C) Waddington's landscape.

Development and ageing of the hematopoietic system

There are only a few certain facts in life, we are conceived, and we die, between that we develop and age. This is also true for the hematopoietic system. Interestingly, the hematopoietic system develops in three consecutive but overlapping waves, where the definitive HSCs (dHSCs) responsible for life long homeostasis is not developed until the third and final wave (Figure 3) [13].

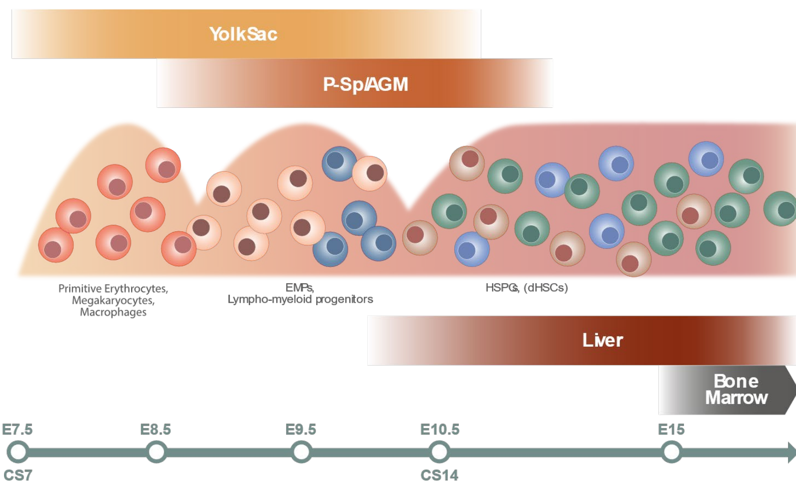


Figure 3. Timeline of development of the hematopoietic system. Embryonic day (E), Carnegie stage (CS). Reprinted from [48], original draft by Veronika Žemaitė.

Before the dHSCs are formed, waves of erythroid, myeloid and lymphoid cells are formed to cover the embryo's basic needs. The first wave consists of erythrocytes, macrophages and megakaryocytes, while during the second erythro-myeloid progenitors (EMPs) are formed, followed by progenitors with lymphoid potential [13]. In mice, the first wave initiates in the yolk sac at E7 (embryonic day) while the second wave forms at E8 with EMPs (also in the yolk sac). Lymphoid progenitors form at E9.5 in both the yolk sac and the embryo proper [49-51]. The dHSC together with hematopoietic progenitors are formed at E10.5 in the aorta-gonad-mesonephros (AGM) region [52]. dHSCs and HSC-independent progenitors arise through the endothelial to hematopoietic transition (EHT), where the hematopoietic cells bud off from endothelial cells lining the dorsal aorta [53]. EHT has also been shown to be responsible for the development of EMPs and it has been suggested that the EHT is also responsible for the development of the embryonically derived myeloid and lymphoid cells.

After the dHSCs are formed they migrate to the fetal liver (FL) at around E12 where they massively expand before finally migrating to the BM at E17.5 [15, 16]. The BM serves as the major BM niche for the rest of life and BM HSCs form the basis of hematopoietic hierarchy responsible for lifelong homeostasis. Interestingly, also some of the early embryonically derived cell populations persists into adulthood [54, 55].

Even though the development of the hematopoietic system has been extensively studied, most of these studies have been performed in mouse. How well the murine system captures human hematopoietic development still remains to be resolved. The lack of studies in the humans is the result of both practical and ethical restraints. To circumvent this problem and investigate the earliest steps of human development, induced pluripotent stem (iPS) cells have been used to model the EHT [56, 57]. Additionally, with the advent of single-cell technologies human developmental hematopoiesis have been studied at an un-presented resolution, also re-capturing the general themes seen in mouse [58-60].

These novel methods together with FACS, in-vivo and in-vitro analysis have been able to define certain features specific to the developing hematopoietic system. Such as that the fetal HSCs are more cycling compared to their adult counterparts, a result of the need for HSCs to form the entire hematopoietic system [61]. Additionally, the fetal derived progenitor populations have been shown to be more multi-/bi-potential, meaning that a single cell can form several different lineages [40, 50], suggesting that these fetal derived progenitor cells are less restricted to a specific lineage. Additionally, the FL cells are also more biased towards forming lymphoid cell types [62], possibly due to the need for the development of immune system in early life.

Three weeks after the HSCs have migrated into the BM, a switch to adult hematopoiesis has been identified in the murine system [15, 16] and recent findings

have shown that this fetal to adult transition occurs through a gradual change in gene expression, independent of niche [63]. Once the hematopoietic system has been set up in the BM, it remains relatively unchanged, contributing to the homeostasis of the hematopoietic system for the remainder of life.

However, as the individual ages, so does the hematopoietic system, if these changes are driven by cell intrinsic of extrinsic changes is a highly contested area of research [9, 64, 65]. No matter what the driving factor is, the resulting phenotype is well characterized. As the HSPCs age, the HSCs lose their reconstitution potential, and the immunophenotypic HSCs increase in proportion relative to other hematopoietic cells in the BM [66-69]. This dichotomy between HSC cell numbers and ability of the HSCs to reconstitute a hematopoietic hierarchy have been explained by changes in cell cycle status [70, 71].

Furthermore, the ageing changes HSPCs in their lineage potential where the young HSPCs have a balanced output of myeloid, erythroid and lymphoid cell types and the aged cells are biased towards a myeloid out-put [66-68, 72]. Additionally, the hematopoietic progenitor populations become increasingly lineage restricted with age [40]. Where the young and fetal HSPCs display bi-/multipotent lineage out-put, their aged counterparts produce mostly unilineage out-put. Whether this is caused by cell intrinsic, extrinsic i.e., niche or the underlying heterogeneity, still needs to be resolved.

Some proponents suggests that the changes observed in ageing is driven by changes in the niche. For example, it has been shown that subjecting aged hematopoietic cells to a young microenvironment rejuvenates the cells [73, 74]. However the rejuvenation by external stimuli have recently been suggested to be more difficult than previously described [75]. Some studies suggest that the hematopoietic cells themselves drive ageing though intrinsic factors, where epigenetic factors, DNA repair, metabolic dysregulation, mitochondrial stress and altered cell polarity have been described as important contributors [64, 76-81]. Interestingly, aged human HSCs displayed substantial epigenetic changes linked to leukemic transformation, identifying a possible epigenetic link to leukemia initiation [77]. Additionally, reprogramming of the aged hematopoietic cells into iPS, which removes a cell's epigenetic signature, followed by differentiation into HSCs, rejuvenated the hematopoietic compartment to levels seen in young mice [82]. Due to the interconnectivity of the hematopoietic cells and the niche, where both the niche and HSPCs can modulate affect each other's function. Therefore, the most likely the answer lies between the two extremes, where both cell intrinsic and extrinsic factors contribute to produce the aged phenotype.

Leukemia

As with all things in life sometimes things go astray, and when things go astray within the hematopoietic system it could prove life threatening. Cancer is one of the leading causes of death in the developed world, and if cancer occurs within the hematopoietic system it is referred to as a leukemia or lymphoma [83]. Leukemia is marked by the accumulation of undifferentiated, defective progenitor cells, called blast cells, this accumulation of blast cells displaces normal cells leading to defective hematopoiesis. Leukemia is broadly classified according to the speed of progression into chronic and acute leukemia subtypes, where acute leukemia progresses faster, compared to chronic leukemia. However, if the chronic leukemia is left untreated, they can progress into an accelerated phase and finally into blast crisis, mirroring their acute counterparts [84].

Leukemia is also separated by the lineage where the accumulation of defective progenitors occurs, being separated into myeloid and lymphoid (either B or T) leukemia. Interestingly, myeloid leukemia is more prevalent in older patients (>50) while lymphoid leukemia occurs more frequently in infants and children [85, 86]. This reflects well the differences seen in the fetal and aged HSPCs, where fetal cells are lymphoid biased while the aged cells are myeloid biased [62, 67, 72]. This also extends to the childhood and infant leukemia where the first hit mutation has been detected in neonatal blood spots prior to overt leukemia onset, suggesting an in-utero origin of leukemia initiation [85, 87].

These definitions subdivide the leukemia into four general classes: Chronic Myeloid Leukemia (CML), Acute Myeloid Leukemia (AML), Chronic Lymphoid Leukemia (CLL) and Acute Lymphoid Leukemia (ALL). Outside of these strict definitions there is also the Mixed Lineage Leukemia (MLL) [88], where the leukemia first presents as either an ALL or AML, but upon treatment or at a more advanced stage relapses can switch to a leukemia of the opposite lineage [88].

The tendency of leukemias to relapse following treatment argues for that a small subset of therapy resistant cells remains following treatment and can rebuild the entire leukemic system. This observation has led researchers to propose the Leukemic Stem Cell (LSC) model. The LSC model proposes that all leukemic cells originate from one primitive leukemic founder cell the LSC. LSCs can like the HSC form all leukemic cells through differentiation as well as self-renew to maintain the LSC pool. If this LSCs are not eradicated by the treatment the leukemia will remerge and cause a relapse (Figure 4A-B). If the LSC is formed through a mutation in an HSC or by initiation in an early progenitor which then acquire stem cell like properties have not been fully resolved [84]. However, it has been shown that within AML and CML the leukemia initiating cells reside within the Lin-CD38-/low population, a population containing HSCs, MPPs and LMPPs, and not so frequently in the CD38+ progenitor populations in both CML and AML [89, 90]. Furthermore, within CMLs, there have been reports of non-leukemic cells harboring the leukemia

initiating mutation suggesting an HSC-derived cell of origin [91, 92]. However, in an inducible mouse model for AML, leukemia could only be initiated within a specific differentiation window, between the granulocyte/monocyte/lymphoid progenitors (GMLP) and GMP while leukemic transformation of the HSC population instead led to loss of function of the HSCs [93]. This raises the question if different leukemia could have different cells of origin.

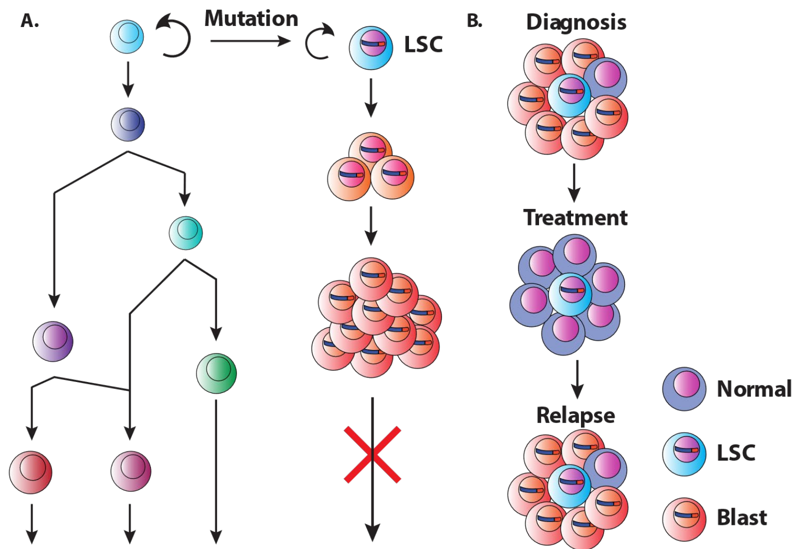


Figure 4. Development and relapse of leukemia according to the LSC model. A) Initiation of leukemia within primitive a hematopoietic population. B) Leukemic heterogeneity during diagnosis, treatment and relapse.

CML

CML is a leukemia caused by the fusion protein BCR-ABL1 resulting from the translocation t9;22, also referred to as the Philadelphia chromosome (Ph) [94, 95]. The BCR-ABL1 fusion protein is a constitutively expressed tyrosine kinase which drives the leukemic transformation resulting in a myeloproliferative disease. Incidences of CML is relatively rare, with one to two individuals of 100.000 being diagnosed each year. With patients generally being above 60 years old at the time of diagnosis [86].

A CML diagnosis was for a long time a death sentence, unless treated with BM transplantation. However, approximately 20 years ago, the tyrosine kinase inhibitors (TKIs) were introduced. Notably, the first TKI introduced was imatinib, which have since been followed by dasatinib, nilotinib, bosutinib and ponatinib. This revolutionized the treatment of CML, and the disease is now considered a semi chronic condition with near normal life-expectancy [96]. However, this favorable prognosis is only achieved through continuous treatment using TKIs for the

remainder of the patient's life, which is both economically costly and can cause side-effects leading to a reduced life quality for the patients. Common side-effects of TKI treatment include fatigue, cramps, nausea and joint pain [97].

These side-effects together with the high treatment costs have led researchers to investigate if treatment can be discontinued. Interestingly, in patients with a deep molecular response, i.e., low or undetectable levels of BCR-ABL1 following TKI treatment, some patients can achieve treatment-free remission (TFR) with deep molecular response following more than two years of discontinuation, suggesting that treatment can be stopped in specific cases [98]. This is supported by a study with the largest patient cohort to date (n=755) where 50% achieved TFR after 24 months of treatment [99]. While these studies show that some of the patients with good response can be operationally cured of CML. However, a large majority of patients never reaches TFR.

The relapse of TKI treated patients in deep molecular response [98, 99] together with the observation of CML cells being able to engraft in murine models [90], have led to the hypothesis that the CML is derived from an LSC which can reconstitute the leukemia upon treatment cessation. To find and specifically target CML LSCs extensive characterization of the primitive CML population have been performed. Much of this characterization have relied on developing immunophenotypic definitions of the LSCs to remove interference from normal HSPCs residing in the BM. Several cell surface markers have been proposed to define the LSCs, and early studies showed that the LSCs expressed CD34 and were low or negative for CD38 [90]. However, these markers are also used to define primitive normal progenitors, thus additional markers are needed to distinguish between normal and CML stem cells. Several markers have been suggested to have these properties, among them are CD33 (SIGLEC3), IL1RAP, CD26 (DPP4), CD25 (IL2RA), CD36, CD123 and CD117 (c-KIT) [100-107].

Although not all these markers are exclusive to CML cells, CD33 is upregulated while CD117 is downregulated in CML, compared to normal HSPCs. Furthermore, CD26 has been shown to be temporarily downregulated during TKI treatment only to bounce right back up upon treatment discontinuation. This precludes these markers from being used to directly target the LSCs, however they could be used in combination to prospectively isolate LSCs for molecular and functional interrogation.

B-ALL

ALL is a leukemia mainly effecting children and adolescents where adults only contribute to about 20% of the cases [108]. Moreover, ALL represents 85% of the childhood leukemia where B-ALL is the most common form, representing approximately one third of all cases of ALL [85, 109]. B-ALL is marked by an over

proliferation of B-cell precursors, leading to symptoms such as fatigue, bruising and fever [108].

B-ALL is mainly caused by chromosomal abnormalities such as translocations or hyperdiploidy [85], where the ETV6-RUNX1 is the most common translocation, followed by BCR-ABL1 and MLL fusions. Furthermore, ETV6-RUNX1 and MLL translocations have been shown to occur in-utero, suggesting a fetal origin of some specific leukemia subsets [87, 110].

Cell identity, cell states and cell types

The identity of a cell has classically been determined by its function, which have been interrogated through in-vivo and in-vitro assays [2]. However, with the advent of modern molecular analysis it has been possible to study the molecular drivers of cell identity in an unprecedented detail. Through these molecular assays, several genes have been described to mark certain cell types within the HSPCs (Figure 5). Their function has been defined using genetic and pharmacological perturbations together with in-vivo and in-vitro read-outs to determine which lineages and features are affected.

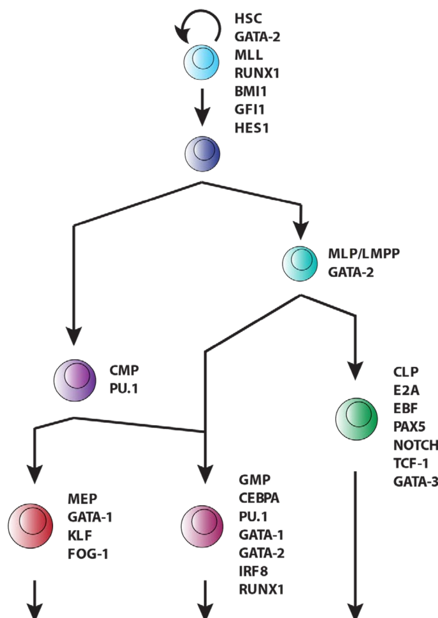


Figure 5. Important genes within early hematopoiesis regulating cell identity. Adapted from [2, 38]

Through these studies it has been shown that the molecular make-up of a cell determines its function and potential. However, the function of a cell is also determined by its state. For instance, it has been shown that HSCs have a reduced self-renewal capacity and ability to engraft while actively cycling [111, 112], whereas dormant HSCs provides the best long-term engraftment [113, 114]. While HSCs cycle infrequently, with estimates being from every three months to around two years [115, 116], it has been suggested from in-vivo studies that after four rounds of division the HSCs lose their self-renewal capacity [114, 117]. The HSC's exit from dormancy has been shown to correlate with CDK6 expression, where low CDK6 expression correlates with dormancy [118]. Other features have also been connected to dormancy, including metabolic state, low MYC expression, low protein synthesis and high p57 expression [119-121]. While cell cycle state effects the potential output of the HSC, it has been shown that HSCs can return to dormancy following stress hematopoiesis [114, 122]. This implies that a single HSC can give rise to different outputs depending on the time of investigation.

Similar results have been shown for lineage commitment, where a single aged HSC gave rise to myeloid restricted progeny within the primary recipient while upon secondary transplantation was able to produce multi-lineage reconstitution [123]. Although the exact mechanism behind this altered phenotype has not thoroughly investigated, it provokes an interesting point of the relative plasticity of lineage commitment and the effect of cell states.

Interestingly, studies have shown that lineage commitment might occur earlier than previously thought, with cells being committed as early as within the immunophenotypic HSCs [23, 124-128]. However, cells may also retain the potency for other lineages much further down in the hierarchy where supposedly lineage restricted progenitors have shown to be able to give rise to cells of an alternative lineage. Although some studies have been able to detect molecularly bi-/multi-potent progenitor populations, where lineage affiliated genes of opposing lineages have been found to be co-expressed [40, 42, 50, 129], others have failed to detect these multipotent states using sc-RNAseq [43, 60]. There could be several explanations for this discrepancy, including methodological differences as well as the cell type investigated. However, the molecular events underlying lineage commitment and restriction is crucial to understand development and disease and therefore further investigation of lineage priming and commitment will be of importance.

The lack of definitive proof of the existence of molecular bi-/multi-potent progenitor populations have led to investigation of other molecular features, where methods investigating chromatin accessibility have become prominent. Many of the initial studies relied on measuring mRNA of genes intimately linked to lineages and cell identities, thus defining cell types and states based on their transcriptional profile. While in reality cell identity is defined by many other features, all contributing to defining a cell's specific functions [130, 131].

Epigenomics, Transcriptomics and proteomics

Cell identity and molecular make-up can roughly be divided into five general omics categories: Genome, Epigenome, Transcriptome, Metabolome and Proteome [132]. Genomics revolves around studies of the DNA-sequence and in the context of hematopoiesis the main genomic focus has been on mutational differences between healthy and diseased states. In leukemia this induces studies on how translocations effect different cell states and populations with the goal to identify mutational drivers of the disease. Where genomics studies the actual DNA-sequence, the field of epigenomics revolves around the chromatin states and structure of the genome, i.e., which parts of the genome are in use within a given cell or cell-type [133, 134]. Transcriptomics investigate which RNA species is being transcribed from the genome [135] and proteomics investigates which proteins are present and their relative frequencies [136]. Finally, the metabolomics studies the metabolites present within a cell, which is the end-product of the cellular processes, and how these affects a cell's growth, survival and function. Where epigenetics, transcriptomics and proteomics have been extensively used to define cell identity, metabolomics have mainly been used to describe cell states.

How a genome containing several billion base-pairs can be correctly used by a cell is as fascinating as it is complex. The solutions lies, to a large extent in compartmentalization, where the DNA is condensed into chromatin complexes of nucleotides wrapped around histone proteins that in turn can be further structured, ultimately forming the chromosomes and nuclear structure [134]. Transcription of DNA into RNA requires the chromatin to be accessible and the accessibility is controlled by transcription factors (TFs), chromatin re-modelers, histones and chromatin-modifying enzymes [134]. TFs are proteins recognizing specific DNA sequences, allowing them to directly bind to DNA to assert their function. Incidentally, many of the genes shown to be lineage and cell type specific within the hematopoietic system are TFs. Some TFs, referred to as pioneering factors, can recognize DNA-sequences and bind to inaccessible DNA which in in turn leads to the recruitment of additional TFs and chromatin modifiers to open the chromatin. This allows for, for example, transcription start sites (TSS) and gene bodies to be accessible for transcription [133, 137]. Additionally, other sites within the chromatin, although not gene coding, are also accessible. These sites are known as cis-regulatory elements and include promoters, enhancers and silencers [138].

Where promoters sit immediately upstream of the TSS and exert the main transcriptional control, the enhancers and silencers can be both distal and proximal and either up- or downstream of the gene which they control. Furthermore, promoters are often universally accessible, and it is generally regarded that, the cis-regulatory elements regulate cell-type specific gene expression. The underlying explanation for this is believed to be that cis-regulatory elements are often bound and controlled by lineage-determining TFs whereas promoters are under control of

a more basal transcriptional machinery [139]. Silencers repress the expression of the gene while enhancers, like their name implies, enhances the expression [138]. To exert their function, enhancers and silencers physically interact with the gene through folding of the chromatin moving distal elements into close proximity of the promotor. A gene can be controlled by several different enhancers, thereby achieving fine tuning of transcription.

By integrating over 434.000 reported enhancers from four different databases [140-143] the GeneHancer database could define approximately 285.000 human enhancers [144], compared to the approximately 20.000 protein coding genes [145] the complexity of the enhancer landscape, where several enhancers can control each gene, allows for interrogation of the molecular heterogeneity between cell populations with a much-improved resolution. Interestingly, the accessibility of enhancer elements is to a large majority cell-type specific and have been able to resolve the human HSPC populations to a higher degree compared to transcriptional data [146]. Moreover, certain types of transcribed enhancers have been shown to proceed coordinated RNA transcription during differentiation [147]. This implies that the enhancer landscape could more readily capture the lineage potential, cell states and cell identities compared to gene centric approaches.

However, one large caveat of utilizing the enhancers to define lineages, cell states and cell identities is that we simply know much less about them compared to regular genes, both in terms of function and genomic location. Where genes and their functions have been meticulously investigated for the last decades with established methods, the importance of enhancers and the techniques to thoroughly study them are just emerging. For example, it is not fully understood how the loss of one enhancer effects the entire landscape, and if any redundancies exist within the system. However, with novel methods as CRISPR-Cas9 genome editing, enabling specific point-mutations, deletions and repression, the functional relevance of specific enhancers can be elucidated [148].

After RNA is transcribed, it is not simply translated to protein but also subject to splicing. The nascently transcribed RNA contains both introns and exons, where splicing is used to remove the introns to form the final mRNA which first then can translated into protein. Moreover, most proteins also have isoforms formed through alternative splicing [145, 149]. Alternative splicing occurs when some of the exons within the RNA are included or excluded, resulting in an isoform of the protein. These isoforms have different functions, thus increasing the biodiversity and complexity generated from one gene and approximately 95% of all genes has an alternatively spliced product [149, 150]. One example of alternative splicing within the hematopoietic system is the protein CD45 encoded by PTPRC gene. CD45 is a pan-hematopoietic marker, marking all hematopoietic cell types except the erythrocytes and plasma cells. While its isoform CD45RA is used to separate the HSPCs into LMPPs, GMPs and CLPs [35, 36, 40]. Emphasizing the importance of alternatively spliced products and the resulting isoforms.

Additionally, some proteins are also post-translationally modified (PTM). Cells often utilizing PTMs to increase protein diversity, without the need for upstream genomic information, since these modifications can drastically alter the properties and interactions of proteins, one common example of this is methylation and acetylation of histones [133]. These PTMs can in turn affect chromatin accessibility and TF occupancy, thus creating a full circle back to the DNA and its structure. The molecular mechanisms controlling cell identity is complex and we have just begun to unravel how epigenetics, transcriptomics, and proteomics interconnected and how they together define the identity and function of a cell.

To explain how these mechanisms define how a cell differentiate from an HSC to an effector cell, the analogy from Waddington can been extended. Here the hills and valleys within the Waddington landscape is created by genes, proteins, and epigenetic factors. Thus, the possible routes the ball can roll down is decided by where it currently is i.e., its cell identity, and what hills and valleys are in front of it i.e., its potential.

Methods commonly used to study hematopoiesis

The study of human hematopoiesis utilizes a myriad of techniques to investigate the function, potential and identity of hematopoietic cells. Here follows a brief description of the main methods used within this thesis together with some of their caveats and interpretations.

Flow cytometry and fluorescent activated cell sorting (FACS)

Few technologies have influenced the field of hematopoiesis as much as flow cytometry and its subcategory FACS analysis [2, 151]. Flow cytometry and FACS utilize lasers in combination with liquids in laminar flow to assay the light scattering and fluorescence spectra of cells. By staining cells with monoclonal antibodies conjugated with fluorophores the relative frequencies of different proteins in the cells and on the cell's surface can be measured.

While the early FACS and flow cytometers could only measure relatively few markers modern instruments can assay up to 40 different fluorophores [152]. This can be achieved by combining multiple lasers of different wavelengths, filters and detectors, allowing for the high complexity of the hematopoietic system and its cell types to be investigated. Additionally, by using FACS, cell types expressing various combinations of different proteins can be prospectively isolated. This allows for enrichment of rare cell fractions which subsequently can be investigated for their function and molecular signature without interference of confounding cell types. Advances in FACS and flow cytometry technologies have paved the way for breakthroughs within the hematopoietic field, making the two highly intertwined.

In-vivo analysis of human HSPC

As described previously, the HSC is defined by its ability to self-renew and differentiate into all hematopoietic lineages. To study this capacity of HSPCs in-vitro transplantation assay has been extensively used [1, 2, 22, 130]. Here the host is myeloablated, normally either by lethally or sub-lethally irradiation to remove hematopoietic cells and to make room for the transplanted cells in the BM niche and mitigating graft-host complications. The investigated cell populations are then typically transplanted into the host by injecting them into the blood stream allowing the cells to migrate into the BM to engraft.

Once inside the BM, the HSCs can start to self-renew and differentiate to repopulate the hematopoietic hierarchy. The engraftment is measured within the blood stream by repeated bleedings to investigate which cells are produced. Here flow cytometry or FACS is used to separate any remaining host cells from the transplanted cells, as well as using different cell surface markers of differentiated cells to evaluate the

lineage potential of the transplanted cells. Interestingly, the self-renewal capacity of HSCs can be propagated through serial transplantations, where the engrafted cells from the primary recipient can be transplanted into secondary and even tertiary recipients. Through this, HSCs has been operationally defined as being able to engraft and produce multilineage reconstitution for 16 weeks in a primary recipient and additionally for 16 weeks in a secondary recipient.

This highly structured and formalized view of hematopoiesis and HSCs have allowed researchers to study the murine hematopoietic system to large extent. However, the human counterpart has been lagging behind, mainly due to the inability to perform in-vivo experiments because of ethical and practical reasons. To allow for human in-vivo engraftment in an experimental setting, researchers have relied on immunodeficient mouse strains. The earliest successful attempts to engraft human stem/progenitors in mice were performed in 1988, where sustained production of donor B and T cells were achieved in the Scid-hu model [153]. This model only achieved lymphoid engraftment, however, by supplementing the mice with human myeloid cytokines IL3, GM-CSF and SCF allowed for also myeloid reconstitution [154]. Next, by backcrossing the Scid mouse with nonobese diabetic (NOD) mice carrying defects in innate immunity, researchers were able to improve engraftment levels of all human cells including myeloid cells [155].

Although, engraftment was improved, and two lineages could be assessed, the NOD-Scid model innately produced activate NK cells and was also predisposed to develop leukemia thereby preventing the study of long-term engraftment. To circumvent these problems, the model was further refined by either a deletion (NSG) or a truncation (NOG) of the IL2RG gene [17, 156]. The loss of IL2RG function prevents development of B-, T- and NK-cells and additionally prevents leukemia initiation, thus creating a mouse model where high long-term human engraftment can be assayed.

Even though, these models have been extensively used to define the function and immunophenotype of human HSPCs, a few caveats remain. Most importantly the murine models still produce a lymphoid biased cellular output, confounding analysis of myeloid biased populations [18]. To alleviate this lineage-bias, several humanized mice models with transgenic human cytokines have been produced [20, 21]. Additionally, there are mouse models where human Mesenchymal Stem Cells (MSCs) are used to create human BM niches in ossicles within the mouse [19]. These humanized models have been able to make the lineage output from human cells more balanced, but how well this captures the true potential of human HSCs is still unknown.

Another caveat is that the transplantation assays is modeling stress hematopoiesis, where the HSC is required to reconstitute the entire hematopoietic system. While this interrogates the potential output of HSCs it does not reveal how the HSCs contribute to steady state hematopoiesis. Additionally, even though the life

expectancy between human and mouse is substantial, the self-renewal potential is measured over 32 weeks in both species. While this represents approximately 33% of a mouse life it only represents 0.8% of a human life, leading to a huge difference in the time where self-renewal capacity is measured. However, few PhD-studies last for 26 years which would be the equivalent time for humans, thus reaching equivalent time for self-renewal between murine and human would be practically impossible.

In-vitro assays to evaluate cell potential

Even though, the in-vivo assays are efficient when it comes to investigating human HSC self-renewal capacity, they do not capture the balanced output seen with murine transplantations [18]. Additionally, aged and leukemic samples have been shown to give poor engraftment, limiting their use in in-vivo studies. Also, the use of animal models is cumbersome and results in suffering of the mice. To circumvent these issues, several in-vitro assays have been developed.

In general, in-vitro cultures tries to mimic the native environment of the human or murine niche to allow cells to differentiate and self-renew. This is done by adding media containing all the nutrients required as well as cytokines. Cytokines are signaling molecules regulating cell differentiation and growth by binding to receptors on the cell surface. By combining different sets of cytokines, the cells properties can be modulated to produce different outcomes i.e., lymphoid-, myeloid-, erythroid differentiation or maintenance of self-renewal [157-159]. This makes the in-vitro culture an effective and versatile tool when investigating cell potential, whereby modulating the content of the in-vitro culture different cell outcomes can be produced.

Even though the in-vitro assays are versatile where differentiation can be controlled through the composition of the culture, there are still a few important issues to consider. For one, the in-vitro cultures rely heavily on the composition of cytokines and media, and thus slight alterations in cytokine concentration could affect the output drastically and balanced output of all lineages is difficult to achieve. Therefore, careful design of culture media composition is needed to prevent one or two lineages to take over, disrupting read out of true lineage potential. Additionally, the immunophenotype of cells are affected by the culture where certain proteins can be up- or downregulated due to the interaction with the plastics used. This can be alleviated by using co-culture systems, where the cells of interest are cultured together with stromal cell lines [160]. The use of stromal cell lines emulates the cell-to-cell interactions usually provided by the niche.

The in-vitro culture systems are powerful tools to interrogate lineage potential, and with its versatility and ease of use many different conditions can be tested. However,

to fully understand the in-vivo function of hematopoietic cells, in-vitro studies are often unsatisfactory.

Single cell molecular technologies

A cell's potential and function are ultimately defined by its molecular signature, where the epigenome, transcriptome and proteome control how the cell responds to external stimuli. Thus, to properly investigate the function of a cell, methods to study the molecular events within a single cell can be powerful, and during the last decade several techniques interrogating different molecular features of single cells has been developed. Early methods suffered from limitations in the small amounts of DNA, RNA and protein within each cell and was therefore performed on bulk populations purified using prospective isolation.

However, measures in bulk are averaged over the entire measured population resulting in that contribution from small subpopulations are often missed. This led to the development of increasingly sensitive methods for assaying the molecular features of cells. With the development of polymerase chain reaction (PCR) and DNA-sequencing suddenly DNA and RNA could be almost indefinitely copied, and their sequences be read to link their abundances to genes and phenotypes. This in turn enabled molecular characterization of smaller and smaller cell populations.

While the bulk molecular analysis was able to define genes and molecular features intimately linked to cell populations, they could not separate the internal heterogeneity of cell population [151, 161]. In-vivo and in-vitro assays suggest that the cells within a given population may not have the same potential and function [26, 28, 40, 41]. It remained unknown though, whether this was an artifact of the assays used or constituted true functional and molecular heterogeneity.

Not until single-cell molecular technologies were developed the molecular basis of this functional heterogeneity could be defined. Indeed, extensive molecular heterogeneity in several of the immunophenotypically defined HSPC populations including CMPs, MPPs, HSCs, GMPs from both human and mouse has been observed [24, 43, 44]. These early studies focused on measuring RNA, however with time a plethora of methods that interrogate accessible DNA, methylation, protein abundance, etc. have been developed.

So why have the single-cell assays become the methods of choice to investigate the molecular underpinnings of hematopoiesis? While the bulk assays offer a high degree of sensitivity (to capture genomic features) it does not dissect the underlying cellular heterogeneity of a population. This becomes an issue when several cell types, each with a unique molecular make-up, constitute a target population. This can be alleviated by sorting purified populations, however, as described previously we know now that even immunophenotypically pure populations have some inherent cellular heterogeneity [24, 44]. Additionally, during fetal development and

leukemic hematopoiesis additional heterogeneity is caused by different on-going waves of differentiation, either the fetal developmental waves or leukemic differentiation [13, 52, 84, 89]. To delineate this heterogeneity single-cell methods can be used.

Single-cell RT-qPCR

One of the earliest single-cell molecular methods implemented in the exploration of hematopoietic heterogeneity was the single-cell RT-qPCR (sc-qPCR) [24, 48, 162, 163]. Here the single cells are sorted into a plate, lysed to release the RNA and by using reverse transcription, the RNA is translated into cDNA. Next the cDNA is generally subjected to several rounds of PCR amplification to increase the amount of cDNA product. Following this the amount of cDNA can be measured using qPCR and the relative amount of mRNA in the cell can be quantified [129, 164]. These early methods relied on regular qPCR for the measurement of mRNA transcripts, limiting the number of genes being investigated. However, with the introduction of microfluidics more genes and cells could be investigated in a high-throughput fashion [162, 163].

Microfluidics relies on small fluidic channels and reaction chambers to perform the qPCR reaction in microliter volumes. The small amount of sample and reactants needed allows for extensive multiplexing of cells and assayed genes. Most commonly either 48x48 or 96x96 chips are used, enabling over 9000 reactions to be performed in one experiment [163]. Additionally, the use of gene specific primers makes the assay highly sensitive and able to detect low expressed genes such as TFs with few false negatives. These gene specific primers also allow for detection of fusion transcripts commonly occurring in leukemia, thus allowing for discrimination of normal and leukemic cells [106].

Although, the use of gene specific primers makes the sc-qPCR highly sensitive, this is also the main drawback of the method. Since sc-qPCR only measures the gene expression from a preselected panel, it requires careful planning and prior knowledge to enable conclusive results [48]. Therefore, the method is appropriate for well characterized systems, or to be used in combination with either bulk or single-cell RNA-seq to enable high quality data. Additionally, even though it was considered high-throughput when it was introduced. 96 cells per experiment is relatively low where sc-RNAseq now has the capacity to measure up to 100.000 cells per experiment [165-167]. In conclusion, sc-qPCR still has its applications, especially when investigating well characterized systems and genes with a small number of transcripts per cell, but it has been largely supplanted by sc-RNAseq in recent years.

Single-cell RNA-sequencing

The introduction of sc-RNAseq revolutionized the study of cellular heterogeneity and has been especially important for the study of primitive hematopoietic

heterogeneity [130, 168], leading to a new theoretical model for hematopoietic differentiation [44, 130]. The advantage of sc-RNAseq compared to its precursor the sc-qPCR is its ability to "unbiasedly" detect all transcripts within each cell, thus not requiring any previous knowledge of the transcripts contributing to cellular heterogeneity. Although, the unbiased nature of sc-RNAseq also has some drawbacks where the methods have shown to be prone to dropouts, i.e., not detecting a transcript [169, 170]. This has led to the development of several sc-RNAseq technologies using different methods for isolation of single-cell and amplification of RNA.

Most sc-RNAseq methods follow the same general workflow, where first the RNA is reverse transcribed into cDNA much like in sc-qPCR. However, while sc-qPCR relies on gene specific primers to detect transcription, sc-RNAseq uses universal primers amplifying all reverse transcribed RNAs. This allows for unbiased amplification of all transcripts. Next the cDNA is fragmented into smaller fragments which are then given sequencing adapters allowing the sequences to be read by Next Generation Sequencing (NGS). The transcripts from individual cells are also tagged with cell specific barcodes before sequencing. After sequencing the transcripts originating from each single cell are separated based on their attached barcodes, enabling the transcriptomic profile of each cell to be interrogated.

In any single-cell library preparation protocol, the method used for isolation single cells influences when cell barcodes are introduced. Therefore, the different sc-RNAseq methods are commonly divided by their method of cell isolation. Three broad categories can be used to describe the cell isolation strategy; plate-based, droplet-based or through combinatorial indexing. In the plate-based systems cells are either hand-picked or FACS sorted into each well where the cells are uniquely marked with cell indexes followed by amplification. The plate-based format allows for extensive variation in the preparation steps and additional manipulation by the end user. This have resulted several different plate-based sc-RNAseq protocols [171-173], where the SMART-seq2 have become the dominant method, partly due to its high sensitivity [174] and ability to read full length transcripts, enabling isoform detection [171, 174]. Additionally, the use of FACS single-cell sorting in most plate-based scRNA-seq methods, enables the protein expression of each cell to be captured using index-sorting. This provides an additional layer of information which can be used correlate the RNA expression levels to the immunophenotype of the cell.

While the plate-based methods offer high sensitivity and the ability to link RNA expression to immunophenotype, they are limited in the number of cells captured. To circumvent the issue of throughput the droplet-based methods were developed. The droplet-based systems rely on microfluidics to create water in oil emulsions, and many scRNA-seq methods employ this method [165, 166, 175]. Briefly, the microfluidics captures cells in a microliter droplet containing all reagents needed for reverse transcription and cDNA synthesis, including cellular barcodes. These

droplets are collected and thereafter reverse transcription and cellular barcoding is carried out within the droplet. Next the cDNA is released from the droplets and amplification, fragmentation and adaptor ligation are performed enabling the resulting library to be sequenced. The advantage of the droplet-based system is the throughput, where thousands of cells can be prepared in one experiment.

The high throughput of cells is also the advantage of the combinatorial indexing approach were sequential indexing and mixing is used to uniquely label each cell for sequencing [167, 176]. In brief, bulk populations of cells in different wells are reverse transcribed with a unique barcode in each well. Next, these cells from all wells are pooled and then sorted into a new plate and split over several wells, where secondary indexing as well as a second strand synthesis and tagmentation is carried out. By relying on a split-pooling approach the combinatorial indexing can potentially capture several thousands of cells in one experiment. This can be achieved by only using the standard instruments available in a typical lab without requiring expensive single-purpose machines. But can be laborious and requires extensive pipetting, additionally the capture rate is substantially lower compared to droplet-/plate-based systems [174].

Each of these methods have its own advantages and disadvantages, while plate-based methods offer high sensitivity and the possibility to link transcriptome to immunophenotype. While droplet-based and combinatorial indexing approaches allows for vast amounts of cells to be interrogated enabling atlas level datasets to be generated [59, 167]. Therefore, the different methodologies should be applied to different kinds of research questions, where plate-based are well equipped to interrogate smaller populations with similar cell types. And the high-throughput methods are appropriate when large uncharacterized populations are to be interrogated for small subpopulations.

Single-cell ATAC-seq

RNAseq is a good tool to investigate the transcriptome of single cells and to identify cell subpopulations. However, recent studies have shown that cell identity is largely defined by its epigenetic profile [146, 147]. In many ways epigenetic profiling can be superior to transcriptomic analysis when defining cell identities. The epigenetic profile of cells can be assayed by several different techniques, including ATAC-seq, ChIP-seq, DNase-seq, as well as chromatin conformation capture-seq (3C-seq) based techniques like HiC-seq, among others [177]. These methods interrogate different aspects of the epigenome, where ATAC-seq and DNase-seq investigate open chromatin through DNA digestion/tagmentation, ChIP-seq utilizes immunoprecipitation to interrogate proteins bound to the DNA, including TFs and methylated or acetylated histones. 3C-seq based methods and its derivates can be used to investigate long range chromatin interactions. These methods have been used to understand the chromatin architecture and how the epigenome is regulated [177].

ATAC-seq was recently described as a rapid and sensitive method for assaying accessible chromatin, enabling analysis of far fewer cells than alternative methods at the time [178]. This was achieved by using a hyperactive Tn5 transposase loaded with sequencing adaptors which could both fragment and tag the genome in one step. When this transposase was incubated together with nuclei only the transposase accessible sites were tagmented due to steric hindrance [178]. Next the tagmented DNA can be amplified through PCR using adaptor specific primers, generating a sequencing ready library. These simple steps reduced both processing time and removed steps where material losses was an issue in previous protocols.

Due to its relative simplicity and sensitivity the ATAC-seq protocol was readily adapted to assay single-cells [179, 180]. The scATAC-seq methods relied on the same technical framework as scRNA-seq and thus can also be divided in plate-based, droplet-based and combinatorial indexing categories of cell isolation, where the droplet-based and combinatorial indexing have become the methods of choice due to the capacity of these technologies to assay thousands of cells.

Single-cell multi-omics

The ability to investigate the transcriptome, proteome and epigenome of single cells have transformed how modern studies are performed. However, these methods only measure a specific feature of the cells, i.e., abundances of transcripts/proteins or chromatin accessibility. While this is highly informative, it only allows for inferences of the other states not assayed. To enable several features to be assayed within each single cell the single-cell multi-omics technologies were developed [181].

Early single-cell multi-omics relied on measuring surface protein levels together with the transcriptome, where applications utilized index-sorting to capture the immunophenotype of each sorted cell [44]. This enabled correlation of transcriptional identity to immunophenotypic identity. While this is an inderict multi-omic approach it was limited in the number of surface proteins interrogated and is only applicable to the plate-based methods. Thus, precluding it from being used in the high throughput methods.

To enable immunophenotypic characterization of cell together with high throughput transcriptional analysis CITE-seq, ABseq and REAP-seq was developed [182-184]. These protocols were developed at the same time in different labs and utilized the same principle for detecting surface proteins. Here the cells are stained with specific antibodies conjugated to oligonucleotides which mimics RNA, enabling detection though sequencing. These oligonucleotides contain an antibody specific sequence used to separate different epitopes, together with a PCR handle sequence used for amplification, and finally, a sequence allowing for the oligo to be captured in the drop-based systems, where both REAP-seq and CITE-seq use a poly-A sequence [182, 184].

The antibody specific sequence is between 6-15bp long enabling 10^3-10^9 different combinations, allowing for extensive multiplexing. For example, above 200 different surface proteins have been analyzed together with mRNA in individual cells making the assay a truly multi-omic approach [185].

Other multi-omic approaches combining transcriptome, epigenome and genome have also been developed [181]. To date, the epigenome has been assayed both using chromatin accessibility and methylation, and the genome have been assayed by both targeted and whole genome technologies [181]. Single-cell multi-omic analysis is a relatively novel field and new protocols are developed each day, multiplying the data which can be derived from each cell. By interrogating multiple features of cell identity within a single cell the causal relationship of these features can be dissected to define the identity of a cell more reliably than either method alone.

Analysis of single-cell data

As single-cell methods have evolved from assaying a few features from a small number of cells to assaying thousands of features in millions of cells the demand on the analysis of the generated data has increased. This has led to the development of a multitude of different single-cell analysis tools and at the time of writing this thesis more than a thousand tools for the analysis of scRNA-seq have been published [186, 187]. These tools can perform a myriad of different analyses, ranging from specific analysis to whole pipelines. Common pipelines or toolkits are Seurat [188] and Scanpy [189] which both perform the commonly used steps in single-cell data analysis.

Whereas early qPCR-based studies focused on a set list of predefined genes of interest where co-expression patterns were investigated [27, 161, 163], this required little or no bioinformatics. However, as more genes and cells could be analyzed more advanced bioinformatic methods were required to describe the cellular heterogeneity. First, the sequencing data produced from the sc-RNAseq needs to be aligned to a reference genome to assess which transcripts are expressed and to demultiplex each individual cell from the rest using the cell specific barcodes. Additionally, where sc-qPCR relied on housekeeping genes to assess if the cell passed quality thresholds, the sequencing-based methods offered additional challenges since confounding factors could be caused by read-depth differences between cells, gene coverage and mitochondrial contamination [190]. Therefore, analysis pipelines usually start with a quality control (QC) step, where cells are filtered based on reads/unique molecular identifiers (UMIs) per cells, genes/cells and mitochondrial contamination to only include viable cells.

Another issue which arises with the development of scRNAseq where the need to unbiasedly define the genes responsible for cellular heterogeneity. This issue has been largely solved by investigating the highly variable features, or highly variable

genes (HVGs) in the context of RNA-seq, within the investigated population. How many of these HVGs which are included in the following analysis is arbitrary and will depend on the data. Including too few may reduce the cellular heterogeneity observed and too many may cause technical artifacts to influence the analysis. Another popular method of gene/feature selection is retaining highly dropout genes [191].

Next, Principal Component Analysis (PCA) is usually used to reduce the dimensions of the data [190]. This is possible due to the inherent low dimensionality of single-cell RNA datasets, where co-regulation of genes enables them to be reduced into a few dimensions [190, 192]. This reduced gene expression matrix can then be utilized to visualize the data. Many different visualization methods have been developed, including additional dimension reduction techniques. This is to optimally display the multidimensional data in two- or three-dimensions where each cell receives a coordinate in this low dimensional space to be displayed using a scatter plot. Several different methods for dimensional reduction is used and among the most popular are t-distributed stochastic neighbor embedding (t-SNE) [193], Uniform Approximation and Projection (UMAP) [194], or other graph-based methods such as SPRING [195].

Once a low dimensional representation has been created further downstream analysis can be performed. The downstream processing can include clustering, cluster annotation, trajectory analysis, differential gene expression testing, projection analysis and gene regulatory network analysis [190]. These steps depend on the biological question investigated. However, cells are typically clustered to define cell states/types that explain the cellular heterogeneity within the population. After clustering, the groups are usually annotated, which can be done manually by investigating cluster specific genes or by label transfer [196, 197].

As differentiation has been shown to be a continuous process [44, 130], cell trajectory analysis has risen to prominence. Trajectory analysis tries to infer developmental/differentiation by minimizing the distances between different neighboring cells in the population, thus traversing the cellular heterogeneity. This puts all cells along a trajectory or multiple linear trajectories, which can have several different endpoints signifying branching trajectories. Here each cell will have a score of where in the trajectory the cell resides, which is called a pseudotime value. The pseudotime values of cells can be used to identify temporal (in an arbitrary time scale) patterns of gene expression as cells proceed from relatively undifferentiated to a more differentiated state. Several methods have been developed to perform trajectory analysis [198], with Monocle [199] and Slingshot [200] being frequently used. These inferred trajectories have also been shown to capture true biological relationships, where transcribable barcodes were used to capture the ground state of lineage relationships [201].

While the steps presented here apply to analysis of sc-RNAseq data, conceptually similar steps of analysis are also performed on sc-ATACseq data as well. Although additional pre-processing steps are required for sc-ATACseq, because a pre-defined set of features are not present and has to be defined in a data specific manner. In standard pre-processing pipelines, both RNAseq and ATACseq require alignment of the sequence reads to a reference genome to determine their identities, RNAseq simply counts the number of identified transcripts while ATACseq requires peak calling. Peak calling is used to determine if a site of the genome is accessible, this is done by piling sequences on top of each other and if a genomic region has an increased enrichment of aligned reads a peak is called, and that region added to a list of accessible regions in that dataset [202].

Another challenge of ATACseq is that once the peaks have been called, they only correspond to areas of the genome. Therefore, to make biological sense of these areas they are overlapped with areas of know function such as, gene bodies, promotors and enhancers among others, referred to as peak annotation [202]. Additionally, ATACseq allows for TF footprinting, where short stretches of inaccessible/occluded chromatin flanked by two peaks can be used to detect TF binding. By analyzing the sequence of these inaccessible sites these can be linked to specific TFs through sequence similarity to previously defined TF binding sites [203].

Following peak annotation, the sc-ATACseq data follows similar analysis as RNA-seq, where dimensional reduction, clustering and additional analysis follows. Here sc-ATACseq specific tools are usually applied where chromeVAR, Cicero, Signac, archR and SnapATAC are commonly used pipelines [204-206].

Ethical considerations

As this thesis both concerns the use of animal experiments and the use of human material several ethical concerns can be raised. The use of animals in research is a highly debated topic, where some wants to eliminate it completely while others simply want to reduce the use of animals. The work in this thesis follows the latter approach by implementing the three R's, Replacement, Reduction and Refinement [207]. Where Replace entails using alternative methods to animal in-vivo experiments such as in-vitro assays to investigate the question at hand. The second R "Reduce" stresses the point of when animal experiments are implemented great care should be taken to only use as few animals as possible to get the same result. Additionally, to optimize output from an animal experiment, making sure that as much information one can derive from an experiment is captured. The final R Refinement is to reduce the stress and discomfort of animals in the experiment. This is done by proper handling of animals, and strict definitions of humane endpoints

eliminating unnecessary suffering. As this thesis mainly focuses on molecular characterization of human cells much of these issues can be avoided. However, due to the strict functional definition of HSCs *i.e.,* self-renewal and differentiation over an extended time, these in-vivo experiments could not be avoided. These types of experiments were reduced by extensive analysis of the molecular data to define candidate populations together with utilization of in-vitro assays to analyze differentiation potential.

While the use of human material for molecular analysis reduces the need for in-vivo analysis, it itself presents other ethical considerations. This is due to the use of human samples in both health and disease, where the donor's integrity could be compromised by the misuse of the samples. First and foremost, great care should be taken to inform the donors of what the samples will be used for, allowing them to make an informed consent to the donation. Next, all samples were anonymized to make sure that the identity of the patient could not be associated with the data generated.

However, many of the samples used where subject to molecular analysis utilizing NGS where the genomic make-up of the patient could be assessed. This produces extensive personal data of the individual. It is therefore of upmost importance that this data is kept on secure servers to prevent this data from being propagated into the wider society. While this ensures that the data is not misused, it also prevents the scientific community from assessing it, precluding peer review and the use by other scientists. To circumvent these issues while still preserving the donor's integrity, special databases have been created to enable secure transfer of data. Additionally, processed data which have removed integrity compromising information can be utilized and freely shared.

Aims of the thesis

The general aim of this thesis was to study how the cellular heterogeneity of human HSPCs is constituted and how different contexts influences the heterogeneity. Throughout this study several different contexts have been investigated, among them are development, ageing and leukemia. These states were also investigated utilizing different single-cell technologies including sc-qPCR, sc-RNAseq and sc-ATACseq to investigate the molecular drivers of human HSPC heterogeneity. How these contexts and technologies were used to investigate the heterogeneity of human HSPCs is summarized in these specific questions;

1. How does the heterogeneity of human HSPCs change during ontogeny and ageing? (Paper I and Paper II)
2. What enhancers control HSPC heterogeneity and HSC identity? (Paper I)
3. Can a therapy insensitive CML population be found within the primitive HSPC population? (Paper III)

Summary of articles included in the thesis

Paper I

Sc-genomic- analysis, including both sc-RNAseq and sc-ATACseq have suggested that the early stages of the human hematopoietic hierarchy forms as a Continuum of LOw primed Un Differentiated (CLOUD) HSPCs [44, 45]. However, the molecular features which are responsible for this cellular heterogeneity have not been extensively studied. Additionally the hematopoietic system have been shown to change drastically during aging [67], where the immunophenotypic HSCs have been shown to increase in frequency while losing potency, demonstrated by impaired engraftment capacity of myeloablated hosts [12, 69]. As well as a shift of lineage output from balanced to myeloid biased [66, 67, 72]. The mechanism behind these age-related changes is still an active area of research, where both cell intrinsic and extrinsic changes are thought to be drivers [67, 74]. Moreover, these studies of age-related changes of HSPC heterogeneity have so far relied on immunophenotypic characterization rather than sc-genomics.

To investigate the molecular features responsible for the cellular heterogeneity we performed CITE-seq [182] and sc-ATACseq [179]. CITE-seq allows for simultaneous investigation of the transcriptome and "surface-ome" of the cells while sc-ATACseq can interrogate the epigenome. To investigate the surface-ome of human HSPCs and how it changes during ageing, we first performed immunophenotypic screens of 342 cell-surface markers in both CB and aged BM (aBM). The markers which were able to divide the immunophenotypic HSC population together with markers which showed differential expression between CB and BM were selected for CITE-seq analysis together with conventional markers of HSPCs.

CITE-seq experiments were performed on CB, young BM (yBM) and aBM to investigate the changes in heterogeneity associated with ageing. A reference map of yBM HSPCs were created on which the CB and aBM HSPCs were projected using SCARF [208] to directly compare the HSPC heterogeneity at different stages of life. Of note this showed an enrichment of molecular HSCs with age, echoing the results achieved by previous immunophenotypic characterizations [66, 72]. Additionally,

we observed that ageing resulted in an enrichment of differentiated Ery and My primed clusters together with a depletion of Ly-primed clusters. Interestingly, multipotent populations also showed a marked depletion in aBM, mirroring previous results where multipotency have been show to decrease with age [40]. Next, to compare the molecular drivers of ageing we performed cluster specific differential gene expression testing, between CB, yBM and aBM. This showed a continuous upregulation of ageing-related genes from CB to yBM and finally to aBM. These results imply a continuous nature of ageing where the ageing process starts early and is propagated during life. Similar results have recently been shown in murine fetal development where no active switch between fetal and adult hematopoiesis occur but rather a continuous upregulation of adult phenotype was observed [63]. These results suggest that the HSPCs are continuously ageing, introducing global effects on gene expression and heterogeneity.

Next, we wanted to investigate the molecular underpinning of the human HSPC heterogeneity. Recent studies have shown that cis-regulatory elements are superior to mRNA analysis in defining cell types [146]. Additionally transcribed enhancers have been shown to lead transcriptional change during differentiation [147]. These findings motivated us to investigating the epigenetic landscape of human HSPCs, focusing especially on enhancers.

To investigate the epigenetic landscape of human HSPCs we performed sc-ATACseq on yBM CD34+ cells. This produced a map similar to the CITE-seq derived yBM map, with cell states related to My-, Ery- and Ly-potential, together with several primitive clusters, related to HSCs and MPPs. Both the CITE-seq data and the sc-ATACseq data was subjected to trajectory analysis, which expectantly showed similar results. However, an earlier separation of the megakaryocyte/erythroid (MegE) trajectory from lympho-myeloid counterparts were observed, suggesting that epigenetic profiling more readily captures differentiation states.

In light of this increased sensitivity, we chose to investigate how enhancer accessibility changed along the differentiation trajectories. Through this analysis we defined 650 enhancers that were enriched within the most primitive populations of all trajectories analyzed, possibly defining HSC identity. Accordingly, these enhancers were enriched for previously defined HSC-related TF binding sites (TFBS), including HOX-, STAT- and SOX- family members. Next, to enable functional validation of the enhancer signature the sc-ATACseq data was integrated with the CITE-seq data, revealing that the enhancer signature could be enriched by CD11A- and CD35+. Indeed, sc-ATACseq analysis of prospectively isolated Lin-CD34+CD38-/lowCD35+ cells resulted in significant enrichment of the HSC enhancer signature.

To functionally validate the HSC-related enhancer signature we prospectively isolated CD35+ and CD11A- subfractions of HSPCs and performed both single cell

in-vitro differentiation analysis and in-vivo xenotransplants. Importantly, enrichment of HSC potential was noted in the CD35+ population that harbor the HSC enhancer signature, where self-renewal and multilineage output was observed. Interestingly, CD11A expression showed a continuum of lineage potential, where low expression correlated with erythroid potential while high expression was linked to lympho-myeloid potential.

In summary this study both investigated the age-related changes in heterogeneity of human HSPCs and revealed a set of 650 HSC-related enhancers. These enhancers could be captured by CD35, enabling further characterization of these enhancers.

Paper II

ALL is a leukemia mainly associated with infants and children and has been shown to be initiated in-utero [85, 87, 110]. To understand the processes behind infant and childhood leukemia initiation the differences between the fetal and adult counterpart hematopoiesis must first be understood. Therefore, extensive studies have been performed to investigate how the hematopoietic system develops [13, 52]. However, these studies have mainly been performed on murine development and has shown that the hematopoietic system develops in three waves [13, 52]. The dHSCs are not formed until the third and final wave, preceded by waves of HSC independent erythroid, myeloid and lymphoid cells [13, 50-52]. The dHSCs migrate into the FL where they rapidly expand before finally moving into the BM where they stay for the remainder of life [16, 52, 62].

The advent of sc-methods has resulted in accumulated knowledge of human fetal hematopoiesis at the single-cell level [58-60, 209]. However, direct comparisons of the heterogeneity differences of adult and fetal HSPCs have been largely overlooked.

To investigate the heterogeneity of early human primitive hematopoiesis we performed CITE-seq [182] on Lin-CD34+ FL cells from Carnegie stage 16 (CS), CS22 and 9 post-conception-weeks (pcw) timepoints. We constructed a reference UMAP from CS22 cells, since these represent the middle timepoint. As expected, we detected clusters related to erythroid, myeloid and lymphoid lineages as well as primitive cell types representing HSC and MPP populations. CITE-seq enables unbiased direct linking of mRNA expression ro immunophenotype without previous enrichment of populations, we investigated classically defined immunophenotypic populations and how well they captured the underlying cell heterogeneity. Interestingly we noted that markers of myeloid cell populations, CD123 [41] and CD135 [27], showed promiscuous expression within the population, where CD135 were generally expressed on all cells, while CD123 were expressed within the lymphoid cell populations. These data showed that both these markers do not reliably capture myeloid primed cells within early human hematopoiesis.

In light of these findings, we proceeded to investigate the capacity of lymphoid associated cell surface markers, CD45RA, CD10, CD7 and IL7RA, to capture lymphoid cell types [27, 34, 35, 210, 211]. By focusing on the CLP population and using a combination of CD38+CD45RA+ cells together with either CD10+, CD7+ or IL7RA+ populations it was evident that IL7RA capture most of the late B-cell primed population, while CD10 marks the more primitive lymphoid primed cells. While CD7 captured both lymphoid clusters almost equally, together with myeloid primed cell types. These data combined with the data of myeloid associated markers, emphasizes the need for careful consideration when comparing immunophenotypic populations across ontogeny.

To perform direct comparison of human cells across ontogeny we utilized the CB and yBM data from Paper I. By projecting the FL CS16 and 9pcw samples together with the CB and BM samples upon the CS22 reference map using scarf [208], we noted several changes in heterogeneity. Of note, the HSC population was enriched with age, while lymphoid populations were decreased mirroring the results seen in Paper I. Interestingly, a multipotent progenitor population referred to as MPP-I received almost no mapping from CB and BM cells, while peaking in CS22. To further investigate the molecular signature of these MPP-I cells differential gene expression analysis was performed compared to the two other primitive populations (HSC and MPP-II). This showed an increased expression of heat shock genes, and gene signatures related to erythro-myeloid cell types. Additional GSEA analysis of MPP-I compared to HSC and MPP-II cells showed enrichment of Myc targets, MTORC1 signaling, unfolded protein response and oxidative phosphorylation. This data points to the MPP-I being a transient fetal derived erythro-myeloid progenitor population. Unfortunately, none of the cell surface makers included in the CITE-seq panel could be used for prospective isolation of this population, precluding further functional validation.

Next, we utilized the projection analysis to investigating the molecular differences between FL CS22 cells and BM cells. BM cells were assigned to a cluster in the FL reference map based on their mapping, and the BM and FL cells in corresponding clusters where then subjected to differential gene expression analysis. This allowed us to define the differentially expressed genes of FL an BM independent of ontogeny related variations in immunophenotype. Through GSEA analysis, an upregulation of proliferation signatures and Myc targets were observed in FL cells, while immune signaling were downregulated.

By focusing on the genes commonly upregulated in FL together with genes enriched in BM a fetal gene signature could be generated. Since childhood B-ALL have been shown to be initiated in-utero [85] we hypothesized that the fetal gene signature would be enriched in B-ALLs of fetal origin. To test this hypothesis, we used a publicly available data set of 2000 ALL patients [212]. First, we investigated patients with the KMT2A-AFF1 (MLL-AF4) translocation, a translocation known to induce B-ALL in infants, children and adults. By applying our fetal gene signature, the samples could be separated by age, indicating that the infant leukemias retained the fetal signature upon leukemia initiation. To validate these findings, we interrogated other driver mutations which replicated the results achieved in MLL-AF4.

In summary, this study investigated the heterogeneity changes occurring during ontogeny, showing an enrichment of HSCs and depletion of lymphoid progenitors during development. Additionally, the conventional surface markers used to purify progenitor populations within human HSPCs are subject to developmental dependent changes. Finally, we showed that childhood leukemias are enriched for a fetal derived gene signature.

Paper III

CML is a leukemia driven by the BCR-ABL1 translocation, the resulting fusion gene is a constitutively active tyrosine kinase. Through the proliferation-promoting actions of this fusion protein a clonal expansion of myeloid progenitors with defective differentiation occurs, leading to an accumulation of undifferentiated progenitors. This causes a life-threatening condition and prior to the introduction of TKI therapy CML was associated with poor prognosis. Nowadays, patients receiving TKI therapy have an almost normal life expectancy, nevertheless if treatment is discontinued patients relapse [99, 213]. This is thought to occur due to therapy resistance of subclones with the capacity to self-renew and reconstitute the entire leukemic system following therapy. Due to these stem cell like properties of these therapy resistant cells they are referred to as LSCs.

LSCs have been shown to reside within the hematopoietic CD34+CD38- population, that also harbor normal HSCs, MPPs and LMPPs [90]. The immunophenotypic similarity of HSCs and LSCs have led researchers to try to find additional surface markers which separates LSCs from normal cells. These studies have defined several markers which enables enrichment of leukemic cells from normal, including IL1RAP, CD26, CD25, CD33 and CD123 [100, 101, 103, 105, 107]. Despite their ability to sperate leukemic and normal cells the specificity of these markers for distinct subpopulations within the CD34+CD38-/low population have not been fully resolved. Additionally, previous studies have relied on immunophenotypic characterization of CML at diagnosis, thus the potential of these conventional CML-markers to capture TKI-insensitive cells have not been fully investigated.

Here, we resolved the heterogeneity of the primitive LSC population at diagnosis and at one to three months of TKI treatment. First to enable discovery of novel LSC specific markers, immunophenotypic screens of 342 individual surface markers were performed. Patient samples at diagnosis and normal healthy age-matched controls were compared to find markers specifically up or downregulated within the CD34+CD38-/low population. Following individual validation, nine markers were shown to be aberrantly expressed in CML, where CD11c, CD25, CD26, CD32, CD276, IL1RAP, ITGB7, and TIM3 were upregulated while CD117 was downregulated.

Next, to interrogate the heterogeneity of the primitive CD34+CD38-/low population we performed sc-qPCR. Here, single cells from patients at diagnosis and following one to three months of TKI treatment together with age matched normal controls were sorted for Lin-CD34+CD38-/low stem and progenitor cells. These were then subjected to sc-qPCR analysis against a panel of 96 gene-specific primers. The panel covered housekeeping genes, stem cell-related genes, lineage specific genes, cell cycle genes, two BCR-ABL1-specific primers, as well as cell surface genes selected

either from the immunophenotypic screens or from literature as established CML LSC markers.

In total, we established a dataset of 2151 single cells, which were clustered into seven subpopulations, including four clusters with increasing myeloid related signature, one expressing lymphoid related genes, one MegE and finally one primitive population. As expected CML cells were enriched within the myeloid clusters as well as the MegE cluster, while healthy control cells dominated the lymphoid, primitive and early myeloid clusters. Next, to investigate the changes in CML heterogeneity induced by TKI therapy the proportion of BCR-ABL1+ cells in each cluster were compared at diagnosis and following TKI treatment. Interestingly, while BCR-ABL+ cells of other clusters remained relatively constant, we could observe a 5.5-fold relative increase of cells within the primitive population following treatment.

To define the immunophenotype of TKI insensitive LSCs, mRNA expression of the included surface markers was compared to BCR-ABL1 expression before and after treatment. Here all markers except CD32 significantly correlated to BCR-ABL1 expression at diagnosis, while only IL1RAP and CD26 mRNA expression was significantly correlated to BCR-ABL1 following TKI therapy. Additionally, CD117 was negatively correlated with BCR-ABL1 expression at both time-points. IL1RAP in BCR-ABL1+ cells were exclusively expressed within the late myeloid and MegE clusters following therapy, while CD26 was downregulated in most population but still detected on BCR-ABL+ cells within the primitive population. Contrarily, CD117 was completely absent on primitive BCR-ABL1+ cells following TKI treatment.

To further investigate the ability of IL1RAP, CD26 and CD117 to prospectively isolate TKI insensitive LSCs, Lin-CD34+CD38-/low cells from three patients at diagnosis and following three months of TKI treatment were additionally stained with IL1RAP, CD26, CD45RA, CD90 and CD117, and index-sorted prior to sc-qPCR analysis. This enabled direct correlation of molecular signature to immunophenotype. Intriguingly, by linking the BCR-ABL1 positive cells from the molecular analysis, with the FACS data from index-sorting the primitive TKI insensitive population was demonstrated to be captured by the Lin-CD34+CD38-/lowCD45RA-CD117-CD26+ immunophenotype, representing a 10-fold enrichment over conventional Lin-CD34+CD38-/low prospective isolation.

In summary, by combing thorough immunophenotypic screens with sc-qPCR we could show that a primitive TKI insensitive CML subpopulation can be captured within Lin-CD34+CD38-/lowCD45RA-CD117-CD26+ cells. Enabling further studies of the molecular drivers of TKI insensitivity.

Paper IV

In connection with Paper III we also published a method paper detailing the protocol used to analyze single-cell gene expression in combination with immunophenotypic characterization. This protocol described the process from initial cell staining to final correlation of gene expression profiles to immunophenotype.

In brief, cells are first thawed and stained with the surface markers used for sorting and index-sorting analysis. Here we stained cells with markers for lineage positive cells (CD2, CD3, CD19, CD56, CD123, CD14, CD16 and CD235a), CD34, CD38, CD45RA, CD90 and CD49F. Next cells were sorted based on Lin-CD34+CD38-/low immunophenotype to enrich for primitive HSPC populations, while expression of CD45RA, CD90 and CD49F were recorded by index-sorting. These cells were sorted into 96-well plates, where preamplification of a selected panel of genes was performed. Next, the pre-amplified cDNA, individual gene specific primers and reactants for qPCR were mixed in the 96x96 GE chip. The chip was then loaded onto the BioMark HD system, were sc-qPCR was performed.

Following sc-qPCR the gene expression matrices was loaded into the analysis tool SCExV [214], where clustering and dimensional reduction using PCA was performed. This approach enables heterogeneity analysis, where molecularly similar cells are grouped together, allowing for cell identities to be defined. Furthermore, these cell identities can then be linked to immunophenotype by FACS data from index-sorted cells to the cell identity defined by SCExV.

In summary, here we present a detailed protocol together with a video containing the most important steps within our sc-qPCR analysis protocol. Ensuring increased reproducibility and enabling broad distribution of the protocol to the scientific field community.

General discussion and future perspectives

The hematopoietic system is the one of the most well-characterized adult stem cell systems to date and is the conceptual paradigm to which all stem cells are compared. The unparalleled characterization of HSCs has largely been achieved through extensive studies in the murine system and early adoption of novel technologies [2, 38, 130], where the use of FACS paved the way for HSCs and their progeny to be purified to a large extent, enabling functional and molecular analysis of the drivers of cell identity.

With such a well characterized system one would think that only minor advancements could be made. However, with the introduction of sc-methodologies, especially the sc-RNAseq, the whole paradigm of the hematopoietic system was overturned [44, 130]. Now differentiation is seen as a continuous process instead of the stepwise hierarchical model. The described heterogeneity of HSC and its progenitors have wide effects on the previously acquired body of knowledge, including cell identity, ageing, development and leukemia progression. We therefore sought to capture the hematopoietic heterogeneity in different states of hematopoiesis.

In Paper I we described the immunophenotypic, transcriptional and epigenomic heterogeneity in HSPCs and how these contributes to cell identity. Through this analysis we defined a set of enhancers characterizing human HSC identity, and cells enriched with this signature could be captured by CD35. While in-vivo and in-vitro analysis of these CD35+ cells showed increased stem cell properties the functions of the individual enhancers were not investigated. Perturbing these enhancers and, investigate the contributions of enhancers to the HSC identity, could potentially lead to important new information regarding the unique properties of stem cells. This could be performed by using pooled CRISPR screens where the enhancers could be either removed, inhibited or activated [148]. This would enable identification of enhancers intimately linked to HSC identity and function, which could assist in identifying the determinants of cell identity.

Additionally, in Paper I we defined changes in HSPC heterogeneity caused by ageing. While many of these changes have been observed before through immunophenotypic characterization, such as increased frequency of HSCs and

decrease of lymphoid progenitors [67], we also noted substantial heterogeneity changes within the immunophenotypically defined HSPCs. CB immunophenotypic progenitors had increased frequencies of contaminating multipotent populations as compared to their yBM and aBM counterparts, possibly explaining the difference in multi-/bi-potential seen with age in functional experiments [40]. These results have wider implications since most previous studies in ageing has relied on comparisons of bulk analysis of immunophenotypic populations, likely confounded by the change in heterogeneity. It would therefore be interesting and important to replicate using single-cell methods. Especially interesting is to study the epigenetic alterations connected to ageing, since these have been shown to predispose aged cells to leukemic transformation [77]. Moreover, we observed a gradual acquisition of age-related gene expression from CB to aBM, recapturing results in the murine system [63]. Consequently, it would be of interest to investigate if the same trend of gradual acquisition of aged phenotype could be observed in human fetal development.

In Paper II we investigated the heterogeneity of first trimester FL HSPCs, where we noted a fetal specific MPP population. This MPP was enriched with a erythro-myeloid gene signature, possibly signifying a transient fetal progenitor population from an earlier developmental wave [13]. It would therefore be of interest to investigate even earlier timepoints of development to pin-point the cell population of origin. Additionally, it would be interesting to explore the HSPC population of Downs-syndrome FL since it has been shown that Down-syndrome is associated with increased risk of acute megakaryocyte-erythroid leukemia together with an increase of MegE progenitors [215]. These phenotypic changes could be the result of a clonal expansion of the fetal specific MPP population observed within our data.

Additionally, we noted an enrichment of fetal signature within infant B-ALL patients as compared to adult patients. Suggesting that a fetal gene signature is retained after leukemia initiation. Since this analysis was performed on bulk RNA-seq data it would be of great interest to compare the changes in heterogeneity of HSPCs between infant and adult patients. This would enable investigation of enrichment of specific infant/adult subpopulations.

Finally, in Paper III we detected a primitive TKI insensitive subpopulation in CML HSPCs, which could be enriched using Lin-CD34+CD38-/lowCD45RA-CD117-CD26+. This immunophenotype could be used for several different applications, including, determining therapy response, enrichment of primitive LSCs for molecular characterization and development of targeted therapies. Which would all have a valuable impact on treatment of CML patients.

Throughout this these papers we have utilized methods enabling linking of immunophenotype to transcriptomic analysis. In papers III-IV this was achieved by index-sorting, while papers I-II made use of CITE-seq. Additionally, in paper I we bioinformatically linked the CITE-seq data to sc-ATACseq facilitating

immunophenotypic enrichment of an epigenetically defined subpopulation. All these studies exemplify the power of multi-omic approaches to define cell identity and cell state. It is therefore not surprising that the single cell field is moving towards developing more multi-omic methods.

A plethora of novel multi-omics methods which allows for several different omic-features to be assayed in each single cell have recently been developed [181]. How these novel approaches will affect our view of the hematopoietic hierarchy is still unknow. But they offer an intriguing possibility to identify the drivers of cell identity and cell state. Especially the dynamics of enhancer-gene regulation could be investigated, something which has previously only been studied in bulk [147]. Here, the single-cell resolution could give new insights pertaining to changes occurring within a single cell, and if previous identified dynamics applies there.

One major caveat of all sc-methods is that they capture static snapshots of the cell state. Therefore, the dynamics of differentiation is only inferred from the data using bioinformatic methods [198-200]. These methods are useful to determine molecular transitions during differentiation but does not offer direct evidence of where a cell is heading to and how long it stays in one state. Alternative approaches of directly measuring clonal lineage contribution have been developed based on transcribed barcodes, detectable through sequencing [201, 216-218]. While still in its infancy these technologies offer intriguing strategies to study the cell fates of specific cells as defined by transcriptomic analysis and could be used to inform further refinements of the hematopoietic hierarchy, development and leukemia progression.

While two of the methods relies on inducible mouse models to track clonal lineage relationships [216, 217], the method from Weinreb et al. relies on lentiviral barcoding and could thus be applied within the human setting [201]. However, this method requires ex-vivo manipulation of cells thus precluding use in the native environment. To analyze clonal contribution of cells within a steady state human hematopoiesis somatic mutational analysis has been used [116, 219-221]. These studies utilized single cell sorting of HSPCs followed by in-vitro expansion and whole genome sequencing, enabling identification of cell specific mutations. These mutations were then used in targeted mutational analysis to investigate development and ageing. While approaches like these offer interesting insights in the clonality of hematopoiesis they are precluded from linking cell states to the mutational analysis. Here multi-omic analysis combining genome, epigenome and transcriptome could be used to define cell states and the clonal architecture of the hematopoietic system [181].

Novel technological developments are continuously developed measuring more and more features and cells. As shown throughout this thesis novel technological advances have always led to a new understanding of the hematopoietic hierarchy. Therefore, the study of hematopoiesis will always move forward, and the principles discovered here will influence and deepen our understanding of biology in general.

Populärvetenskaplig sammanfattning

Det hematopoetiska systemet är ett stort och komplext system som innefattar alla blodceller [1]. Blodet i vår kropp har många funktioner, däribland transportera syret vi andas genom kroppen, läka sår och skydda oss från infektioner. Denna stora uppsjö av funktioner kommer från fler olika specialiserade celler som alla kommer från ett och samma ursprung, den hematopoetiska stamcellen [1, 2]. De hematopoetiska stamcellerna bor i benmärgen där den producerar alla blodceller i kroppen. Stamcellen har unika funktioner som tillåter den att dels skapa en kopia av sig själv, vilket gör att stamceller bibehålls genom hela livet. Dels kan stamcellen differentiera, vilket är en process som innebär att stamcellen kan förlora sin förmåga att kopiera sig själv och successivt bli en av de specialiserade cellerna, det vill säga en cell som skyddar oss från infektioner, läker sår eller transporterar syre.

Historiskt har denna process beskrivits som ett hierarkiskt system där stamcellen är överst och under den finns flera underordnade cellpopulationer, och genom att gå ner ett eller flera steg blir cellen mer och mer specialiserad och till slut en av de specialiserade cellerna [2]. Där de celler som är längst upp i hierarkin brukar kallas primitiva, det vill säga odifferentierade. Den hematopoetiska hierarkin har förfinats med flera olika populationer med olika potential, så kallade progenitorer. Detta har gjort till stor del genom att använda specifika protein på cellernas yta för att dela upp dem i olika grupper/populationer. Genom att sedan undersöka funktionen av dessa cellgrupper/-populationer har olika progenitor populationer definierats. Men med nya metoder som tillåter molekyläranalys av individuella celler, så kallad en-cells-analys, har synen på den hematopoetiska hierarkin ändrats. När dessa metoder användes på de primitiva cellerna i det hematopoetiska systemet, det vill säga cellerna längst upp i hierarkin, kunde man se att det inte fanns några tydliga populationer av celler. Det var snarare en stor blandning av celler som kontinuerligt skaffade sig en mer specialiserad identitet, utan att gå igenom några tydligt definierade steg [44, 130]. Denna nya syn på den hematopoetiska hierarkin, där tidigare populationer av celler med specifik funktion nu ses som en blandning av celler med olika funktioner, detta brukar kallas cellulär heterogenitet. Den cellulära heterogeniteten i de primitiva hematopoetiska cellerna påverkar flera tidigare studerade områden inom hematopoes, därför är målet med denna avhandling är att studera dessa i kontexten cellulär heterogenitet.

När vi åldras, åldras också våra hematopoetiska celler. Med åldrande så ökar mängden stamceller, men de blir också sämre på att återskapa vårt hematopoetiska

system [66, 67]. Dessutom ändras vilka celler som produceras av stamcellerna, celler kopplade till det adaptiva immunförsvaret minskas samtidigt som celler kopplade till det nativa immunförsvaret ökar. Där det nativa immunförsvaret är första försvaret som ospecifikt anfaller alla främmande organismer, medan det adaptiva immunförsvaret specifikt anfaller vissa typer av organismer som det har tränats för att känna igen. Dessa förändringar tros skapas inom de primitiva cellpopulationerna. I studie I undersöker vi om dessa funktionella förändringar beror på förändringar i cellulär heterogenitet i de primitiva hematopoetiska cellerna vid olika åldrar. Genom att undersöka heterogeniteten med molekylär en-cells-analys av dessa olika åldrar kunde vi se att det sker substantiella förändringar med ålder. Molekylärt definierade stamcellerna ökade, medan celler kopplade till det adaptiva immunförsvaret minskade med åldern.

Vidare i studien ville vi undersöka vilka molekylära signaler som definierar stamceller och valde att fokusera på enhancers. Enhancers är områden i vårt DNA som kan styra genuttryck och som har visats vara ytterst celltypsspecifika [146, 147]. Därför var enhancers ett bra alternativ för att definiera stamcellers identitet. Genom att använda en molekylär en-cells-metod som gör att man kan detektera enhancers så kunde vi definiera 650 enhancers kopplade till stamcellsidentitet. Vi kunde senare visa att celler som uttryckte ett visst protein som kallas CD35 var signifikant anrikade med dessa 650 enhancers. Detta tillät oss att sortera ut dessa CD35 positiva celler och testa dem för stamcellsaktivitet. När vi undersökte cellerna visades att de både kunde göra kopior av sig själva och differentiera ut till alla hematopoetiska celler. Detta bevisade att celler med denna enhancersignatur har stamcellsfunktion.

I studie II ville vi undersöka hur den cellulära heterogeniteten förändrades under utvecklingen av det hematopoetiska systemet. Det hematopoetiska systemet uppstår innan vi föds, och uppstår i tre olika vågor av hematopoes [13, 52]. Den hematopoetiska stamcellen uppstår i den tredje och sista vågen. I de två tidigare vågorna skapas celler som kan transportera syre och celler kopplade till det nativa immunförsvaret som behövs för fosterutvecklingen. När de första hematopoetiska stamcellerna skapas flyttar de först till levern där de expanderar för att sedan emigrera till benmärgen där de lever resten av livet [16]. Denna process är inte bara intressant ur ett utvecklingsperspektiv utan också i blodcancer, som också kallas leukemi. Anledningen till det är att vissa barnleukemier har visats ha skapats redan under utvecklingen innan barnet fötts [85, 87].

För att undersöka den cellulära heterogeniteten under utvecklingshematopoes utförde vi samma en-cells-analys som i studie I, på primitiva celler i flera tidiga stadier av fosterutvecklingen. Detta visade att de generella populationerna som vi observerat i vår tidigare analys av vuxna primitiva celler också existerade i fosterproverna. Men när vi jämförde den cellulära heterogeniteten närmare mellan de primitiva cellerna i foster deras vuxna motsvarighet kunde vi se en ökning av stamceller med ålder. Samtidigt som fosterproverna hade en ökning av celler

kopplat till det adaptiva immunförsvaret. Ytterligare kunde vi se en fosterspecifik grupp celler av omogna progenitorer som inte fanns i de vuxna proverna. Dessa celler hade en molekylär signal som liknande de syretransporterande och de nativa immunförsvars cellerna. Detta kan tolkas som att denna population av omogna celler kommer från ett tidigare stadie av fosterutveckling, men eftersom vi inte kunde hitta något specifikt protein för denna population kunde vi inte testa cellerna funktionellt.

Härnäst, ville vi undersöka om skillnaderna mellan fostercellerna och celler från vuxna kunde skilja på leukemier som uppstår i vuxna och barn, eftersom vissa typer av leukemi uppstår redan i fosterutvecklingen [85, 87]. För att göra detta så skapade vi en gensignatur som var specifik för de fetala cellerna. Denna signatur kan skilja patienter med en viss typ av leukemi efter ålder. Detta visade på att barn med viss typ av leukemi behåller en gensignatur från fosterutvecklingen när de utvecklar leukemisjukdomen. Detta resultat kan leda till vidare studier som kan påvisa i vilka celler leukemin uppstår och kanske leda till förbättrade behandling för dessa patienter.

I studie III undersöker vi Kronisk Myeloisk Leukemi (KML). Denna leukemi uppstår främst hos äldre patienter och beror på fusionsproteinet BCR-ABL1 [84]. Detta fusionsprotein skapas genom att två olika kromosomer (9 och 22) binds ihop vilket då skapar en blandning av två olika protein - BCR och ABL. Det leder till en överproduktion av vissa cell progenetorer som inte kan differentiera som vanligt, vilket ger en ackumulation av dessa omogna celltyper som då hindrar normal hematopoes. KML kan behandlas med en viss typ av läkemedel så kallade TKI-inhibitorer som stoppar fusionsproteinets aktivitet. När KML-patienter behandlas med sådana läkemedel så får de en mer eller mindre normal livslängd, men det har visast att om de slutar ta läkemedlet så återkommer leukemin. Det har dock visats att vissa patienter kan sluta ta läkemedlet och ändå inte få tillbaka leukemin.

Anledningen till att patienterna får ett återfall tros bero på en grupp av primitiva läkemedelsresistenta KML-celler överlever behandlingen. De cellerna som överlevt behandlingen bygger sedan upp hela KML systemet igen. På grund av dessa förmågor hos de primitiva cellerna har de börjats kallas leukemistamceller [84, 89]. Om dessa celler skulle kunna behandlas specifikt så kan KML-patienter bli helt friska. Problemet är dock att dessa celler ser ut och beter sig väldigt likt de normala hematopoetiska stamcellerna. Vi försöker därför i denna studie beskriva heterogeniteten inom primitiva celler i KML vid diagnos och under behandling för att hitta denna population av primitiva behandlingsresistenta celler. Det kan möjliggöra vidare karakterisering av dem. För att göra det använde vi en-cellsmetoder som kan skilja på leukemiska och normala celler samtidigt som vi undersöker deras molekylära signatur efter primitiv eller differentierad signal.

Genom denna analys kunde vi detektera en liten population av primitiva KML-celler som ökade under behandling jämfört med vid diagnos. Vidare analys av dessa celler

visade att de hade specifika proteiner uttryckta på cellytan. Genom att kombinera dessa ytproteiner kunde vi visa att de kunde anrikas till en hög grad jämfört med tidigare protokoll. Resultaten från denna studie kan användas till att jämföra de primitiva cellerna för att utveckla metoder för behandling av KML. De skulle också kunna användas för att optimera behandling av patienter för att se om de kan avbryta sin läkemedelsbehandling utan att få tillbaka KML.

I kombination med studie III så publicerade vi också en metodartikel som beskriver tillvägagångssättet för att utföra en-cells-analys av KML-celler. Denna metod presenteras i studie IV. Artikeln är av värde för att öka reproducerbarheten av studien och för att tillåta andra forskningsgrupper att utföra samma analys.

Sammanfattningsvis beskriver denna avhandling de nya rönen inom hematopoes och hur en-cellsanalys har revolutionerat fältet. Vi använder dessa metoder för att beskriva förändringar som sker med den cellulära heterogeniteten i förhållande till åldrande och bildandet av hematopoes. Dessutom visar vi hur heterogeniteten förändras inom leukemi under behandling, och hittat en kombination av ytprotein som kan användas för att anrika de celler som är behandlings resistenta för att vidare studera dessa.

Acknowledgement

Very few things in life are solely your own accomplishments, as such this thesis and the work herein would and could not have been done without the help of others. These others are not only the many dear colleagues who have helped with the science but also family and friends who make life into the thrilling experience that it is.

First and foremost, I would like to thank Göran, my primary guide through the wide field of hematopoiesis. Your encouragement and support have been instrumental for my development throughout these years. You've let me explore novel techniques and let me influence where the projects are going, something which has allowed me to develop into an independent researcher and for that I will be ever grateful.

Secondly, I would like to then my many co-supervisors, Micke, Shamit and Ram. While I've not utilized your knowledge and experience to its full extent, it has always felt good to know that you are there to support me if anything was needed. Micke, your work as the department head made the environment a great place to work at during the whole PhD experience. Shamit, your introduction to the field of bioinformatics during my early days here helped me a lot in later stages and let us not forget your excellent reviews of the Stamstället food. Finally, Ram, a great fountain of knowledge, always willing to discuss science and other topics in extensive depths.

Let us not forget the entire group both past and current members, Rebecca, Parashar, Ram, Rasmus, Eva, Mojgan, Linda, Aurelié and Kristin, without you this work could not have been performed. Additionally, all of you have made these years very fun and entertaining. Rebecca and Parashar, our time in Australia were one of the best times during the PhD studies. Parashar, your bioinformatics skills are unparalleled, and our time together have been both very educational and fun. Rebecca, the one to ask when it comes to CML, always great hearing your random stories. Mojgan, you are truly the nicest person I've ever had the pleasure of meeting, hope everything goes will now in the end of your PhD studies. Eva, I don't know how many hours I've bothered you with questions regarding molecular techniques and cell culture, but you have always been willing to help. Linda, always fun to talk with you, and thank you for all your help. Rasmus, the newest addition, it is nice to see how you have developed and I think Tool is going to be amazing! Aurelié arguably the scariest member of our group, and I think our first encounter

where you are cursing me out for taking too long in the cell culture. But I've come to realize that the bark is worse than the bite. Finally, Kristin although retired now, your help ordering and fixing things in the lab helped a lot and made things run smoothly.

While the group and its members has been instrumental for my development and the work presented herein many other people have also contributed. Especially, Charlotta with whom I worked with during paper II, which have been an intense experience where I think we produced one of the fastest drafts of an article ever achieved. Although, prior to us working together I've bothered her extensively, but she has always kindly replied. Jonas U, my "little" gym buddy thanks for not only keeping my mind in shape but also my body, I would never have tried the "chin-up burpy" if it was not for you.

David B and Ewa S, both of you are always very helpful when I come and ask random questions concerning hematopoiesis and different assays. It is great to have both of you here in the department. Christine, thank you for organizing the many great seminars and always being very nice to talk to.

I would also like to take the time and thank the many collaborators I've had though out the years. Each collaboration has been an educational experience, introducing me to such wide areas of science as breast cancer, RNA splicing, erythropoiesis, lung cancer and many more. It has been truly inspirational to see the breadth and depth of science performed.

Also, the other PIs, post docs and PhD student in our department who has given great insights on my projects during department presentations. You have all helped to foster a truly special environment where one can go and ask anyone for tips and tricks on most techniques while also having a good time at and outside of work. I will try to include you all, and any omissions are not a sign of you not being important but only a reflection on how many nice and fun people are in the department. The Italian cluster Giulia, Stefano and Nicola, thanks for teaching me how to eat Italian food the "proper" way, and for not teasing me too much for liking Lambrusco. The Nordic cluster Hugo, Maria J, Trine, Johanna and Alex R, always a pleasure of talking and hanging out with you all. Now for the rest, Mo, Anna C, Magda, Roberto, Ariana, Niklas, Sandro, Dimitra and anyone I've forgotten it has been great getting to know you all!

Of course, any research into hematopoiesis would not be possible without working flow cytometry and FACS machines. Therefore, I would like to thank the wonderful ladies of FACS, Anna F, Zhi and Theia, for almost always having working and perfectly set up machines, you've also been very nice when I've come with stupid questions regarding FACS. Also, the CTG people, Ulrich and Julia, for their expert work with my single-cell experiment. Additionally, I would like to thank the people

who has helped me with irradiation for the in-vivo work, Alex D, Isabell, Hongzhe and Anna R, it has been very nice of you to take the time from your busy schedule to help me. Also, the bioinformatics core, Stefan, Johan Ro and Oscar always nice talking to you. Johan Ri, thank you for all the sample throughout the years, it has been great working with you.

I would also like to thank the awesome people of A12, although I left you for B12 you have always welcomed me back. So, Svetlana, David Y, Alex M, Taha, Els, Pavan, Valli, Sara W and Fabio thank you for providing a nice and welcoming "neighborhood". Here I should also mention the "old-guard" of PhDs who finished while I was just starting, Roman, Maria D, Roger and Carolina, you have all been very helpful when I was young and naive. Additionally, the B9 people Jonas F, Julie and Isaac thanks for your friendship and the good times.

As I now think I've thanked anyone work-related, I will now start with the friends, not to say that the aforementioned people are not included in the friend column. To improve readability, I'll start with the oldest friends and continue to the "fresh" faces of my life. Gustaf, Martin and Rasmus, we've been friends for more than twenty years now and I would not have been the same person without anyone of you. We have done some fun and stupid things throughout the years and some time I think it is a small miracle that we survived. But I'm hoping that we will be in each other's lives for more than double the time we already spent. Carl, Ludvig and William, the three years we spent at Katedral was only the beginning, we have spent many fun hours playing games and talking about everything and nothing, and I think that we will have many more fun years ahead. Anton, Anna "Apa", Daniel S, Filip O and Anders, it has been great knowing all of you! Now for the freshest faces, Jörn, Johan, Daniel A, Filip S and Tora, thoroughly enjoyed the time together in AF borgen. Who knew that drinking beer until the sun went up (during winter in Sweden), would give you lifelong friends?

Emma, thank you for being with me for the last part of this PhD ride, it would not have been the same without you. How you stuck with me during these last few months of finalizing the thesis will always amaze me.

The foundation of anyone lies with the family, and I would definitely not be the person I am today without my family. My brother, thanks for keeping me active by taking me with you for MTB rides, and for your cute kids, hopefully I will have more time to sped with them now. Finally, and perhaps most importantly, my parents to which I owe everything. They have given me one of the best life imaginable, encouraging me to be inquisitive and to apply myself, two important aspects of doing anything really, but also for doing a PhD. Your support during these years have been instrumental.

Thank you!

References

1. Seita, J. and I.L. Weissman, *Hematopoietic stem cell: self-renewal versus differentiation.* Wiley Interdiscip Rev Syst Biol Med, 2010. **2**(6): p. 640-53.
2. Doulatov, S., et al., *Hematopoiesis: A Human Perspective.* Cell Stem Cell, 2012. **10**(2): p. 120-136.
3. Ramalho-Santos, M. and H. Willenbring, *On the Origin of the Term "Stem Cell".* Cell Stem Cell, 2007. **1**(1): p. 35-38.
4. Till, J.E. and E.A. McCulloch, *A direct measurement of the radiation sensitivity of normal mouse bone marrow cells.* Radiation research, 1961. **14**(2): p. 213-222.
5. Becker, A.J., C.E. Mc, and J.E. Till, *Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells.* Nature, 1963. **197**: p. 452-4.
6. Siminovitch, L., E.A. McCulloch, and J.E. Till, *THE DISTRIBUTION OF COLONY-FORMING CELLS AMONG SPLEEN COLONIES.* J Cell Comp Physiol, 1963. **62**: p. 327-36.
7. Wu, A.M., et al., *Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system.* J Exp Med, 1968. **127**(3): p. 455-64.
8. Zakrzewski, W., et al., *Stem cells: past, present, and future.* Stem Cell Research & Therapy, 2019. **10**(1): p. 68.
9. Crane, G.M., E. Jeffery, and S.J. Morrison, *Adult haematopoietic stem cell niches.* Nature Reviews Immunology, 2017. **17**(9): p. 573-590.
10. Thomas, E.D., et al., *Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy.* N Engl J Med, 1957. **257**(11): p. 491-6.
11. Granot, N. and R. Storb, *History of hematopoietic cell transplantation: challenges and progress.* Haematologica, 2020. **105**(12): p. 2716-2729.
12. Wang, J.C.Y., M. Doedens, and J.E. Dick, *Primitive Human Hematopoietic Cells Are Enriched in Cord Blood Compared With Adult Bone Marrow or Mobilized Peripheral Blood as Measured by the Quantitative In Vivo SCID-Repopulating Cell Assay.* Blood, 1997. **89**(11): p. 3919-3924.
13. Palis, J., *Hematopoietic stem cell-independent hematopoiesis: emergence of erythroid, megakaryocyte, and myeloid potential in the mammalian embryo.* FEBS Letters, 2016. **590**(22): p. 3965-3974.
14. Anasetti, C., et al., *Peripheral-blood stem cells versus bone marrow from unrelated donors.* N Engl J Med, 2012. **367**(16): p. 1487-96.

15. Gao, X., et al., *The hematopoietic stem cell niche: from embryo to adult.* Development, 2018. **145**(2).
16. Ema, H. and H. Nakauchi, *Expansion of hematopoietic stem cells in the developing liver of a mouse embryo.* Blood, 2000. **95**(7): p. 2284-2288.
17. Ito, M., et al., *NOD/SCID/γcnull mouse: an excellent recipient mouse model for engraftment of human cells.* Blood, 2002. **100**(9): p. 3175-3182.
18. McDermott, S.P., et al., *Comparison of human cord blood engraftment between immunocompromised mouse strains.* Blood, 2010. **116**(2): p. 193-200.
19. Reinisch, A., et al., *A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells.* Nature Medicine, 2016. **22**(7): p. 812-821.
20. Rongvaux, A., et al., *Human thrombopoietin knockin mice efficiently support human hematopoiesis in vivo.* Proceedings of the National Academy of Sciences, 2011. **108**(6): p. 2378-2383.
21. Nicolini, F.E., et al., *NOD/SCID mice engineered to express human IL-3, GM-CSF and Steel factor constitutively mobilize engrafted human progenitors and compromise human stem cell regeneration.* Leukemia, 2004. **18**(2): p. 341-347.
22. Spangrude, G.J., S. Heimfeld, and I.L. Weissman, *Purification and characterization of mouse hematopoietic stem cells.* Science, 1988. **241**(4861): p. 58-62.
23. Kent, D.G., et al., *Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential.* Blood, 2009. **113**(25): p. 6342-6350.
24. Wilson, Nicola K., et al., *Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations.* Cell Stem Cell, 2015. **16**(6): p. 712-724.
25. Kiel, M.J., et al., *SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells.* Cell, 2005. **121**(7): p. 1109-1121.
26. Notta, F., et al., *Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment.* Science, 2011. **333**(6039): p. 218-21.
27. 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, 2010. **11**(7): p. 585-593.
28. Manz, M.G., et al., *Prospective isolation of human clonogenic common myeloid progenitors.* Proceedings of the National Academy of Sciences, 2002. **99**(18): p. 11872.
29. Civin, C.I., et al., *Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells.* The Journal of Immunology, 1984. **133**(1): p. 157-165.
30. Ishii, M., et al., *Development of a high-resolution purification method for precise functional characterization of primitive human cord blood–derived*

CD34–negative SCID-repopulating cells. Experimental Hematology, 2011. **39**(2): p. 203-213.e1.
31. Bhatia, M., et al., *A newly discovered class of human hematopoietic cells with SCID-repopulating activity.* Nature Medicine, 1998. **4**(9): p. 1038-1045.
32. Anjos-Afonso, F., et al., *CD34(-) cells at the apex of the human hematopoietic stem cell hierarchy have distinctive cellular and molecular signatures.* Cell Stem Cell, 2013. **13**(2): p. 161-74.
33. Bhatia, M., et al., *Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5320-5.
34. Galy, A., et al., *Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset.* Immunity, 1995. **3**(4): p. 459-73.
35. P.M. Lansdorp, H.S., and Cj. Eaves, *Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow.* J Exp Med, 1990. **172**(1): p. 363-6.
36. Majeti, R., C.Y. Park, and I.L. Weissman, *Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood.* Cell stem cell, 2007. **1**(6): p. 635-645.
37. Murray, L., et al., *Enrichment of human hematopoietic stem cell activity in the CD34+Thy- 1+Lin- subpopulation from mobilized peripheral blood.* Blood, 1995. **85**(2): p. 368-378.
38. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology.* Cell, 2008. **132**(4): p. 631-44.
39. Orkin, S.H., *Diversification of haematopoietic stem cells to specific lineages.* Nat Rev Genet, 2000. **1**(1): p. 57-64.
40. Notta, F., et al., *Distinct routes of lineage development reshape the human blood hierarchy across ontogeny.* Science, 2016. **351**(6269): p. aab2116.
41. Mori, Y., et al., *Prospective isolation of human erythroid lineage-committed progenitors.* Proceedings of the National Academy of Sciences, 2015. **112**(31): p. 9638-9643.
42. Pronk, C.J.H., et al., *Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy.* Cell Stem Cell, 2007. **1**(4): p. 428-442.
43. Paul, F., et al., *Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.* Cell, 2015. **163**(7): p. 1663-77.
44. Velten, L., et al., *Human haematopoietic stem cell lineage commitment is a continuous process.* Nat Cell Biol, 2017. **19**(4): p. 271-281.
45. Buenrostro, J.D., et al., *Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation.* Cell, 2018. **173**(6): p. 1535-1548.e16.
46. Nestorowa, S., et al., *A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation.* Blood, 2016. **128**(8): p. e20-e31.
47. Waddington, C.H., *The strategy of the genes*. 2014: Routledge.

48. Karlsson, G., M.N.E. Sommarin, and C. Böiers, *Defining the Emerging Blood System During Development at Single-Cell Resolution.* Frontiers in Cell and Developmental Biology, 2021. **9**(932).
49. Yoshimoto, M., et al., *Autonomous murine T-cell progenitor production in the extra-embryonic yolk sac before HSC emergence.* Blood, 2012. **119**(24): p. 5706-5714.
50. Böiers, C., et al., *Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells.* Cell Stem Cell, 2013. **13**(5): p. 535-48.
51. Ghosn, E., et al., *Hematopoietic stem cell-independent hematopoiesis and the origins of innate-like B lymphocytes.* Development, 2019. **146**(15).
52. Dzierzak, E. and A. Bigas, *Blood Development: Hematopoietic Stem Cell Dependence and Independence.* Cell Stem Cell, 2018. **22**(5): p. 639-651.
53. Ottersbach, K., *Endothelial-to-haematopoietic transition: an update on the process of making blood.* Biochem Soc Trans, 2019. **47**(2): p. 591-601.
54. Yuan, J., et al., *Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis.* Science, 2012. **335**(6073): p. 1195-200.
55. Ginhoux, F., et al., *Origin and differentiation of microglia.* Frontiers in cellular neuroscience, 2013. **7**: p. 45-45.
56. Guibentif, C., et al., *Single-Cell Analysis Identifies Distinct Stages of Human Endothelial-to-Hematopoietic Transition.* Cell Reports, 2017. **19**(1): p. 10-19.
57. Kennedy, M., et al., *T Lymphocyte Potential Marks the Emergence of Definitive Hematopoietic Progenitors in Human Pluripotent Stem Cell Differentiation Cultures.* Cell Reports, 2012. **2**(6): p. 1722-1735.
58. Domcke, S., et al., *A human cell atlas of fetal chromatin accessibility.* Science, 2020. **370**(6518): p. eaba7612.
59. Popescu, D.-M., et al., *Decoding human fetal liver haematopoiesis.* Nature, 2019. **574**(7778): p. 365-371.
60. Ranzoni, A.M., et al., *Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis.* Cell Stem Cell, 2021. **28**(3): p. 472-487.e7.
61. Bowie, M.B., et al., *Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect.* The Journal of Clinical Investigation, 2006. **116**(10): p. 2808-2816.
62. Benz, C., et al., *Hematopoietic Stem Cell Subtypes Expand Differentially during Development and Display Distinct Lymphopoietic Programs.* Cell Stem Cell, 2012. **10**(3): p. 273-283.
63. Li, Y., et al., *Single-Cell Analysis of Neonatal HSC Ontogeny Reveals Gradual and Uncoordinated Transcriptional Reprogramming that Begins before Birth.* Cell Stem Cell, 2020. **27**(5): p. 732-747.e7.
64. Eva, M.-R. and F. Maria Carolina, *Understanding intrinsic hematopoietic stem cell aging.* Haematologica, 2020. **105**(1): p. 22-37.

65. de Haan, G. and S.S. Lazare, *Aging of hematopoietic stem cells.* Blood, 2018. **131**(5): p. 479-487.
66. 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, 2011. **108**(50): p. 20012-20017.
67. Pang, W.W., S.L. Schrier, and I.L. Weissman, *Age-associated changes in human hematopoietic stem cells.* Seminars in Hematology, 2017. **54**(1): p. 39-42.
68. Dykstra, B., et al., *Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells.* J Exp Med, 2011. **208**(13): p. 2691-703.
69. Woolthuis, C.M., et al., *Aging Impairs Long-Term Hematopoietic Regeneration after Autologous Stem Cell Transplantation.* Biology of Blood and Marrow Transplantation, 2014. **20**(6): p. 865-871.
70. Nygren, J.M. and D. Bryder, *A novel assay to trace proliferation history in vivo reveals that enhanced divisional kinetics accompany loss of hematopoietic stem cell self-renewal.* PLoS One, 2008. **3**(11): p. e3710.
71. Janzen, V., et al., *Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a.* Nature, 2006. **443**(7110): p. 421-6.
72. Rundberg Nilsson, A., et al., *Human and Murine Hematopoietic Stem Cell Aging Is Associated with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias.* PLoS One, 2016. **11**(7): p. e0158369.
73. Mayack, S.R., et al., *Systemic signals regulate ageing and rejuvenation of blood stem cell niches.* Nature, 2010. **463**(7280): p. 495-500.
74. Ya-Hsuan, H. and M.-F. Simón, *Microenvironmental contributions to hematopoietic stem cell aging.* Haematologica, 2020. **105**(1): p. 38-46.
75. Ho, T.T., et al., *Aged hematopoietic stem cells are refractory to bloodborne systemic rejuvenation interventions.* Journal of Experimental Medicine, 2021. **218**(7).
76. Beerman, I., et al., *Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging.* Cell Stem Cell, 2013. **12**(4): p. 413-425.
77. Adelman, E.R., et al., *Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia.* Cancer Discovery, 2019: p. CD-18-1474.
78. Florian, M.C., et al., *Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation.* Cell Stem Cell, 2012. **10**(5): p. 520-30.
79. Ho, T.T., et al., *Autophagy maintains the metabolism and function of young and old stem cells.* Nature, 2017. **543**(7644): p. 205-210.
80. Flach, J., et al., *Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells.* Nature, 2014. **512**(7513): p. 198-202.
81. Mohrin, M., et al., *A mitochondrial UPR-mediated metabolic checkpoint regulates hematopoietic stem cell aging.* Science, 2015. **347**(6228): p. 1374-1377.

82. Wahlestedt, M., et al., *Clonal reversal of ageing-associated stem cell lineage bias via a pluripotent intermediate.* Nature Communications, 2017. **8**(1): p. 14533.
83. Siegel, R.L., et al., *Cancer Statistics, 2021.* CA: A Cancer Journal for Clinicians, 2021. **71**(1): p. 7-33.
84. Houshmand, M., et al., *Chronic myeloid leukemia stem cells.* Leukemia, 2019. **33**(7): p. 1543-1556.
85. Greaves, M., *A causal mechanism for childhood acute lymphoblastic leukaemia.* Nature Reviews Cancer, 2018. **18**(8): p. 471-484.
86. Rohrbacher, M. and J. Hasford, *Epidemiology of chronic myeloid leukaemia (CML).* Best Practice & Research Clinical Haematology, 2009. **22**(3): p. 295-302.
87. Gale, K.B., et al., *Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots.* Proc Natl Acad Sci U S A, 1997. **94**(25): p. 13950-4.
88. Slany, R.K., *The molecular biology of mixed lineage leukemia.* Haematologica, 2009. **94**(7): p. 984-993.
89. Bonnet, D. and J.E. Dick, *Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell.* Nat Med, 1997. **3**(7): p. 730-7.
90. Eisterer, W., et al., *Different subsets of primary chronic myeloid leukemia stem cells engraft immunodeficient mice and produce a model of the human disease.* Leukemia, 2005. **19**(3): p. 435-441.
91. Biernaux, C., et al., *Detection of major bcr-abl gene expression at a very low level in blood cells of some healthy individuals.* Blood, 1995. **86**(8): p. 3118-3122.
92. Bose, S., et al., *The Presence of Typical and Atypical BCR-ABL Fusion Genes in Leukocytes of Normal Individuals: Biologic Significance and Implications for the Assessment of Minimal Residual Disease.* Blood, 1998. **92**(9): p. 3362-3367.
93. Ugale, A., et al., *Hematopoietic Stem Cells Are Intrinsically Protected against MLL-ENL-Mediated Transformation.* Cell Reports, 2014. **9**(4): p. 1246-1255.
94. Nowell, P.C. and D.A. Hungerford, *Chromosome studies in human leukemia. II. Chronic granulocytic leukemia.* J Natl Cancer Inst, 1961. **27**: p. 1013-35.
95. Rowley, J.D., *Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining.* Nature, 1973. **243**(5405): p. 290-3.
96. Soverini, S., et al., *Targeting Leukemic Stem Cells in Chronic Myeloid Leukemia: Is It Worth the Effort?* Int J Mol Sci, 2021. **22**(13).
97. Mughal, T.I. and A. Schrieber, *Principal long-term adverse effects of imatinib in patients with chronic myeloid leukemia in chronic phase.* Biologics : targets & therapy, 2010. **4**: p. 315-323.

98. Ross, D.M. and T.P. Hughes, *Treatment-free remission in patients with chronic myeloid leukaemia.* Nature Reviews Clinical Oncology, 2020. **17**(8): p. 493-503.
99. Saussele, S., et al., *Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial.* The Lancet Oncology, 2018. **19**(6): p. 747-757.
100. Järås, M., et al., *Isolation and killing of candidate chronic myeloid leukemia stem cells by antibody targeting of IL-1 receptor accessory protein.* Proceedings of the National Academy of Sciences, 2010. **107**(37): p. 16280-16285.
101. Kobayashi, C.I., et al., *The IL-2/CD25 axis maintains distinct subsets of chronic myeloid leukemia-initiating cells.* Blood, 2014. **123**(16): p. 2540-2549.
102. Bocchia, M., et al., *Residual Peripheral Blood CD26(+) Leukemic Stem Cells in Chronic Myeloid Leukemia Patients During TKI Therapy and During Treatment-Free Remission.* Front Oncol, 2018. **8**: p. 194.
103. Herrmann, H., et al., *CD34(+)/CD38(-) stem cells in chronic myeloid leukemia express Siglec-3 (CD33) and are responsive to the CD33-targeting drug gemtuzumab/ozogamicin.* Haematologica, 2012. **97**(2): p. 219-26.
104. Niklas, L., et al., *CD36 defines primitive chronic myeloid leukemia cells less responsive to imatinib but vulnerable to antibody-based therapeutic targeting.* Haematologica, 2018. **103**(3): p. 447-455.
105. Herrmann, H., et al., *Dipeptidylpeptidase IV (CD26) defines leukemic stem cells (LSC) in chronic myeloid leukemia.* Blood, 2014. **123**(25): p. 3951-3962.
106. Warfvinge, R., et al., *Single-cell molecular analysis defines therapy response and immunophenotype of stem cell subpopulations in CML.* Blood, 2017. **129**(17): p. 2384-2394.
107. Nievergall, E., et al., *Monoclonal antibody targeting of IL-3 receptor α with CSL362 effectively depletes CML progenitor and stem cells.* Blood, 2014. **123**(8): p. 1218-1228.
108. Terwilliger, T. and M. Abdul-Hay, *Acute lymphoblastic leukemia: a comprehensive review and 2017 update.* Blood cancer journal, 2017. **7**(6): p. e577-e577.
109. Parkin, D.M., et al., *The international incidence of childhood cancer.* Int J Cancer, 1988. **42**(4): p. 511-20.
110. Wiemels, J.L., et al., *Prenatal origin of acute lymphoblastic leukaemia in children.* Lancet, 1999. **354**(9189): p. 1499-503.
111. Ramshaw, H.S., et al., *Engraftment of Bone Marrow Cells Into Normal Unprepared Hosts: Effects of 5-Fluorouracil and Cell Cycle Status.* Blood, 1995. **86**(3): p. 924-929.

112. Gothot, A., et al., *Functional Heterogeneity of Human CD34+ Cells Isolated in Subcompartments of the G0 /G1 Phase of the Cell Cycle.* Blood, 1997. **90**(11): p. 4384-4393.
113. Foudi, A., et al., *Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells.* Nature Biotechnology, 2009. **27**(1): p. 84-90.
114. Wilson, A., et al., *Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair.* Cell, 2008. **135**(6): p. 1118-1129.
115. Catlin, S.N., et al., *The replication rate of human hematopoietic stem cells in vivo.* Blood, 2011. **117**(17): p. 4460-4466.
116. Lee-Six, H., et al., *Population dynamics of normal human blood inferred from somatic mutations.* Nature, 2018. **561**(7724): p. 473-478.
117. Bernitz, J.M., et al., *Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions.* Cell, 2016. **167**(5): p. 1296-1309.e10.
118. Laurenti, E., et al., *CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells.* Cell Stem Cell, 2015. **16**(3): p. 302-313.
119. Cabezas-Wallscheid, N., et al., *Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy.* Cell, 2017. **169**(5): p. 807-823.e19.
120. Zou, P., et al., *p57Kip2 and p27Kip1 Cooperate to Maintain Hematopoietic Stem Cell Quiescence through Interactions with Hsc70.* Cell Stem Cell, 2011. **9**(3): p. 247-261.
121. Laurenti, E., et al., *Hematopoietic stem cell function and survival depend on c-Myc and N-Myc activity.* Cell Stem Cell, 2008. **3**(6): p. 611-24.
122. Takizawa, H., et al., *Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation.* J Exp Med, 2011. **208**(2): p. 273-84.
123. Yamamoto, R., et al., *Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment.* Cell Stem Cell, 2018. **22**(4): p. 600-607.e4.
124. Belluschi, S., et al., *Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors.* Nature Communications, 2018. **9**(1): p. 4100.
125. Muller-Sieburg, C.E., et al., *Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness.* Blood, 2004. **103**(11): p. 4111-4118.
126. Beerman, I., et al., *Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion.* Proceedings of the National Academy of Sciences, 2010. **107**(12): p. 5465.
127. Morita, Y., H. Ema, and H. Nakauchi, *Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment.* Journal of Experimental Medicine, 2010. **207**(6): p. 1173-1182.

128. Gekas, C. and T. Graf, *CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age.* Blood, 2013. **121**(22): p. 4463-4472.
129. Miyamoto, T., et al., *Myeloid or Lymphoid Promiscuity as a Critical Step in Hematopoietic Lineage Commitment.* Developmental Cell, 2002. **3**(1): p. 137-147.
130. Laurenti, E. and B. Göttgens, *From haematopoietic stem cells to complex differentiation landscapes.* Nature, 2018. **553**(7689): p. 418-426.
131. Trapnell, C., *Defining cell types and states with single-cell genomics.* Genome Res, 2015. **25**(10): p. 1491-8.
132. Mincarelli, L., et al., *Defining Cell Identity with Single-Cell Omics.* Proteomics, 2018. **18**(18): p. e1700312.
133. Minnoye, L., et al., *Chromatin accessibility profiling methods.* Nature Reviews Methods Primers, 2021. **1**(1): p. 10.
134. Klemm, S.L., Z. Shipony, and W.J. Greenleaf, *Chromatin accessibility and the regulatory epigenome.* Nature Reviews Genetics, 2019. **20**(4): p. 207-220.
135. Lowe, R., et al., *Transcriptomics technologies.* PLOS Computational Biology, 2017. **13**(5): p. e1005457.
136. Graves, P.R. and T.A.J. Haystead, *Molecular biologist's guide to proteomics.* Microbiology and molecular biology reviews : MMBR, 2002. **66**(1): p. 39-63.
137. Lambert, S.A., et al., *The Human Transcription Factors.* Cell, 2018. **172**(4): p. 650-665.
138. Bulger, M. and M. Groudine, *Enhancers: The abundance and function of regulatory sequences beyond promoters.* Developmental Biology, 2010. **339**(2): p. 250-257.
139. Heinz, S., et al., *Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities.* Molecular Cell, 2010. **38**(4): p. 576-589.
140. Zerbino, D.R., et al., *The ensembl regulatory build.* Genome Biol, 2015. **16**(1): p. 56.
141. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-61.
142. Visel, A., et al., *VISTA Enhancer Browser--a database of tissue-specific human enhancers.* Nucleic Acids Res, 2007. **35**(Database issue): p. D88-92.
143. *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
144. Fishilevich, S., et al., *GeneHancer: genome-wide integration of enhancers and target genes in GeneCards.* Database : the journal of biological databases and curation, 2017. **2017**: p. bax028.
145. Salzberg, S.L., *Open questions: How many genes do we have?* BMC Biology, 2018. **16**(1): p. 94.

146. Corces, M.R., et al., *Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution.* Nat Genet, 2016. **48**(10): p. 1193-203.
147. Arner, E., et al., *Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells.* Science, 2015. **347**(6225): p. 1010-4.
148. Gasperini, M., J.M. Tome, and J. Shendure, *Towards a comprehensive catalogue of validated and target-linked human enhancers.* Nature Reviews Genetics, 2020. **21**(5): p. 292-310.
149. Wang, E.T., et al., *Alternative isoform regulation in human tissue transcriptomes.* Nature, 2008. **456**(7221): p. 470-476.
150. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nature Genetics, 2008. **40**(12): p. 1413-1415.
151. Jacobsen, S.E.W. and C. Nerlov, *Haematopoiesis in the era of advanced single-cell technologies.* Nature Cell Biology, 2019. **21**(1): p. 2-8.
152. McKinnon, K.M., *Flow Cytometry: An Overview.* Current Protocols in Immunology, 2018. **120**(1): p. 5.1.1-5.1.11.
153. McCune, J.M., et al., *The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function.* Science, 1988. **241**(4873): p. 1632-9.
154. Lapidot, T., et al., *Cytokine Stimulation of Multilineage Hematopoiesis from Immature Human Cells Engrafted in SCID Mice.* Science, 1992. **255**(5048): p. 1137-1141.
155. Shultz, L.D., et al., *Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice.* The Journal of Immunology, 1995. **154**(1): p. 180.
156. Shultz, L.D., et al., *Human Lymphoid and Myeloid Cell Development in NOD/LtSz-scid IL2Rγ^{null} Mice Engrafted with Mobilized Human Hemopoietic Stem Cells.* The Journal of Immunology, 2005. **174**(10): p. 6477-6489.
157. Pike, B.L. and W.A. Robinson, *Human bone marrow colony growth in agar-gel.* Journal of cellular physiology, 1970. **76**(1): p. 77-84.
158. Sutherland, H.J., et al., *Differential Regulation of Primitive Human Hematopoietic Cells in Long-Term Cultures Maintained on Genetically Engineered Murine Stromal Cells.* Blood, 1991. **78**(3): p. 666-672.
159. Bourgine, P.E., et al., *In vitro biomimetic engineering of a human hematopoietic niche with functional properties.* Proceedings of the National Academy of Sciences, 2018. **115**(25): p. E5688.
160. Magnusson, M., et al., *Expansion on Stromal Cells Preserves the Undifferentiated State of Human Hematopoietic Stem Cells Despite Compromised Reconstitution Ability.* PLOS ONE, 2013. **8**(1): p. e53912.
161. Wills, Q.F., et al., *Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments.* Nat Biotechnol, 2013. **31**(8): p. 748-52.

162. Warren, L., et al., *Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR.* Proceedings of the National Academy of Sciences, 2006. **103**(47): p. 17807.
163. Teles, J., T. Enver, and C. Pina, *Single-cell PCR profiling of gene expression in hematopoiesis.* Methods Mol Biol, 2014. **1185**: p. 21-42.
164. Bengtsson, M., et al., *Quantification of mRNA in single cells and modelling of RT-qPCR induced noise.* BMC Molecular Biology, 2008. **9**(1): p. 63.
165. Macosko, Evan Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
166. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.
167. Cao, J., et al., *Comprehensive single-cell transcriptional profiling of a multicellular organism.* Science, 2017. **357**(6352): p. 661-667.
168. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell.* Nature Methods, 2009. **6**(5): p. 377-382.
169. Hicks, S.C., et al., *Missing data and technical variability in single-cell RNA-sequencing experiments.* Biostatistics, 2018. **19**(4): p. 562-578.
170. van Dijk, D., et al., *Recovering Gene Interactions from Single-Cell Data Using Data Diffusion.* Cell, 2018. **174**(3): p. 716-729.e27.
171. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nature Methods, 2013. **10**: p. 1096.
172. Jaitin Diego, A., et al., *Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types.* Science, 2014. **343**(6172): p. 776-779.
173. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.* Cell Rep, 2012. **2**(3): p. 666-73.
174. Ding, J., et al., *Systematic comparison of single-cell and single-nucleus RNA-sequencing methods.* Nature Biotechnology, 2020. **38**(6): p. 737-746.
175. Zilionis, R., et al., *Single-cell barcoding and sequencing using droplet microfluidics.* Nature Protocols, 2017. **12**(1): p. 44-73.
176. Rosenberg, A.B., et al., *Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.* Science (New York, N.Y.), 2018. **360**(6385): p. 176-182.
177. Mehrmohamadi, M., et al., *A Comparative Overview of Epigenomic Profiling Methods.* Frontiers in Cell and Developmental Biology, 2021. **9**(1990).
178. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nature Methods, 2013. **10**(12): p. 1213-1218.
179. Buenrostro, J.D., et al., *Single-cell chromatin accessibility reveals principles of regulatory variation.* Nature, 2015. **523**(7561): p. 486-490.
180. Cusanovich, D.A., et al., *Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing.* Science, 2015. **348**(6237): p. 910-914.

181. Lee, J., D.Y. Hyeon, and D. Hwang, *Single-cell multiomics: technologies and data analysis methods.* Experimental & Molecular Medicine, 2020. **52**(9): p. 1428-1442.
182. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature Methods, 2017. **14**(9): p. 865-868.
183. Shahi, P., et al., *Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding.* Scientific Reports, 2017. **7**(1): p. 44447.
184. Peterson, V.M., et al., *Multiplexed quantification of proteins and transcripts in single cells.* Nature Biotechnology, 2017. **35**(10): p. 936-939.
185. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587.e29.
186. Zappia, L. and F.J. Theis, *Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape.* bioRxiv, 2021: p. 2021.08.13.456196.
187. Zappia, L., B. Phipson, and A. Oshlack, *Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database.* PLOS Computational Biology, 2018. **14**(6): p. e1006245.
188. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902.e21.
189. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1): p. 15.
190. Luecken, M.D. and F.J. Theis, *Current best practices in single-cell RNA-seq analysis: a tutorial.* Molecular Systems Biology, 2019. **15**(6): p. e8746.
191. Andrews, T.S. and M. Hemberg, *M3Drop: dropout-based feature selection for scRNASeq.* Bioinformatics, 2019. **35**(16): p. 2865-2867.
192. Heimberg, G., et al., *Low Dimensionality in Gene Expression Data Enables the Accurate Extraction of Transcriptional Programs from Shallow Sequencing.* Cell Systems, 2016. **2**(4): p. 239-250.
193. Van der Maaten, L. and G. Hinton, *Visualizing data using t-SNE.* Journal of machine learning research, 2008. **9**(11).
194. McInnes, L., J. Healy, and J. Melville, *Umap: Uniform manifold approximation and projection for dimension reduction.* arXiv preprint arXiv:1802.03426, 2018.
195. Weinreb, C., S. Wolock, and A.M. Klein, *SPRING: a kinetic interface for visualizing high dimensional single-cell expression data.* Bioinformatics, 2018. **34**(7): p. 1246-1248.
196. Lotfollahi, M., et al., *Mapping single-cell data to reference atlases by transfer learning.* Nature Biotechnology, 2021.
197. Kiselev, V.Y., A. Yiu, and M. Hemberg, *scmap: projection of single-cell RNA-seq data across data sets.* Nature Methods, 2018. **15**(5): p. 359-362.
198. Chen, J., L. Rénia, and F. Ginhoux, *Constructing cell lineages from single-cell transcriptomes.* Molecular Aspects of Medicine, 2018. **59**: p. 95-113.
199. Trapnell, C., et al., *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.* Nat Biotechnol, 2014. **32**(4): p. 381-386.

200. Street, K., et al., *Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics.* BMC genomics, 2018. **19**(1): p. 477-477.
201. Weinreb, C., et al., *Lineage tracing on transcriptional landscapes links state to fate during differentiation.* Science, 2020. **367**(6479): p. eaaw3381.
202. Yan, F., et al., *From reads to insight: a hitchhiker's guide to ATAC-seq data analysis.* Genome Biology, 2020. **21**(1): p. 22.
203. Vierstra, J. and J.A. Stamatoyannopoulos, *Genomic footprinting.* Nature Methods, 2016. **13**(3): p. 213-221.
204. Chen, H., et al., *Assessment of computational methods for the analysis of single-cell ATAC-seq data.* Genome Biology, 2019. **20**(1): p. 241.
205. Stuart, T., et al., *Single-cell chromatin state analysis with Signac.* Nature Methods, 2021. **18**(11): p. 1333-1341.
206. Granja, J.M., et al., *ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis.* Nature Genetics, 2021. **53**(3): p. 403-411.
207. Russell, W.M.S. and R.L. Burch, *The principles of humane experimental technique*. 1959: Methuen.
208. Dhapola, P., et al., *Scarf: A toolkit for memory efficient analysis of large-scale single-cell genomics data.* bioRxiv, 2021: p. 2021.05.02.441899.
209. Jardine, L., et al., *Blood and immune development in human fetal bone marrow and Down syndrome.* Nature, 2021. **598**(7880): p. 327-331.
210. Hoebeke, I., et al., *T-, B- and NK-lymphoid, but not myeloid cells arise from human CD34(+)CD38(-)CD7(+) common lymphoid progenitors expressing lymphoid-specific genes.* Leukemia, 2007. **21**(2): p. 311-9.
211. Kondo, M., I.L. Weissman, and K. Akashi, *Identification of clonogenic common lymphoid progenitors in mouse bone marrow.* Cell, 1997. **91**(5): p. 661-72.
212. Gu, Z., et al., *PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia.* Nature genetics, 2019. **51**(2): p. 296-307.
213. Holyoake, T.L. and D. Vetrie, *The chronic myeloid leukemia stem cell: stemming the tide of persistence.* Blood, 2017. **129**(12): p. 1595-1606.
214. Lang, S., et al., *SCExV: a webtool for the analysis and visualisation of single cell qRT-PCR data.* BMC Bioinformatics, 2015. **16**(1): p. 320.
215. Roy, A., et al., *Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21.* Proceedings of the National Academy of Sciences, 2012. **109**(43): p. 17579.
216. Bowling, S., et al., *An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells.* Cell, 2020. **181**(7): p. 1693-1694.
217. Pei, W., et al., *Resolving fates and single-cell transcriptomes of hematopoietic stem cell clones by PolyloxExpress barcoding.* Cell Stem Cell, 2020. **27**(3): p. 383-395. e8.
218. Kester, L. and A. van Oudenaarden, *Single-Cell Transcriptomics Meets Lineage Tracing.* Cell Stem Cell, 2018. **23**(2): p. 166-179.

219. Abascal, F., et al., *Somatic mutation landscapes at single-molecule resolution.* Nature, 2021. **593**(7859): p. 405-410.
220. Park, S., et al., *Clonal dynamics in early human embryogenesis inferred from somatic mutation.* Nature, 2021. **597**(7876): p. 393-397.
221. Spencer Chapman, M., et al., *Lineage tracing of human development through somatic mutations.* Nature, 2021. **595**(7865): p. 85-90.

Printed by Media-Tryck, Lund 2021 NORDIC SWAN ECOLABEL 3041 0903

FACULTY OF MEDICINE

Division of Molecular Hematology
Department of Laboratory Medicin

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2021:151
ISBN 978-91-8021-158-1
ISSN 1652-8220