Applying single cell-based technologies to the characterisation of human haematopoietic stem cell heterogeneity



Merve Aksöz

Linacre College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Medical Sciences

Hilary Term 2021

Abstract

Haematopoiesis is essential for all vertebrate life. It involves the generation of terminally differentiated blood cells from a small pool of self-renewing multipotent haematopoietic stem cells (HSCs) through a series of lineage-restricted progenitors. HSCs form the basis of bone marrow (BM) transplantations and enable the treatment of patients with genetic blood disorders and leukaemia. It has long been assumed that individual HSCs have the capacity to produce all blood lineages in equal proportions. However, recently, single cell and serial transplantation studies in mice demonstrated a profound degree of heterogeneity within murine HSCs with a substantial proportion displaying a restricted pattern of lineage output. Consequently, these studies in mice have challenged the traditional view of a functionally homogeneous HSC population and instead suggested that individual HSCs differ in their properties. In contrast, human HSC heterogeneity has not yet been fully elucidated. The aim of this work was to characterise heterogeneity in human adult BM-derived HSCs, particularly focusing on the identification of human HSCs with distinct lineage outputs and their transcriptional differences. For this purpose, human HSCs were purified from adult BM, barcoded with unique DNA sequences encoded by lentiviruses, and transplanted into non-irradiated, immune-deficient NBSGW mice. We performed single-cell RNA sequencing of the engrafted barcoded HSCs and bulk RNA sequencing of their cellular *in vivo* output. We identified several human HSC subtypes with biased and restricted lineage outputs, as well as associated transcriptional differences, revealing correlated functional and molecular heterogeneity of human adult BM HSCs. This work may lead to the identification of potential molecular and/or surface markers that can help to determine HSC subsets with specific haematopoietic fate restrictions. Ultimately, characterisation of the human HSC heterogeneity in terms of the molecular signatures of specific HSC subtypes with distinct lineage outputs may allow us to predict how transplanted HSCs behave in patients.

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor Prof Claus Nerlov for welcoming me in his lab and giving me the opportunity to study DPhil at the University of Oxford. Understanding HSC biology has always been a passion of mine and I am happy to have been able to work on a project that I enjoy. I am grateful for his trust and the opportunity to work on such a big and impactful project. Thank you for the supervision and for guiding me on the way with your endless scientific knowledge. Finally, I would like to thank Claus for kindly arranging funding to support me during the last year of my DPhil.

I also would like to extend my gratitude to the past and present Nerlov lab members, Yiran Meng, Simona Valletta, Mario Buono, Roy Drissen, Xiying Ren, Hongyan Tao, and Alexander Thomas for their friendship, guidance, and help. Especially Simona Valletta has been a good friend to me.

Second, I would like to thank my co-supervisor Prof Paresh Vyas for his scientific inputs on this project and his encouragement on the way. I also would like to express my gratitude to Zahra Aboukhalil and Dimitris Karamitros, who were former members of the Vyas lab, for being kind and helpful when I first started this project. I also want to thank Supat Thongjuea for his advice on computational analyses.

I am also grateful to Ryan Beveridge and Philip Hublitz from the WIMM Virus Screening and Genome Engineering facilities. Ryan has become a good friend to me, and I want to thank him for his help and guidance. I also would like to thank the Flow Cytometry facility for their assistance on my complicated cell sorts. I would like to acknowledge Sally-Ann Clark in particular, for her help, patience, and empathy. I also thank WIMM Single Cell and Sequencing facilities led by Neil Ashley and Timothy Rostron.

I also would like to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) and the British Council, who awarded me a PhD scholarship and supported me during the first three years of my doctoral studies.

I want to thank my friends in the WIMM, WIMM Graduate Association, and the Linacre College for making my time in Oxford unforgettable. I would like to thank Falk Schneider, Lucy Field, Janina Nahler, Caitlin O'Brien-Ball, Liliana Barbieri, Narain Karedla, Martyna Lukoseviciute, and Erdinç Sezgin in particular, for creating a positive atmosphere in the WIMM and being dearest friends to me. I would not have survived the long PhD marathon easily without your support and friendship. Especially, I would like to thank Falk Schneider for his support, help, encouragement, and sharing of the happiest and hardest moments together. I cannot thank you enough. Last but not least, I thank the good friends from Turkey for their friendship over the years and their encouragement on the way.

Finally, I would like to thank my parents, Sibel and Haluk Aksöz, and my brother Berke Aksöz for believing in me and their continuous support on this journey. I could not have achieved this without your encouragement.

Contents

Abstract ..... **II**

Acknowledgements ..... **III**

Contents ..... **IV**

List of figures ..... **VIII**

List of tables ..... **XI**

Commonly used abbreviations ..... **XII**

Chapter 1: Introduction **1**

1.1. Haematopoiesis 2

1.2. Haematopoietic stem cells 3

1.2.1. Clinical relevance of HSCs 4

1.2.2. Prospective isolation of HSCs 9

1.2.3. Human HSC xenotransplantation 12

1.2.3.1. Xenograft models to study human haematopoiesis 14

1.2.4. Viral transduction of HSCs 18

1.3. Mature human haematopoietic populations 21

1.4. Classical human haematopoiesis hierarchy 23

1.5. HSC heterogeneity 26

1.6. *In vivo* lineage tracing 34

1.6.1. Cellular barcoding 39

1.7. Single-cell transcriptomic approaches for studying haematopoiesis and advanced haematopoiesis models 42

1.8. Aims and objectives 49

Chapter 2: Materials and Methods 51

2.1. Animals 52

2.2. Construction and validation of barcoded lentiviral vector libraries 52

2.3. Processing of human bone marrow samples 54

2.4. CD34 enrichment 54

2.5. Fluorescence-activated cell sorting staining and HSC purification 55

2.6. Transduction of HSCs with a lentiviral barcoding library 57

2.7. Intraosseous transplantation assay 58

2.8. Xenograft peripheral blood analysis 59

2.9. Xenograft bone marrow analysis 60

2.10. Single-cell RNA sequencing of HSCs 62

2.10.1. Sample collection and lysis 63

2.10.2. First-strand synthesis 63

2.10.3. Pre-amplification 64

2.10.4. Magnetic purification and quality control of cDNA libraries 65

2.10.5. Tagmentation and library preparation 66

2.11. Bulk RNA-seq of lineage populations 68

2.12. Bioinformatic analysis 70

2.12.1. Pre-processing of RNA-seq data and quality control assessment 70

2.12.2. Barcode detection and analysis 70

2.12.3. Hierarchical clustering 71

2.12.4. Processing single-cell RNA-seq data from HSCs 72

2.12.5. Gene set enrichment analysis 75

Chapter 3: Optimisation of methodologies for cellular barcoding of human adult BM-derived HSCs 76

3.1. Introduction 77

3.2. Results 79

3.2.1. Optimisation of lentiviral transduction conditions and the engraftment of human adult BM-derived HSCs in murine BM 79

3.2.2. Optimisation of sequencing library preparations 94

3.3. Discussion 98

Chapter 4: Identification of human HSC fates and the underlying transcriptional differences 109

4.1. Introduction 110

4.2. Results 112

4.2.1. Identification of human HSC subtypes by their fates 113

4.2.2. Identification of transcriptional differences between human HSC subtypes 119

4.3. Discussion 130

Chapter 5: Conclusions 139

Appendix 143

Bibliography 146

List of figures

Figure 1.2.1. Clinical use of HSCs.....8

Figure 1.4.1. Classical view of human haematopoiesis.....25

Figure 1.5.1. Heterogeneity in HSC properties.....27

Figure 1.5.2. Fates of the five classes of murine HSC clones based on their lineage contribution. 31

Figure 1.6.1. Cellular barcoding of HSCs 40

Figure 1.7.1. Advanced human haematopoietic hierarchy 47

Figure 2.2.1.The map of peGZ2-linkerB322 vector. 53

Figure 2.12.1. Quality control metrics of scRNA-seq data.....73

Figure 2.12.2. Filtering cells on number of expressed genes.....73

Figure 2.12.3. Linear dimensionality reduction. 74

Figure 2.12.4. A summary of performed bioinformatic analysis in this study 75

Figure 3.2.1. Representative FACS plots used to define and purify human HSCs. 81

Figure 3.2.2. Summary of lentiviral barcoding of primary human HSCs.....82

Figure 3.2.3. Human cell identification in murine BM in NBSGW recipients. 83

Figure 3.2.4. Engraftment of human cells in peripheral blood and bone marrow of a murine xenograft 16 weeks post-transplantation. 84

Figure 3.2.5. Human cell engraftment in the BM of NBSGW murine recipients detected by the expression of hCD45 marker by FACS. 87

Figure 3.2.6. EGFP percentages in the human cell compartment in the BM of NBSGW murine recipients detected by FACS. 88
Figure 3.2.7. FACS gating strategy used to identify and sort barcoded ($EGFP^+$) human HSCs in the murine BM. 91
Figure 3.2.8. FACS gating strategy used for the identification and purification of barcoded ($EGFP^+$) human B-Ly, My, Ery, and Mk populations. 93
Figure 3.2.9. Established experimental workflow for final experiments using optimised conditions and timings. 93
Figure 3.2.10. Bioanalyzer electropherograms of cDNA libraries: the optimisation of RNase inhibitor concentration in lysis buffer. 97
Figure 3.2.11. Bioanalyzer electropherograms of cDNA libraries: using different diameters of FACS nozzles. 98
Figure 4.2.1. The cellular *in vivo* output of HSC clones with unique barcode IDs. 116
Figure 4.2.2. Hierarchical clustering of barcoded HSC clones based on their lineage output. 117
Figure 4.2.3. Fates of the 6 classes of human HSCs (as identified by hierarchical clustering) schematically illustrated. 118
Figure 4.2.4. No significant difference between the clone sizes of the HSC subtypes. 119
Figure 4.2.5. Dimensionality reduction visualisation of HSC subtypes on UMAP identified using the hierarchical clustering of HSC clones 120
Figure 4.2.6. Differential gene expression analysis of functional human HSC subtypes revealed transcriptional heterogeneity. 121

Figure 4.2.7. The Mk-restricted HSC subtype was upregulated for multiple HSC gene-sets some of which may predict their Mk output and less proliferative state. 124
Figure 4.2.8. HSC maintenance related genes were significantly upregulated in the Mk-restricted subtype, as revealed by DEG and leading-edge analysis after the GSEA. 126
Figure 4.2.9. Biased/restricted subtypes (Mk-restricted, Ery-restricted, EryMy-biased, and MkEry-biased) might be more self-renewing and primed for Mk and Ery lineage differentiation as compared to the Multilineage subtype. 129
Figure 4.2.10. Representative GSEA plots of gene ontology terms from comparison of biased/restricted subtypes to the Multilineage subtype. 130

List of tables

Table 2.5.1 List of anti-human antibodies used for FACS staining 56

Table 2.10.1 List of primers used in single-cell and bulk RNA sequencing 63

Table 2.10.2 First-strand synthesis reaction mix 64

Table 2.10.3 First-strand synthesis reaction conditions 64

Table 2.10.4 PCR reagents 65

Table 2.10.5 Pre-amplification reaction conditions 65

Table 2.10.6 Tagmentation reaction mix 67

Table 2.10.7 Index PCR reagent mix 67

Table 2.10.8 Index PCR program 67

Table 2.11.1 Reverse transcription mix 69

Table 2.11.2 Pre-amplification mix 69

Table 3.2.1. hCD45% in murine peripheral blood for Exp 1-5 analysed at different time points post-transplantations. 84

Table 3.2.2. hCD45% in murine bone marrow for Exp 1-10 analysed at different time points post-transplantations. 89

Table 4.2.1. The number of human HSCs and HSC clones (per donor) detected in murine xenografts 16 weeks post- transplantation 115

Commonly used abbreviations

B-Ly	B lymphoid
BM	Bone marrow
BSA	Bovine serum albumin
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CRISPR	Clustered regularly interspaced short palindromic repeats
DEG	Differentially expressed gene
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetraacetic acid
EGFP	Enhanced green fluorescence protein
Ery	Erythroid
FACS	Fluorescence-activated cell sorting
FBS	Foetal bovine serum
Flt3-L	Fms-like tyrosine kinase 3-ligand
FMO	Fluorescence minus one
G-CSF	Granulocyte colony-stimulating factor
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GFP	Green fluorescence protein
GM-CSF	Granulocyte-macrophage colony-stimulating factor

GMP	Granulocyte-macrophage progenitor
GvHD	Graft-versus-host disease
GO	Gene ontology
HIV-1	Human immunodeficiency virus-1
Hrs	Hours
HLA	Histocompatibility antigen
HSC	Haematopoietic stem cell
HSPC	Haematopoietic stem and progenitor cell
IL	Interleukin
IMDM	Iscove's modified Dulbecco's medium
LIN-	Lineage negative
LMPP	Lymphoid-primed multipotential progenitors
LogFC	Log fold change
LT-HSCs	Long-term HSCs
LV	Lentivirus
Ly	Lymphoid
MEP	Megakaryocyte-erythroid progenitor
Mk	Megakaryocyte
MNC	Mononuclear cell
MOI	Multiplicity of infection
Mo-MLV	Moloney murine leukaemia
MPP	Multipotent progenitor
mRNA	Messenger ribonucleic acid
My	Myeloid
PB	Peripheral blood

PBE	PBS/BSA/EDTA
PBS	Phosphate buffered saline
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
RBC	Red blood cell
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Reverse transcription
RV	Retrovirus
SCF	Stem cell factor
scRNA-seq	Single-cell RNA sequencing
ST-HSCs	Short-term HSCs
T-Ly	T lymphoid
TPO	Thrombopoietin
UCB	Umbilical cord blood
vWF	von Willebrand factor
WBC	White blood cells
7AAD	7-amino-actinomycin D

Chapter 1: Introduction

1.1. Haematopoiesis

The haematopoietic system is one of the best characterised cellular differentiation hierarchies in mammals and responsible for lifelong maintenance of blood and immune cell production. In adults, approximately one trillion (10^{12}) mature haematopoietic cells are produced daily in bone marrow (BM), and each cell type has distinct physiological functions including oxygen delivery, thrombopoiesis, and innate/adaptive immunity [1, 2]. This system is tightly regulated by the activity of a population of undifferentiated cells referred to as haematopoietic stem cells (HSCs) which mainly reside in the BM. HSCs sit at the apex of the developmental hierarchy. They have the ability to self-renew to generate daughter cells while also contributing to the production of specialised blood and immune cells through a series of intermediate progenitors [1]. Haematopoietic progenitors are the offspring of HSCs which proliferate extensively and committed to differentiating into mature cells. These, then, proceed to the generation of terminally differentiated blood cells which comprise red blood cells (erythroid), short-lived myeloid lineages (granulocytes, macrophages), megakaryocytes/platelets, and long-lived lymphoid cells of the adaptive (B and T cells) and innate (natural killer cells) immune system. Therefore, the function of the haematopoietic system is a direct reflection of the HSC pool function and represent an example of how stem cells sustain and regenerate complex, self-renewing tissues during the lifespan of an organism. Disruption in the homeostasis of this system may lead to a range of haematological diseases such as leukaemia and lymphoma, anaemia, and immunodeficiency. Thus, the study of haematopoiesis, to understand the HSC function in homeostasis and perturbations

that underlie haematological disorders, as well as to develop treatments, has been the focus of intense work for almost a century.

1.2. Haematopoietic stem cells

HSCs are defined by their inherent capacity to self-renew, which refers to their ability to give rise to daughter HSCs without differentiation, and their multipotency to differentiate into all blood cell lineages throughout adult life. They are the only long-term self-renewing haematopoietic cell type and can be classified as quiescent or homeostatic based on their cycling behaviour [3-5]. HSCs undergo self-renewal in two ways: symmetric and asymmetric cell divisions. During symmetric division, an HSC gives rise to two new stem cells, which leads to expansion of the HSC pool, whereas in the asymmetric division one stem cell and one differentiated cell are formed that results in maintenance of the HSC pool while increasing the number of committed/differentiated cells [6, 7]. HSCs can also produce two differentiated cells via symmetric cell division.

HSCs are a rare population of cells that localise in the endosteal regions of the BM microenvironment (i.e., hypoxic microenvironment) with unique vasculature termed the “niche” [8]. Their balance between quiescence and expansion is regulated by cell-intrinsic (e.g., DNA mutations, unequal segregation of cell determinants, and epigenetic alterations) and cell-extrinsic (i.e., imposed by the BM microenvironment: cell-cell contacts, signalling pathways, cytokines, and chemokines) factors [9-13]. Steady-state haematopoiesis is mostly maintained by the restricted progenitors generated by HSCs. HSCs divide infrequently, approximately once every 135 days and 280 days in mouse and human, respectively [14-16]. HSCs are primarily

sustained in the G_0 phase of the cell cycle throughout life and rely on glycolysis in their low oxygen microenvironments [17]. This protects them from potential mutagenic hazards of DNA replication and cell division [18, 19]. They undergo division for replenishing haematopoietic cells and regenerating the system mostly after haematopoietic stresses such as injury, blood loss, infection, inflammation, and transplantation [16, 20, 21]. Additionally, a small number of HSCs can be detected in the bloodstream too, and their release from the BM niche into circulation can be enhanced by stimulation with cytokines in a process called mobilisation [22, 23]. The complementary process of mobilisation is referred to as homing which indicates the HSCs returning to their specialised BM niches from blood [24, 25]. Homing is also a pivotal step towards successful engraftment after HSC transplantation that is facilitated by chemotactic factors released in the niche. HSCs introduced into the bloodstream can engraft in the host, repopulate the HSC pool, and produce mature progeny [26].

1.2.1. Clinical relevance of HSCs

The search for HSCs started when the lethal consequence of the Hiroshima and Nagasaki atomic bombings in 1945 was found to be due to the haematopoietic failure (i.e., compromised ability to regenerate the blood system). Subsequently, great efforts have been made to recognise the effects of radiation exposure by giving a lethal dose of X-rays to adult mice and exploring how to restore the BM function afterwards. It was found that the effects of radiation could be prevented by injecting BM or spleen cells from non-irradiated donors [27-29]. After the observation that the damaged BM could be repopulated by infusion of healthy BM cells, Donnall Thomas and colleagues successfully performed the first BM

transplantation into leukaemia patients in 1957 [30]. Since then, the haematology field has expanded a lot. In 1961, Till and McCulloch published their first ground-breaking discoveries and indicated that the marrow cells transplanted into irradiated mice generated colonies in mouse spleen, confirming the existence of multipotential HSCs. In 1962, Goodman and Hodgson reported the presence of haematopoietic stem and progenitor cells (HSPCs) in peripheral blood (PB) by transplanting mouse PB leukocytes into lethally irradiated murine hosts, which proliferated and gave rise to lymphoid and myeloid cells [31]. The discovery that the administration of granulocyte colony-stimulating factor (G-CSF) to donors increases the number of stem cells in the PB led to the routine use of mobilised PB-derived HSCs in clinics [32]. By the late 1980s, umbilical cord blood (UCB, i.e., placental blood) was recognised as an alternative HSC source following their transplantation into a Fanconi anaemia patient (rare inherited bone marrow failure syndrome) and haematopoietic reconstitution [33].

Nowadays, autologous (i.e., patient's own cells) or allogeneic (i.e., from a donor) HSC transplantations have become standard therapy for patients with leukaemia, lymphoma, autoimmune diseases, and genetic defects, such as sickle cell anaemia, thalassemia, and Wiskott–Aldrich syndrome [34-37]. HSCs can be harvested from either a healthy donor's BM (under anaesthesia from the posterior/anterior iliac crest or sternum) or PB by the administration of G-CSF or GM-CSF to promote the circulation of stem cells into the bloodstream [38]. Before transplantation, patients usually receive conditioning regimens (chemotherapy and/or radiation) to ablate abnormal cells (known as myeloablation) and to suppress immune reactions to permit the engraftment of the transplanted HSPCs (Figure 1.2.1A) [39]. Thus, a

patient's BM can be reconstituted with healthy donor HSCs which can generate functional cells that can replace dysfunctional ones after conditioning treatments and help to restore BM function. Mostly, $CD34^+$ enriched BM or PB mobilised HSPCs are used for both autologous and allogeneic transplantations for a rapid reconstitution after myeloablative therapy [35, 40, 41]. CD34 is a surface protein that is expressed on HSPCs and lost during differentiation [42-44]. Conveniently, $CD34^+$ cells can be enriched by CliniMACS (a method that relies on coating cells with magnetic beads and separates $CD34^+$ cells in the magnetic field) [40].

The engraftment occurs within 7-14 days or 14 to 28 days after autologous and allogeneic transplantations, respectively [38], and neutrophil and platelet recovery are the indicators of this [45]. However, allogeneic transplantation, in most cases, results in an immunological reaction related to histocompatibility antigen (HLA) mismatch caused by the donor T cells against the recipients, called graft-versus-host disease (GvHD) [38, 39, 46]. GvHD is mostly associated with transplantation-related morbidity and mortality. If a suitable HLA-matched adult donor is not available, alternatively, UCB can be used safely. UCB is an easily accessible source of HSPCs and has less-stringent HLA-matching requirements due to the immaturity of lymphocytes at birth [47-49]. However, the limited number of HSCs per UCB unit delays reconstitution of the transplanted cells in the recipient's BM, and this may result in GvHD and/or associated mortality caused by infection [47, 50]. For this reason, UCB transplantations are preferred for the treatment of children. Successful treatment requires a sufficiently high number of HSCs for engraftment to enable repopulation of blood lineages [38, 49, 51, 52]. To overcome cell number limitations, a few strategies have been tested such as co-transplantation of two UCB grafts from

partially HLA-matched donors to improve engraftment and *ex vivo* expansion prior to transplantation [53-57].

As mentioned, limited HLA-matched donor availability could restrict the use of allogeneic BM transplantations for the treatment of immunodeficiencies. HSCs have also long been preferred as a target for gene therapy due to their multipotency and self-renewing capacity. Therefore, genetic engineering of autologous HSCs is considered as an alternative to stem cell transplantation from allogeneic donors. Successful HSC gene therapy was demonstrated to treat inherited diseases such as Wiskott–Aldrich syndrome [58, 59], X-linked severe combined immunodeficiency (SCID-X1) [60, 61], adenosine deaminase defects (ADA-SCID) [62, 63], β-thalassemia [64, 65], and leukodystrophies [66, 67]. The gene therapy workflow involves harvesting patient HSCs, their pre-stimulation with early-acting cytokines, the introduction of a wild-type gene using viral vectors, and their re-transplantation into myeloablated patients (Figure 1.2.1B). The gene modification introduced into the genome of the transplanted HSCs is transmitted to daughter HSCs and downstream progenitor cells by cell divisions and differentiation. This results in the expansion of the gene-corrected cells at greater numbers. Therefore, HSC transplantations provide a great platform for delivering stem-cell-based therapies due to their inherited self-renewal capacity and differentiation potential. Although gene therapy seems to be simple in theory, it involves determining the gene/genes responsible of development of diseases, their cloning in a suitable vector, their optimal delivery and expression in the target cells. Major concerns of the procedure include insertional mutagenesis, off-target effects, viral tropism, immune response, and undesired clinical outcomes [68, 69]. Efficient gene transfer with minimal *in vitro*

manipulation without compromising homing/engraftment activity of HSCs is also crucial for the success of the therapy (viral transduction and *in vitro* culture conditions will be discussed in section 1.2.4). However, *in vitro* HSC manipulations for gene therapy have to be performed in specialised/accredited centres for the safety of the procedure [70]. All these complexities increase the cost of the procedure and subsequently restrict the patient's access to these treatments.

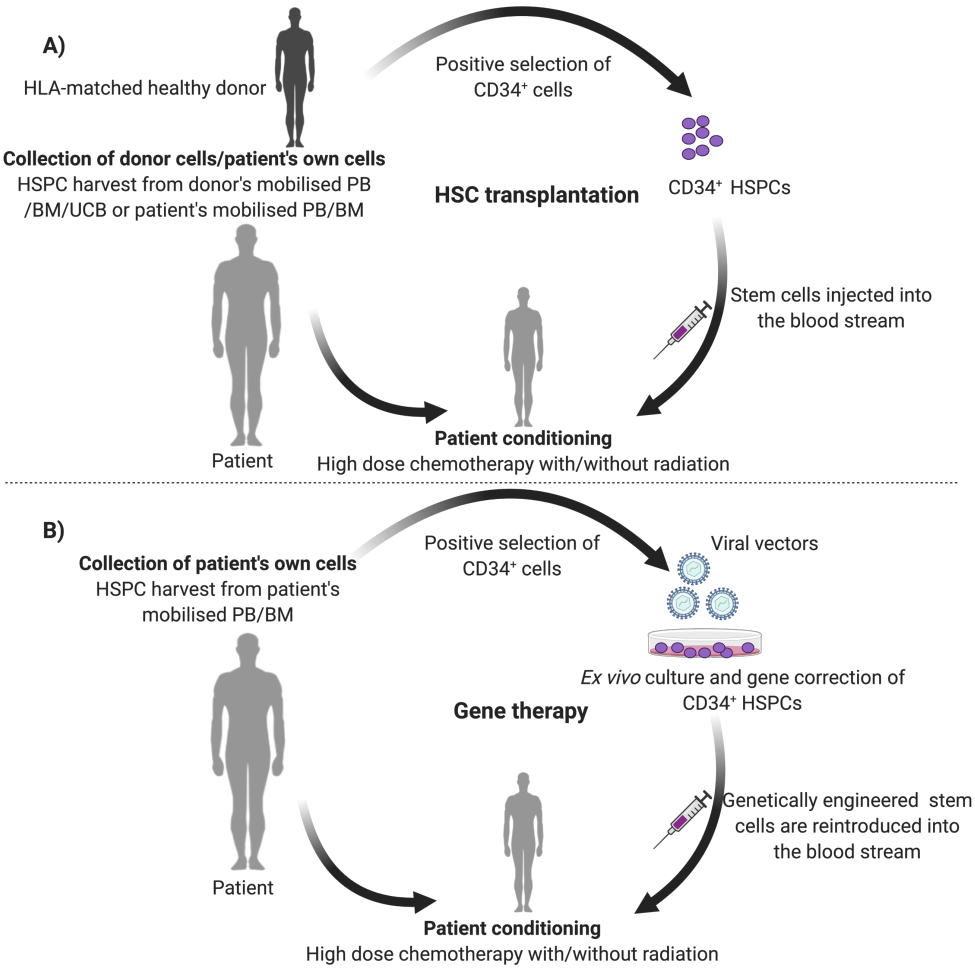


Figure 1.2.1. Clinical use of HSCs.
A) HSPCs from an HLA-matched healthy donor (allogeneic) can be purified and transplanted into the patient after myeloablation. In autologous transplantations, the patient's own cells are collected before the conditioning therapy and infused back to the patient after the treatment. **B)** HSC gene therapy applications are used to treat inherited diseases by introducing the wild-type gene into target cells. Patient HSPCs are harvested, manipulated with viral vectors in the presence of early acting cytokines, and transplanted back to the patient after the conditioning therapy. Created with BioRender.com.

The success of transplantation may vary according to the type and progress of the disease, depletion of patient's cells by radiation, source and dose of HSCs transplanted, and HLA compatibility [38, 39]. If the safety and efficacy of the gene therapy treatments can be demonstrated in a large number of patients with long-term positive outcomes, the transplantation of autologous gene-corrected HSCs/HSPCs provides an alternative to allogeneic transplantations which would be potentially available for many disease types [71]. Furthermore, a better understanding of the properties of a population of donor cells, for instance, the number of reconstituting human HSCs and their capacity to reconstitute the cellular lineages (short-term: neutrophils and platelets; long-term: lymphocytes) will help to reduce morbidity or mortality associated with HSC transplantations [34, 35, 38, 46]. This becomes also relevant for gene therapy applications where the characterisation of HSC population is pivotal for the success of disease treatment.

1.2.2. Prospective isolation of HSCs

The study of haematopoiesis has been advanced with the increased availability of monoclonal antibodies, the use of flow cytometry and fluorescence-activated cell sorting (FACS) technologies, that allowed the characterisation and isolation of HSC/HSPCs as well as differentiated blood cell populations using cell surface markers [72]. FACS is a cell purification technique that uses a laser- and fluidics-based technology to count and sort the cell population of interest, based on the immunophenotype identified by flow cytometry [73]. Several unique proteins are expressed on the plasma membrane that can distinguish one cell population from another in a heterogeneous cell suspension. These cell surface antigens can be targeted using specific fluorophore-conjugated monoclonal antibodies [73-75]. The

flow cytometry/FACS uses laser beams as light sources in combination with photo-multiplier tubes or photodiodes as detectors to measure light scatter and fluorescence emission characteristics of a mixed population of cells [76]. The scattered light is collected at two different angles: forward scatter (FSC) and side scatter (SSC) which conveys information about the size and granularity of the cells, respectively [73, 75]. FACS uses these signals in combination with the fluorescence and applies an electrostatic charge to the liquid stream allowing to sort the cells embedded in droplets into chosen tubes or plates through a nozzle for further analysis (the working principle of flow cytometry and FACS are described in detail elsewhere [75]). Thus, cell populations can be measured based on multiple parameters such as size, granularity, and fluorescence characteristics. FACS is also used to identify and purify cells expressing reporter fluorescent proteins integrated into their genome (e.g., green fluorescent protein, GFP). The number of possible combinations of fluorochromes that can be utilised in a single analysis depends on the type/number of lasers (i.e., having the corresponding wavelength), filters, and detectors that a flow cytometer/FACS instrument carry [77]. This allows researchers to carry out a more complex analysis.

Isolation of HSCs is a multi-step procedure that starts with mononuclear cell isolation and pre-sort enrichment of HSPCs. This increases the speed and efficiency of FACS purification of the HSC population. A few different combinations of markers were previously used to isolate mouse HSCs such as $Lin^{neg/low}$ $Sca1^{+}$ c-Kit^{high} (LSK) and Thy1.1^{low} [78, 79], or Thy1.1^{low} Lin^{neg} $Sca1^{+}$ [80] and Rhodamine123^{low} [81] or LSK $CD34^{low/neg}$ [82], or LSK $CD150^{+}CD48^{-}$ [83] or LSK $Flt3^{-}$ [84] (reviewed in detail in [3]). Mouse long-term HSCs (LT-HSCs) are most commonly defined as LSK

$CD34^{-}Flt3^{-}$. Although some of these cell surface markers are also used for human HSC identification, there are important differences in the expression patterns between these two species. For example, mouse HSCs are $CD34^{low}$ [82], $CD38^{+}$ [85], CD90 $(Thy1)^{low}$ [79], while human HSCs are $CD34^{+}$ [86], $CD38^{-}$ [87], and $CD90^{+}$ [88]. Additionally, the Fms-like tyrosine kinase 3 (Flt-3) receptor is not expressed on adult mouse LT-HSCs, conversely, it is detected on human UCB- and BM-derived HSCs [89, 90]. Furthermore, mouse HSCs are $c\text{-}kit^{+}$, however, human HSCs were reported as $c\text{-}kit^{low}$ [91, 92].

The characterised HSC surface markers in the human system are limited, and only a small number of markers are currently used to purify human HSCs. As in the mouse, the HSC population in human is defined by the absence of expression of lineage markers for differentiated cells; so, they are called lineage negative (Lin^{-}). CD34 has become the widely used cell surface marker for enriching the primitive human haematopoietic stem/progenitor cells as evidenced by clinical transplantations and *in vivo* repopulation assays in xenografts [93, 94]. It was first identified on the cell surface of a phenotypically less differentiated myeloblastic leukemic cell line, KG-1a [93]. CD34 is expressed on a diverse array of tissues in humans, which shows progenitor-like properties (e.g., HSCs and their progenitors, embryonic fibroblasts) and downregulated during differentiation, thus its expression is lost in terminally differentiated cells (i.e., mature blood cells) [42, 44]. Isolation of $CD34^{+}$ human HSPCs can also be achieved by magnetic-activated cell sorting (MACS) using magnetic microbeads coated with CD34 antibodies which is faster than FACS and reduces the time and cost of purifying HSCs [95]. HSCs are further purified by FACS based on the expression of CD38 [87], CD90 [88], and CD45RA

[96]; the combination used to define HSCs is commonly referred to as $Lin^-CD34^+CD38^-CD90^+CD45RA^-$. Recently, Notta *et al* proposed CD49f as an HSC specific marker that can be used in combination with common HSC markers to distinguish LT-HSCs from multipotent progenitors (MPPs) [97, 98].

Collectively, although many cell surface markers are used to identify mouse HSCs, those markers cannot be directly used for the immunophenotypic definition of human HSCs due to the species differences. Accurate immunophenotypic definition of human HSC and progenitor populations is crucial for the downstream functional and molecular analysis of lineage commitment. This is also translated to the design of experimental conditions for *ex vivo* expansion and transduction of human HSCs (i.e., cell surface receptor-cytokine interaction, see section 1.2.4) and therefore, it also has implications in gene therapy trials.

1.2.3. Human HSC xenotransplantation

A widely accepted way to show the “stemness” of a cell is to assess its capacity to differentiate into all blood cell lineages after transplantation into an irradiated recipient animal (multilineage and radioprotective property of HSCs) [99, 100]. The differentiation capacity of HSCs is analysed based on the quantification of donor-derived output in the periphery or BM of the host by flow cytometry. This is also coupled with serial transplantation assay to test long-term self-renewal and multilineage differentiation capacity: mouse/human HSCs that engraft and repopulate for 16 weeks and over in primary and secondary transplants are described as LT-HSCs [101-103].

Haematopoiesis has been well-studied in mice due to its homology with the human system and suitability for genetic modelling to mimic normal and malignant haematopoiesis [1]. Much of our understanding of haematopoiesis has been obtained in mice: indeed, from a technical point of view, it is much easier to retrieve HSCs from mouse BM, rather than from human, a process that requires donors to go under surgery to donate their BM. The extensive amount of work done using mouse BM samples led to the identification of a few combinations of cell surface markers to isolate mouse HSCs to near homogeneity [78-82]. Thus, functional outputs of mouse HSCs can be measured using *in vivo* repopulation assays. In addition, genetic manipulations in murine models have provided a powerful tool and yielded important insights towards understanding haematopoietic gene functions *in vivo* in health and disease. Although the findings from the murine system provide valuable information on HSC properties and overall haematopoietic differentiation, many critical differences exist due to species differences (e.g., immunophenotypic characterisation of HSPCs, cytokine and adhesion molecule pathways, haematopoietic gene regulation networks, adaptive and immune systems) [104]. The human haematopoiesis studies are mostly hampered by technical and ethical constraints (e.g., difficulties in *in vivo* gene manipulations in human) [104, 105]. Yet, with the advances in cell-sorting strategies and improved immunodeficient mouse models with a more humanized microenvironment (this will be explored in the next sub-chapter 1.2.3.1), significant progress has been made towards defining human blood hierarchy.

Most of the work on human HSCs relies on the xenotransplantation of UCB-derived HSCs into murine recipients [55, 98, 106-112]. Although UCB cells are an easily

accessible source of HSCs, there are functional differences between UCB and BM HSCs in terms of composition and properties: previous studies revealed that the ability of UCB-derived stem/progenitor cells to self-renew and/or generate colony-forming cells is greater than their BM counterparts [113-115]. UCB-derived HSPCs engraft and repopulate better upon transplantation into irradiated mice [112, 113, 116-120]. Therefore, most of the xenotransplantation studies are performed using UCB HSCs. However, as described before, its use in clinics is not widespread yet due to the limited number of CD34 expressing cells that can be retrieved from a single UCB unit; thereby, it is mostly preferred for the treatment of children [38, 121]. At the moment adult BM remains more clinically relevant and requires further studies to be characterised.

1.2.3.1. Xenograft models to study human haematopoiesis

The immunodeficient mice can be transplanted with human HSCs that can subsequently contribute to multilineage haematopoiesis in murine BM; this process is termed humanization [108]. The humanized mouse is used as a general term to refer to a mouse that does not have any immune system but is engrafted with human stem cells and thereby repopulated with human T cells, B cells, and other immune cells (constituted with human immunity) [122, 123].

The first successful human haematopoietic tissue transplantation was performed by McCune *et al* in the late 1980s using severe combined immunodeficient (SCID) mice lacking B and T cells due to a homozygous mutation in the *Prkdc*^{scid} gene which results in an inability to give an effective immune response to foreign antigens [124]. By transplanting human foetal tissues and foetal liver haematopoietic cells into SCID

mice, McCune *et al* demonstrated the human B and T-cell differentiation after engraftment [124]. Mosier *et al* transplanted human PB mononuclear cells into SCID mice and established engraftment of human T and B cells [125]. These studies showed that the human HSCs were able to engraft in murine BM and the usefulness of the immunodeficient xenograft models; however, those models lacked myeloid cell engraftment. To this end, Kamel-Reid and Dick transplanted human BM cells into sub-lethally irradiated SCID mice and infused human cytokines which then resulted in the formation of myeloid colonies and B-lymphoid lineage [126]. This study fulfilled the two other key criteria of HSCs: long-term engraftment and multilineage haematopoiesis [126]. To improve the human cell reconstitution, SCID mice were backcrossed onto the nonobese diabetic (NOD) strain, generating the NOD-SCID model. This model reported higher human cell engraftment levels compared to the previously reported models due to the human CD47 ligand expressed on haematopoietic cells, binding to the mouse NOD SIRPα (do-not-eat-me signal) allele on macrophages and preventing their phagocytosis [127]. However, NOD/SCID mice showed a notable risk of thymic lymphoma development in the long-term and NK cells in the BM resisted engraftment [128]. To combat this problem, the NSG mouse model was generated by introducing a deletion in the IL-2R common gamma chain (a key receptor component of IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 signalling) which resulted in a complete loss of murine B, T, and NK cells, allowing long-term engraftment of transplanted cells [1, 110, 129, 130]. NSG mice were reported to provide improved engraftment of $CD34^+$ HSPCs in their BM in comparison to the NOD-SCID model. In addition, it was observed that female NSG mice displayed up to 6-fold higher human HSC engraftment compared to males, meaning that sex-associated factors might play a role in human HSC regulation

[131]. Further modifications to the NSG line were made through the transgenic expression of human myeloid promoting cytokines such as stem cell factor (SCF), interleukin-3 (IL-3), and granulocyte-macrophage colony-stimulating factor (GM-CSF) [132-135]. Although the myeloid cell engraftment was significantly improved compared to the NSG model, nevertheless, human HSPCs lost their self-renewal/serial transfer ability after long exposure to cytokines [136]. In past years, MISTRG mouse model was generated from $Rag2^{-/-}Il2rg^{-/-}$ mice [137] which are genetically engineered to express human cytokine genes including human IL-3, Thrombopoietin (TPO), GM-CSF, SIRPα, and macrophage-stimulating factor (knocked into their respective mouse loci), to enable more efficient human cell reconstitution and myeloid and NK cell engraftments [138]. Recently, efforts have also been made to create humanized ossicle niches (i.e., niche-like structures, a possible new site for engraftment of human HSCs) in NSG mice to facilitate the engraftment of transplanted human cells with the support of secreted niche factors [139, 140]. This was achieved by subcutaneous transplantation of human BM-derived mesenchymal stem cells into NSG mice with or without carrier scaffolds to form humanized ossicle niches 8-10 weeks before human HSC transplantation [140-143]. Although this model was developed to mimic the human BM niches by providing multicellular interaction between mesenchymal stem cells and HSCs, it still lacks the complexity of the human BM environment [144], as well as it relies on prior irradiation before HSC transplantation.

To date, various mouse models have been used for human HSC transplantation and analysis (the most common ones are discussed above), but NSG-derived strains are the most popular due to their high engraftment potential. NSG and many

other strains rely on conditioning treatment (i.e., irradiation) prior to transplantation to deplete the endogenous HSCs and open up niche space for the donor cells. Thus, high levels of human chimerism can be achieved. However, irradiation prevents the study of human HSC function under steady-state conditions and creates a stressed environment for the donor cells (some of the human HSC transplantation studies which were analysed in the pre-irradiated murine models are cited here) [106, 145-147]. Knowing that haematopoiesis is regulated by the BM niche [20] and irradiation causes systemic side effects, engraftment without prior irradiation requirement would be advantageous [148]. Stable and long-term reconstitution of human HSCs in adult nonconditioned mice remained as a challenge until the development of mice with *Kit* receptor mutations [107, 108, 148, 149]. The Thomson and Waskow groups reported that NBSGW (NOD.Cg-Kit$^{W\text{-}41J}$Tyr^{+}Prkdc^{scid}Il2rg^{tm1Wjl}/ThomJ) [148] and NSGW41 (NOD.Cg-Kit$^{W\text{-}41J}$Prkdc^{scid}Il2rg^{tm1Wjl}/WaskJ) [107, 108] immunodeficient mice with loss-of-function *Kit* mutations were able to support stable long-term human stem/progenitor cell engraftment without irradiation. c-Kit (SCF receptor) is required to sustain endogenous murine haematopoiesis. Following its disruption, murine HSCs are impaired and become less competitive in occupying niches: hence, human HSCs can engraft better and repopulate in the murine BM [107, 108]. Therefore, previous irradiation conditioning is not required for this model. Moreover, these mice showed remarkably increased human myeloid cell engraftment and improved differentiation of erythroid and megakaryocyte progenitors compared to the irradiated NSG mice [107, 148]. Currently, they represent the best model for studying multilineage human HSC engraftment under more physiological conditions and have been used in this work.

1.2.4. Viral transduction of HSCs

Viral transduction techniques have been widely used in biological/biomedical research. They are utilised to induce/inhibit gene expression in a variety of cell types from immortalised cell lines to primary cells to understand their function in cellular mechanisms or for *in vivo* lineage tracing approaches. Retroviral vectors (RVs) have long been employed to genetically modify HSCs to express therapeutic genes for the treatment of various diseases [58, 60]. γ-RVs and lentiviruses (LV) are the members of the retroviridae (retroviruses) family, and they are the most frequently used vectors in gene therapy applications to introduce a wild-type gene into the genome to replace an abnormal gene responsible for the development of a certain disease [71, 150]. The majority of γ-RVs and LVs used for therapeutic purposes are based on Moloney murine leukaemia (Mo-MLV) virus and human immunodeficiency virus-1 (HIV-1), respectively. They carry an RNA genome that is converted to DNA in the host cell by the viral reverse transcriptase enzyme [151]. The genes that they carry can stably integrate into the genome of HSCs; this way, the transgene is constitutively expressed in HSCs as well as in their progeny (depending on the choice of promoter/enhancer). However, γ-RVs have the ability to transduce only proliferating cells which restrict their use for *in vivo* gene delivery in non-proliferating cells such as hepatocytes, myofibers, neurons, and HSCs. Therefore, long pre-stimulation, usually for several days, with early acting cytokines is required for the recruitment of cells into the cell cycle. LVs in contrast to γ-RVs can transduce both proliferating and non-dividing cells [152, 153]. In addition, LVs have been considered safer than γ-RVs; this is due to the differences in their integration site selection [154-156]. γ-RVs show a strong bias towards integrating close to the promoter regions (near proto-oncogenes) and CpG islands, where insertion of viral

long terminal repeat (LTR) transcriptional enhancer can activate neighbouring gene expression which increases the risk of insertional mutagenesis [154, 156, 157]. In one gene therapy trial of SCID-X1 disorder using γ-RVs, it was demonstrated that the activation of *LMO2* proto-oncogene was due to the vector integration at proximity to *LMO2* promoter [158]. This resulted in leukaemia development in those patients receiving therapy due to the clonal outgrowth of transfected cells. In contrast LVs predicted safer as compared to γ-RVs, specifically, the third-generation LVs is absent of essential genes required for the lentiviral packaging system, therefore they are replication-incompetent and self-inactivating viruses [154, 159]. They integrate within transcribed genes and that the use of newer generation of LVs strongly reduces the risk of insertional mutagenesis compared to the γ-RVs as promoter/enhancer activity of LV LTR is disrupted [160].

For the viral transduction, *in vitro* manipulation of the cells is an essential step; however, the most appropriate *in vitro* culture conditions and duration may vary according to the cell type, the vector of choice, and the gene of the transfer. Many research and clinical groups reported their optimal transduction conditions for efficient gene delivery into HSCs, whereas there is a great variation among the published protocols [111, 150, 161-163]. The key variables include a vector of choice, pre-stimulation and transduction durations, successive rounds of transductions, the use of transduction enhancers such as polybrene and RetroNectin, cell culture media, cytokine combinations and concentrations. The most commonly used early acting cytokines are SCF, Flt3-ligand (Flt3-L), TPO, IL-3, and IL-6 in both γ-RV and LV mediated transduction [55, 89, 150, 162]. The combination of these growth factors in cell culture media was reported to increase

transduction efficiencies [161, 162] by inducing proliferation of HSPCs and recruiting them into the cell cycle [164-166]. Some other data suggested that although the use of IL-3 elevated the transduction efficiency of progenitors, it also caused a reduction in $CD34^+$ cells [161, 162]. Similarly, the addition of foetal bovine serum (FBS) to the cell culture media was demonstrated to cause HSC differentiation [161, 162]. Furthermore, the use of animal origin serum may cause immunological reactions and create some safety concerns in clinical applications [161]. Overall, there is an overwhelming need for clarification, characterisation, and standardisation of transduction and *in vitro* culture conditions. To date, optimisation for every experimental set up is crucially required.

As mentioned above, γ-RVs have only the ability to infect proliferating cells; therefore, long *in vitro* culture with cytokines to induce cell cycle transit is needed for transduction of HSCs. However, this affects long-term HSC viability by inducing cell differentiation and reducing the reconstitution potential, which negatively impacts the long-term success in gene therapy applications [161, 167]. The utilisation of LVs potentially reduces *in vitro* cell culture time and preserves cellular functions. Despite the potential of LVs to infect quiescent HSCs without the need for proliferation, shortened *in vitro* cytokine stimulation may still enhance HSCs susceptibility to transduction [161]. $CD34^+$ cells in the G_1 or G_2/S/M phase of the cell cycle were transduced better compared to the G_0 $CD34^+$ cells [168]. However, the negative impact of *in vitro* cell culture should be minimised for faster and efficient polyclonal long-term human cell engraftment. Therefore, it is necessary to optimise lentiviral transduction protocols for efficient gene delivery to HSCs while preserving their biological properties at the same time. Yet, BM and mobilised PB HSCs are

more therapeutically relevant for human gene therapy trials, therefore the improved lentiviral transduction of adult BM HSCs can be translated into gene therapy protocols.

1.3. Mature human haematopoietic populations

The regenerative potential of HSCs relies on their ability to give rise to functional mature cell types. Mature haematopoietic populations are produced through a series of divisions from committed progenitors that are generated by HSCs, in a process known as differentiation. Mature cells have a short lifespan: therefore, HSCs are required to self-renew to sustain themselves and generate more of the committed multipotent, oligopotent, and lineage-restricted (unipotent) progenitors to sustain mature blood cells in the BM and blood [169, 170].

Each of those terminally differentiated populations has distinct physiological roles in the organism. The utilisation of multi-colour flow cytometry analysis improved the identification and isolation of different blood lineages. Human leukocytes, also known as white blood cells (WBC), are found in the BM, blood, and lymph tissue and they consist of myeloid and lymphoid cells as well as megakaryocytes. They can be identified with the expression of human CD45, pan-leukocyte marker [171, 172]. Red blood cells (RBCs), also referred to as erythroid (Ery) lineage, constitute the highest proportion of cell type in the blood. Unlike WBCs, the mature erythrocytes lack CD45 expression due to their anucleate nature [43].

Myeloid (My) cells are short-lived WBCs produced through a process called myelopoiesis comprising monocytes, neutrophils, basophils, eosinophils, and mast

cells that contribute to the innate immune system. One of the major functions of My lineage is phagocytosis that refers to their immediate defence against foreign materials or pathogens. They release inflammatory cytokines (e.g. TNFα, IL-1, IL-6, and IL-12) to initiate immune responses [173]. In general, My lineage can be identified by CD33 expression in the human system [174]. Neutrophils are the most abundant type of myeloid cells in the BM. The size of myeloid cells ranges from 15 to 20 μm [175, 176].

Lymphoid (Ly) cells are the long-lived cell population that forms the adaptive and the innate immune system which consists of B cells (B-Ly, $CD19^+$) [177, 178] and T cells (T-Ly, $CD3^+$) [179] as well as natural killer cells (NK, $CD56^+$) [180]. They act against pathogens (antigens) and initiate antibody responses. The cell sizes of B-Ly and T-Ly are estimated between 7 to 10 μm in diameter [175, 181].

Megakaryocytes (Mk) are quite rare and constitute ~0.01-0.05% of the total nucleated cells in BM [182]. They are generated through a process termed megakaryopoiesis. Mk population undergoes multiple rounds of DNA replications without cell division which is referred to as endomitosis [183]. Thus, their nuclei have multiple lobes (2-16) and they display the largest cell diameter (50 to 100 μm) [184]. Mk cells ($CD41^+$) [185] play a pivotal role in thrombopoiesis by releasing platelets into the blood circulation [184]. Platelets are small anucleate cells (1 to 4 μm in diameter) responsible for the blood clotting process. Platelets lack CD45 pan-haematopoietic marker expression as they do not have a nucleus.

RBCs deliver oxygen from the lungs throughout the body via blood vessels and make up the blood. They contain large amounts of haemoglobin protein, an iron-containing biomolecule with high affinity to molecular oxygen. Thus, RBCs can bind and carry oxygen to the tissues as well as are responsible for transporting carbon dioxide back to the lungs. Erythroid cells show a gradual loss of CD45 expression during maturation [43, 186]. While nucleated erythroid precursors and pro-erythroblasts express CD45 dimly, CD45 is not expressed on nucleated erythroblasts or anucleate mature erythrocytes. Erythroblasts are identified by the expression of CD235a and CD71 markers on their cell surface [186]. Their cell size is estimated between 10-15 μm [187]. CD71 expression is lost in mature anucleate erythrocytes and these can be identified as $CD235^+$ [43, 186, 188]. The size of human erythrocytes is similar to the lymphocytes: they are 6 to 8 μm in diameter [175].

1.4. Classical human haematopoiesis hierarchy

The classical tree-like haematopoiesis model proposes that haematopoiesis originates in multipotent HSCs and continues through discrete and progressively lineage-restricted progenitor (oligo-, bi-, and unipotent) populations. According to this model, multipotent HSCs ($Lin^-CD34^+CD38^-CD90^+CD45RA^-$) reside at the top of the hierarchy [3, 96, 169], with the initial bifurcation occurring downstream of the MPP compartment ($Lin^-CD34^+CD38^-CD90^-CD45RA^-$) [80, 96]. MPPs are multipotent haematopoietic cells but do not have the property of long-term self-renewal (i.e., have reduced and incomplete self-renewal activity) [80, 96]. MPPs separated oligopotent cell populations due to their lymphoid restricted (common lymphoid progenitor, CLP, $Lin^-CD34^+CD38^+CD127^+$) [169, 189] or myeloid restricted

(common myeloid progenitor, CMP, $Lin^-CD34^+CD38^+CD45RA^-CD123^+$) lineage potential (Figure 1.4.1) [190, 191]. The CLPs were thought to be the common ancestor of all the lymphoid cells (B, T, and NK cells). On the other hand, the CMPs were further branched into granulocyte-macrophage progenitor (GMP, $Lin^-CD34^+CD38^+CD45RA^+CD123^+$) and megakaryocyte-erythroid progenitor (MEP, $Lin^-CD34^+CD38^+CD45RA^-CD123^-$) populations [190, 191]. According to this model, GMPs were proposed to give rise to all myeloid cell types including granulocytes (eosinophil, basophil, neutrophil, and mast cell) and macrophages; MEPs were thought to generate erythrocytes and platelets (Figure 1.4.1) [169].

Using *in vivo* reconstitution and single-cell *in vitro* assays, this classical model was later revised with the identification of murine lymphoid-primed multipotent progenitor (LMPP) population that lacks Ery and Mk potential [192, 193]. Additional sub-populations with My (pre-granulocyte-macrophage progenitor, preGM) or Ery (colony-forming unit-erythroid, preCFU-E) potential were also suggested in the murine models [194]. The LMPP branching is now widely accepted since it was also identified in the human haematopoiesis (defined as $Lin^-CD34^+CD38^-CD90^-CD45RA^+$) [195].

The classical tree-like haematopoiesis model is built on the cell populations isolated by FACS as bulk, using limited predefined cell surface markers, followed by *in vitro* and *in vivo* clonal assays. Consequently, the classical model assumed that every cell within each isolated population had the same fate, underestimating functional heterogeneity within the identified populations. Additionally, it needs to be considered that the different cell types may be generated through more than one

differentiation route. Therefore, further clarification of the haematopoietic differentiation pathway and appreciation of functional heterogeneity within the classical populations has become essential. Recently, with the development and/or advancement of new assays and technologies, e.g., single-cell transplantations, lineage tracing techniques, and single-cell RNA sequencing, the branching points in the classical haematopoiesis model have been questioned. These technologies and relevant studies will be explained and reviewed throughout the next sections (see section 1.5, 1.6, and 1.7).

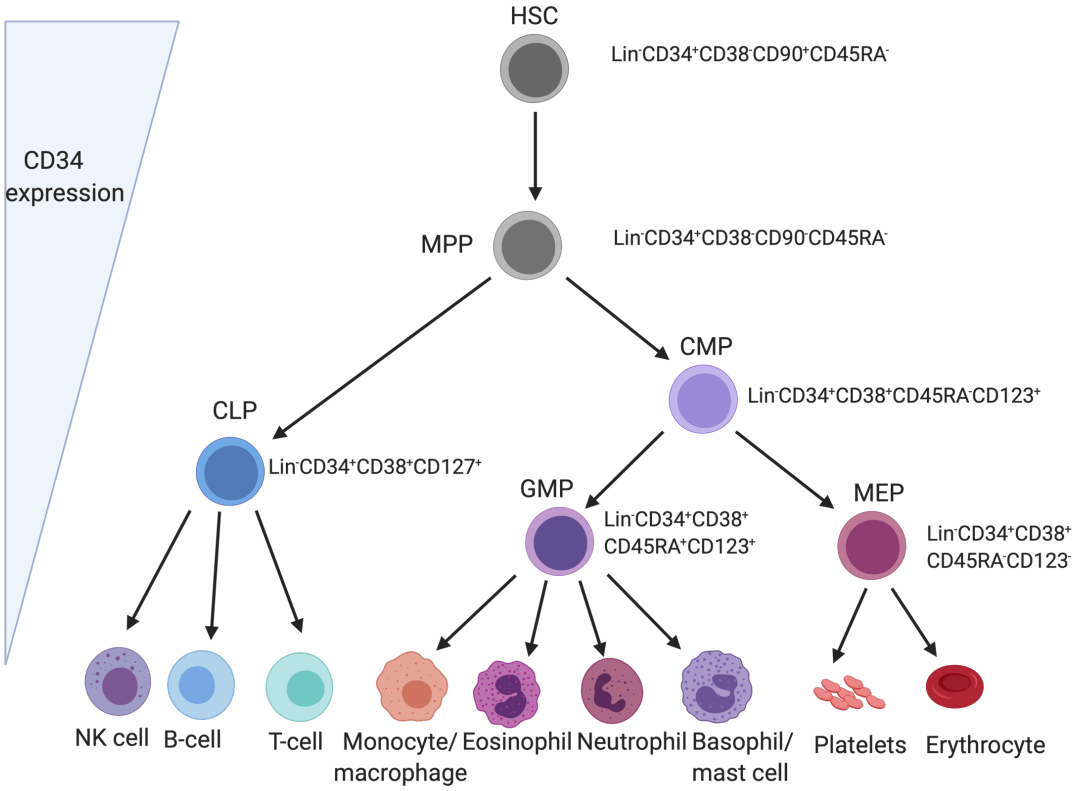


Figure 1.4.1. Classical view of human haematopoiesis
Stepwise (tree-like) differentiation haematopoiesis: CMP-CLP immunophenotype-based model. Multipotent HSCs can self-renew to give rise to daughter HSCs or produce multipotent progenitors (MPPs). The initial branching separates oligopotent cells committed to lymphoid (CLP) or myeloid/erythroid (CMP) fates. Lineage potential decreases progressively, resulting in the production of lineage restricted progenitors and mature blood cells. This model was largely established by Weissman's group [169]. Indicated surface markers are used to isolate stem and progenitor cells. CMP, common myeloid progenitor; CLP, common lymphoid progenitor; GMP, granulocyte-macrophage progenitor; MEP, megakaryocyte-erythroid progenitor; MkP, megakaryocyte progenitor; EryP, erythroid progenitor; NK, natural killer cells. Created with BioRender.com.

1.5. HSC heterogeneity

It has long been assumed that the HSC pool was homogeneous and consequently a pure population of HSCs could be isolated based on the expression of cell surface markers [170]. LT-HSCs have been typically quantified and characterised based on the measurements of their contribution to lymphoid and myeloid lineages in immunocompromised murine hosts for a period of 16-24 weeks and beyond [97, 196]. This approach was based on the assumption that the self-renewal and lineage differentiation capacity of all single HSCs are equal (e.g., an HSC was thought to differentiate into one lymphoid and one myeloid cell) and controlled by a common signalling mechanism [197]. However, HSCs (short- and long-term) differ in their self-renewal capacity thus this generates heterogeneity in the HSC compartment [14, 197, 198]. All daughter HSCs and mature cells produced by single parent HSC compose an HSC clone and the life span of each clone differs [197]. Recent findings in the murine model contradict the historical concept of a homogenous HSC pool where all HSCs produce all blood lineages equally. The growing body of evidence urges us to consider that not all HSCs are alike, but instead differ in their functional and molecular properties such as self-renewal capacity [198, 199], cell cycle kinetics [14, 16], reconstitution kinetics [199], repopulating activity [198, 200], clonal lifespan [197], and lineage contribution [200-205] (see Figure 1.5.1).

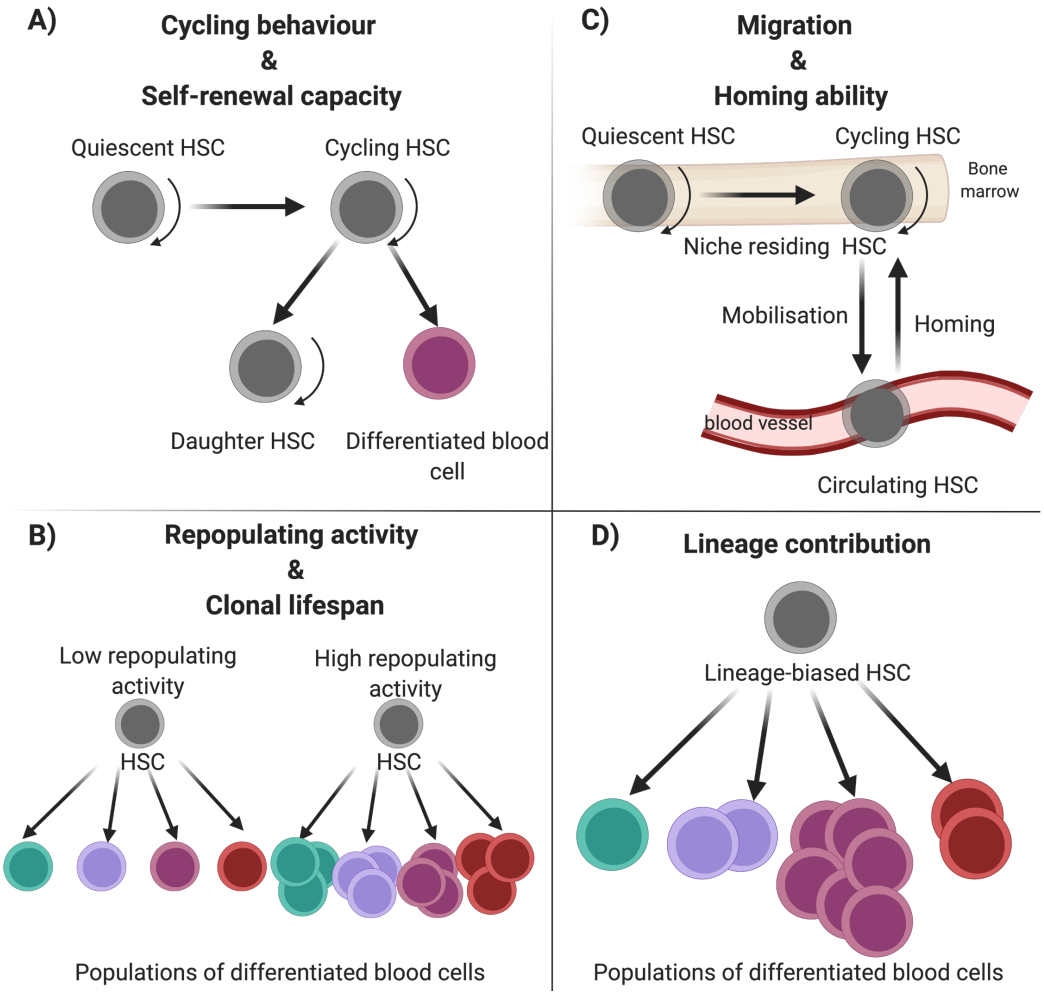


Figure 1.5.1. Heterogeneity in HSC properties
A) HSCs can be classified as quiescent or cycling based on their cell cycle status. They self-renew to give rise to daughter HSCs that again can self-renew and differentiate. Together these mechanisms can expand a single HSC into a large clone that can provide mature cells for the life span of the organisms. **B)** HSC clones differs in their life span due to the number of divisions that each HSC undergo. HSCs also varies in their repopulation ability which describes how much of mature cells can a stem cell produce when transplanted into a recipient animal. Different repopulation capacity of HSCs results in different sizes of clone that is generated. **C)** HSCs are localised in BM niches, but upon stimulation with cytokines, they can enter the bloodstream (mobilisation). The complementary process is called homing/engraftment that describes HSCs returning to the BM niches from the circulation. **D)** It was assumed that all HSCs can generate all blood lineages upon transplantation. However, the majority of HSCs display “biases” towards the generation of certain blood lineages: so-called lineage-biased HSCs. Created with BioRender.com.

Technological advances and refinements in flow cytometry and cell sorting over the last decade led to the purification of long-term murine HSCs to near homogeneity [82, 83]. This innovation also allowed single-cell transplantation studies. These studies have yielded highly valuable insights and unveiled, for the first time, HSC

heterogeneity by showing that one single cell can be isolated, transplanted and that its functional output, expansion capacity, and self-renewal duration can be assessed [198, 200-204, 206]. The initial identification of murine HSC heterogeneity was through tracking lineage output of single HSC-derived clones or single HSCs [200, 201]. These studies reported HSC clones with lineage-biases, based on the measurement of their predominant lineages within the total lineage output of the donor-derived HSCs. According to this, HSCs were classified into subtypes: myeloid-biased, lymphoid-biased, and balanced. The pattern of bias was confirmed to be stable by performing up to three serial transplantations, meaning that these biased subsets retained their self-renewal ability [200, 201]. Therefore, they identified the lineage bias as an intrinsic property of HSCs. Moreover, some additional studies reported the age-related changes in the HSC clonal composition: myeloid-biased HSCs become dominant in the aged HSC compartment and this change was consistent between each mouse, suggesting an intrinsically controlled mechanism [207, 208]. Furthermore, other studies demonstrated that HSCs can be enriched to different subsets (myeloid-biased or lymphoid-biased) with extensive self-renewal activity, based on the CD150 surface expression and Hoechst dye efflux, providing further evidence for the predetermined lineage bias [102, 197, 199, 209].

Earlier studies mentioned above demonstrated that the distinct murine HSC subsets exist with different lineage outputs (e.g., myeloid-biased, balanced, and lymphoid-biased) [200, 210]. These findings were based on the transplantation of single HSCs with a CD45.2 allotype into CD45.1 recipient mouse which enabled tracking of myeloid and lymphoid cells, but not of mature erythrocytes or platelets that devoid

of CD45 expression. Recently, the Nerlov and Jacobsen groups demonstrated the platelet-biased and platelet/myeloid biased HSCs in murine BM by tracking LT-HSCs defined as LSK $CD34^{-}CD150^{+}CD48^{-}$ [203]. Prior to that demonstration, a small group of murine LT-HSCs expressing megakaryocytic genes was identified: one of the most highly expressed megakaryocytic gene was the gene coding for the von Willebrand factor (vWF) [193]. vWF is a protein that mediates platelet aggregation [203, 211]. Therefore, to determine if there is a megakaryocyte/platelet-biased HSC subset, a transgenic reporter mouse model (vWF-EGFP) was developed [203]. They purified vWF-$EGFP^{+}$ LT-HSCs (single-cell, 5, 10, or 50 cells) and transplanted them into lethally irradiated adult recipients, allowing identification of vWF^{+} HSCs that exist with long-term stable (sustained in secondary recipients) platelet-biased output and vWF^{-} HSCs with a lymphoid-biased output [203]. vWF^{+} platelet-biased HSCs also contribute to myeloid lineages substantially as well as generate fewer lymphoid cells in both primary and secondary recipients [203]. While vWF^{+} HSCs reconstituted vWF^{+} and vWF^{-} HSCs in secondary recipients, vWF^{-} HSCs lacked that ability. Collectively, in this study, it was suggested that vWF^{-} HSCs were generated by vWF^{+} HSCs; therefore, platelet-biased HSCs were considered as a more primitive subset and to sit at the top of the haematopoietic hierarchical tree [203]. Moreover, an additional study from the Nerlov and Jacobsen groups using the same transgenic mouse model demonstrated that the aged HSCs which were previously suggested to be myeloid-biased [207, 208], demonstrated platelet-bias [212]. This was accompanied by significant enrichment of platelet-lineage-specific gene expression in HSCs at the single-cell level, suggesting functional and molecular platelet-bias of HSCs [212]. Age-dependent platelet-bias was due to a decrease in lymphoid cells and that the decreased lymphopoiesis directly correlated

with increased platelet priming of HSCs [212]. It was suggested that HSC platelet priming might be suppressing the lymphoid lineage commitment or differentiation epigenetically, as lymphoid gene expression is barely detectable in HSCs [212]. Overall, these studies proposed that the platelet-biased and platelet-myeloid biased populations might have identified as a myeloid-biased population in previous studies since platelet reconstitution could not be evaluated with the mouse models used before [199-201, 207, 208].

Another study used transgenic reporter mice (Kusabira Orange mouse line which expresses orange and red fluorescent protein) to track lineage outputs (lymphoid, myeloid, megakaryocyte/platelet, and erythroid) of transplanted single murine HSCs ($CD150^{+}CD34^{-}LSK$) [204]. They found that the murine HSC subsets committed to generating megakaryocyte, megakaryocyte-erythroid, or megakaryocyte-erythroid-myeloid lineages [204]. Also, by coupling paired daughter cell assays, they proposed that the myeloid branching (branching of megakaryocyte-erythroid-myeloid lineages) may occur directly from the HSCs and these cells were self-renewing oligopotent progenitors rather than multipotent [204]. Similarly, Haas *et al* also observed stem-like megakaryocyte progenitors (referred to as SL-MkPs) within the HSC compartment [213]. SL-MkPs had many shared features with multipotent HSCs i.e., they were small in size and maintained in a primed but quiescent state during homeostasis, yet they were a lineage-restricted population [213]. SL-MkPs became activated upon inflammatory stress, suggesting a potential generation of megakaryocytes directly from HSCs as a response to inflammatory stress to replenish platelets rapidly [213]. Yet, more recent single-cell transplantation studies reported distinct types of self-renewing murine multipotent HSCs that adopted

restricted fates towards the generation of platelet, platelet-erythroid, platelet-erythroid-myeloid, or platelet-erythroid-myeloid-B cell lineages [202] (Figure 1.5.2). This was demonstrated by tracking murine LT-HSCs (LSK $CD34^{-}CD150^{+}CD48^{-}$) purified from Vwf-tdTomato/Gata1-EGFP double reporter mice (*Gata-1* has an important role in megakaryocyte and erythrocyte development) in lethally irradiated primary and secondary murine recipients and performing *in vitro* cell culture assays [202]. Most notably, this study showed that a distinct group of multipotent HSCs (referred to as platelet-restricted) adopted a cell fate prior to differentiation and had the ability to give rise to exclusively megakaryocytes/platelets, without contributing to any other blood lineages [202]. Together, all this evidence suggested that megakaryocytes could be produced independently of other lineages.

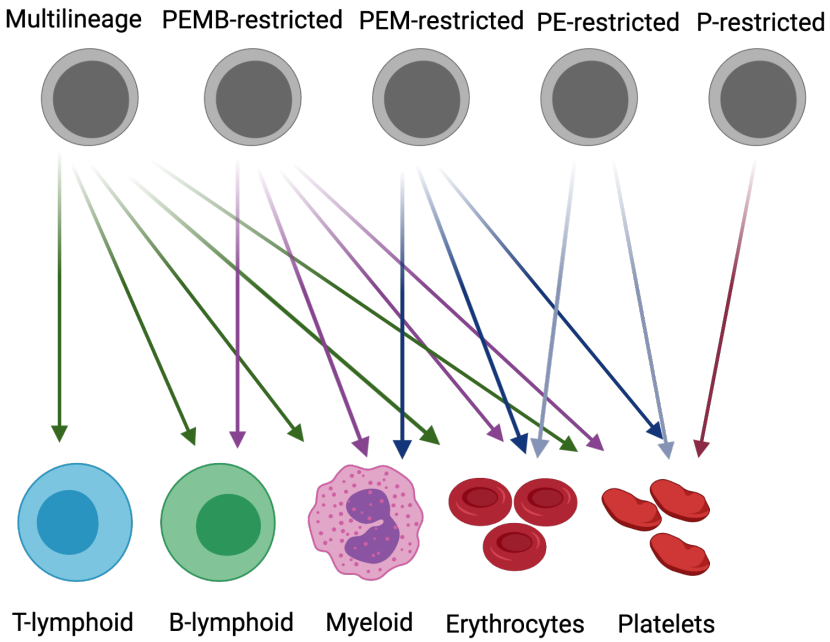


Figure 1.5.2. Fates of the five classes of murine HSC clones based on their lineage contribution.
The most recent study by Carrelha *et al* (2018) described HSCs restricted towards generating platelet (P-restricted), platelet-erythroid (PE-restricted), platelet-erythroid-myeloid (PEM-restricted), and platelet-erythroid-myeloid-B cell (PEMB-restricted). They also reported Multilineage HSCs that replenish all five mature lineages [202].

To date, a large body of evidence supporting intrinsic-related HSC heterogeneity derived from studies in mice; little is known for equivalent human HSC subpopulations. The utilisation of human CD45RA surface marker in combination with CD90 and CD49f in flow cytometry distinguished the short-term self-renewing human MPPs, and the long-term self-renewing human HSCs, showing the first pieces of evidence of multiple haematopoietic cell types with different self-renewal ability and reconstitution kinetics [96, 97]. Yet, the purity of the sorted functional human HSCs based on their immunophenotype is not as efficient as in the mouse, which affects the downstream analysis and processes. As mentioned earlier, the current human HSC definition relies largely on their purification based on a few cell surface marker expressions and their subsequent transplantation into murine recipients. In particular, most of the work was performed using UCB-derived HSCs. Recently, Laurenti *et al* showed that the UCB- and BM-derived HSCs were heterogeneous in terms of their cell cycle properties and that the timing of exit from quiescence was regulated by the CDK6 levels [14]. Although both ST-HSCs ($Lin^-CD34^+CD38^-CD45RA^-CD90^-CD49f^-$) and LT-HSCs ($Lin^-CD34^+CD38^-CD45RA^-CD90^+CD49f^+$) were in a quiescent state (G_0 state), ST-HSCs had higher proliferation and more frequent divisions compared to LT-HSCs [14]. This was due to the upregulation of CDK6 expression (independent of CDK/cyclinD complex) in ST-HSCs, which controls rapid entry to the cell cycle upon mitogenic stimulation. However, CDK6 expression was undetectable in LT-HSCs, limiting the number of divisions they undergo [14]. Another study from the Laurenti group demonstrated heterogeneity in a few studied lineage outputs of UCB-derived HSC/MPPs ($CD19^-CD34^+CD38^-CD45RA^-$) and LT-HSCs using an optimised single-cell colony-forming assay and single cell RNA-seq [205]. They detected two subsets in the HSC/MPP

pool and LT-HSCs based on the differential expression of CLEC9A and CD34 surface markers: CLEC9A^{hi}CD34^{lo} (referred to as Subset 1 HSC/MPP and Subset 1 CD49f^{+}, respectively) fraction were enriched for single cells with My/Ery and Mk potential *in vitro* and upregulated for HSC-related and Ery/Mk lineage priming programmes. CLEC9A^{lo}CD34^{hi} (Subset 2 HSC/MPP and Subset 2 CD49f^{+}) fraction consisted of single cells with My/Ly potential *in vitro* and upregulated for cell cycle genes and lymphoid differentiation-related gene-sets [205]. The above findings, *in vivo* experiments (demonstrating the multipotency of Subset 1 HSC/MPP and Subset 1 CD49f^{+}) and bulk CD49f+ *in vitro* culture experiments (demonstrating Subset 1 CD49f^{+} cells could give rise to Subset 2 CD49f^{+} cells), suggested that the Subset 1 CD49f^{+} hierarchically precede the Subset 2 CD49f^{+} cells [205]. In another study, by Knapp *et al*, using mass cytometric analysis, single-cell *in vitro* culture assays as well as single-cell cellular barcoding of UCB-derived HSPCs/HSCs and their bulk transplantation into irradiated immunodeficient mice, the authors identified a phenotypically distinct and serially transplantable LT-HSC subpopulation (CD33^{+}CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}) [214]. However, this study did not report heterogeneity in lineage output profiles of HSCs [214].

Collectively, the cell fate decisions between self-renewal and differentiation are made at the single-cell level and there is a crucial need to develop reliable methods that trace the single-cell behaviours for understanding haematopoiesis in health and disease in mouse and human. The biological causes of HSC heterogeneity are still a matter of intense research. It is assumed to be mainly regulated by intrinsic mechanisms such as gene transcription, epigenetic alterations, asymmetric cell

divisions, DNA mutation, cell cycle, cell size, and ageing [9-13, 197]. However, it is likely that extrinsic signals by the BM niche (i.e., cell-cell contacts, signalling pathways, cytokines) may play a role in HSC heterogeneity, meaning that the progeny of different HSC subsets could show different sensitivity to these signals. Thus, this behaviour may influence their cell fate choices by supporting survival and proliferation [215, 216]. Considering the high level of heterogeneity identified in murine BM and UCB, human adult HSCs remain poorly characterised, and it is important to understand the behaviour of individual adult HSCs in a polyclonal environment and the number of simultaneously active HSC clones that contribute to haematopoiesis. In this perspective, in the next section, the aspects of *in vivo* lineage tracing and some of these methodologies will be discussed.

1.6. *In vivo* lineage tracing

As discussed in the previous chapter, single-cell transplantation experiments are important for demonstrating what a single cell can do on its own as well as functional and transcriptional differences within phenotypically defined HSC populations. Those experiments contributed to our recognition of HSC heterogeneity; however, they had several limitations. First, it is very costly, time-consuming, and technically challenging to collect robust data (i.e., requiring a large number of mice) to represent the whole population [217]. Also, single-cell transplantations lack testing of cell interactions within the niche and do not provide any information about how HSCs behave in a polyclonal environment. These issues have been partially overcome by adopting *in vivo* lineage tracing techniques. Lineage tracing is the methodology used to track a single cell from its origin and follow its fate in various organisms. The key in lineage tracing is that the label must be transmitted to all progeny of the

founder cell during cell division (not to any other neighbouring cell) and needs to be retained over generations [218]. Furthermore, the labelling procedure should not change the properties of the tagged cell, its output or its interactions with neighbouring cells [218]. To date, live-cell imaging, various vital dyes, radioactive tracers, multicolour reporters, naturally occurring mutations, and genetic tags have been used to conduct lineage tracing. They will be introduced briefly in this section.

Historically, lineage relationships have been observed by imaging single-cells over time using time-lapse microscopy and microinjection of vital dyes into single-cells [219-221]. Those techniques have a few caveats and technical challenges: culture conditions for each cell type being imaged should be optimised without inducing cell-death or changes in cell behaviour. Additionally, an automated cell tracking platform needs to be optimised for the cell of interest and requires in-depth knowledge [218, 222]. Vital dyes (i.e., a dye that stains cells without killing them) were initially used to study the cell fate of *Xenopus* embryos [223]. However, vital dyes are soluble in water and may diffuse into the neighbouring cells, thereby the direct progeny of the founder cell cannot be distinguished [218]. This issue was partially overcome using hydrophobic yet lipid-soluble fluorescent carbocyanine dyes to track the development of trunk neural crests [220] and zebrafish anterior nervous system [221]. However, the dye was diluted after multiple rounds of cell division and made it harder to detect in daughter cells [218]. Radioactive Thymidine analogues (^{3}H-Thymidine, BrdU, or EdU) incorporating into newly synthesised DNA, were utilised to track slow-cycling cells (e.g. HSCs) and their lineages. However, longer tracing periods (over 10-days) using BrdU proved to be toxic to various cell types such as neuronal stem/progenitor and mature haematopoietic cells [224-226].

It was also reported that BrdU toxicity affected HSC behaviour by inducing them to exit from the quiescent state and enter the cell cycle [227]. To use this labelling approach, understanding the division kinetics of the tissue of interest is essential due to BrdU dilution following each cell division.

Some other studies use endogenous randomly occurring somatic mutations such as single nucleotide variants, copy number variants, and microsatellites to infer fate decisions [228-231]. This is a useful methodology especially for retrospective analysis of human haematological malignancies [232, 233]. Although these methods do not require external manipulation to track the cells, the detection of nuclear somatic mutations by whole-genome sequencing in each cell is costly, contains high error rates, and it remains difficult to quantify the contributions of different clones [234, 235]. Recently, mutations in mitochondrial DNA were used to track single cells and determine lineage relationships [236]. Mitochondria have a sufficiently small genome for cost-effective sequencing, it is transcribed, and contain a higher mutation rate compared to the nuclear genome, holding a great promise for retrospective lineage tracing [236].

Multicolour labelling systems represent another method that can be used for *in vivo* lineage tracing. This can be achieved by simultaneous introduction of different coloured lentiviral vectors [237] or genetic recombination of fluorescent-reporter-based systems (such as Brainbow or Confetti) [238, 239]. Although these methods have been successfully used for *in situ* analysis, the number of blood clones that can be tracked in parallel is restricted due to the limitation of the distinct colours.

To increase the number of clones that can be tracked at a given time, several types of genetic marking has been applied including retroviral integration [61, 118, 240, 241], Polylox [242, 243], Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9-based lineage tracing [244, 245], transposon integration [246, 247], and cellular barcoding [106, 248-250]. These methods permit simultaneous tracking of hundreds to thousands of clones using high-throughput sequencing (lineage tracing using the cellular barcoding method will be discussed separately in the next-sub section, 1.6.1). The use of genetic markers for lineage tracing has proved to be advantageous because they can stably be expressed under a strong promoter, are inherited by the progeny of the cell of origin, and do not contaminate neighbouring cells. Lineage tracing using the retroviral integration sites as genetic tags have been used in biological research and clinical settings to track the individual HSC clones [61, 240, 241, 251, 252]. For instance, patients receiving autologous gene corrected HSPCs allow the study of the human haematopoietic system at the clonal level. This method relies on the cutting of genomic DNA of labelled cells with restriction enzymes, amplification of insertion sites by polymerase chain reaction (PCR), and detection by sequencing. The enrichment of a specific integration site has been used as an indicator of clonal expansion [253]. However, this method also has methodological constraints such as bias in fragment amplification of variable length, dependency on the viral integration site, and restriction endonucleases [254]. Lineage tracing methods using Cre recombinases can also genetically label cells based on the activity of tissue or cell-specific promoter (e.g., Brainbow mice). Most recently, Cre-recombinase driven Polylox barcoding was developed by Pei *et al* to track embryonically derived HSC clones [242, 243, 255]. Polylox is an artificial DNA recombination locus consisted of ten

flanking loxP sites and marks cells through tamoxifen-inducible Cre-loxP recombination [255]. Transposon tagging (e.g., Sleeping Beauty) relies on the semi-random acute mobilisation of only one single transposable element in which integration sites serve as a heritable marker [246, 247]. In this lineage tracing concept, transposase expression is inducible and reversible and controlled by Tet-regulated promoters to prevent constant transposon mobilisation. Transposon tagging and Polylox systems generate barcodes *in vivo* that can also be referred to as endogenous barcodes, and the diversity of the barcodes are notably higher (several hundred thousand barcodes) than that of multicolour labelling. CRISPR-Cas9-mediated lineage tracing approaches, similar to transposon tagging and Polylox systems, allows barcode generation *in vivo* via engineering of animal lines with inducible Cas9 expression which was initially introduced in zebrafish embryos (e.g., GESTALT) [245] and more recently in mouse embryos (e.g. CARLIN) [244]. Using the CRISPR-Cas9-mediated lineage tracing method, cells are tagged with unique and heritable sequences that can serve as a barcode, generated through Cas9-induced DNA double-stranded break (DSB) repair system [244, 245]. The components of this mechanism (CRISPR-Cas9 DNA endonuclease, a panel of single-guide RNAs/homing guide RNAs, and an array of synthetic DNA target sites) generates high diversity of barcodes that accumulate over rounds of cell divisions and lineage relationships between cells can be inferred from the shared mutations [244, 245].

Altogether, the methods described above allowed tracking cell fate decision and inferring cell lineage relationships in the haematopoietic system; however, these methods were limited by DNA-based barcode sequencing, which neglects the cell

transcriptome. Recent advances in sequencing technologies allowed researchers to develop mouse models that enable the read out of both lineage output of single cells in parallel with the single-cell transcriptome [244, 247, 256]. These integrated measurements are exciting and likely to shed light on the mysteries of cellular identity and differentiation.

1.6.1. Cellular barcoding

Cellular barcoding can also be referred to as lentiviral barcoding, a powerful *in vivo* lineage tracing tool to assay multiple single cells in parallel and to address individual cell fates (it has been used in this study). Cellular barcoding relies on tagging individual cells of interest with unique DNA sequences termed “barcodes” that are stably integrated into the genome and inherited by their progeny [106, 249, 250, 257, 258]. These unique heritable random DNA sequence tags are introduced into target cells with retro- or lentiviral vectors that express a fluorescent reporter. Following this, uniquely barcoded cells are subsequently transplanted into recipient mice. In this manner, the clonal progeny of the cell of interest can be tracked *in vivo* over time. Thus, different cell types can be purified, and their genomic DNA can be analysed for barcodes with high-throughput next-generation sequencing (NGS) (Figure 1.6.1). To date, it has been used to track T-cell behaviour [259], haematopoietic and cancer cell lines [260, 261], breast tissue development [262], murine HSCs [249, 263-266], murine LMPPs [258, 267] and CMPs [268], rhesus macaque HSPCs [269-273], patient-derived leukaemia cells [274, 275] and UCB-derived HSPCs [106, 214, 248].

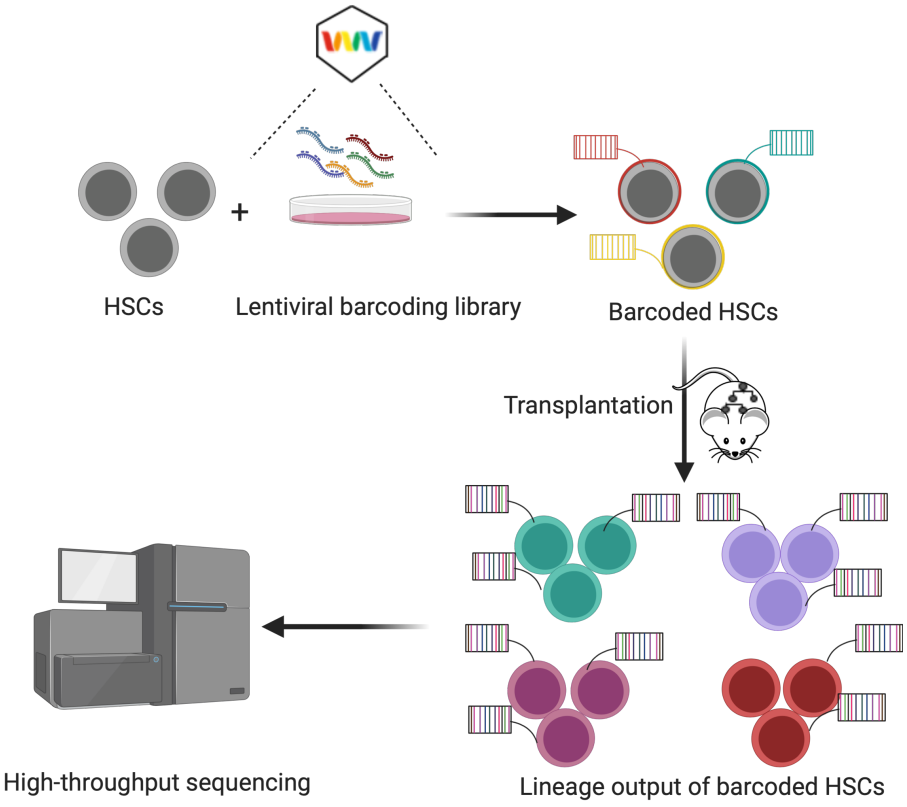


Figure 1.6.1. Cellular barcoding of HSCs
HSCs are transduced with a lentiviral barcode library which introduces a unique genetic label into the genome of each HSC. This is usually followed by their transplantation into a recipient animal. Lineage output of HSCs can be purified from the recipient animal and barcode tags in their genome can be studied using high-throughput sequencing. Created with BioRender.com.

Previous cellular barcoding studies utilised DNA-sequencing based methods for the analysis of barcode representation in various bulk cell lineages [106, 249, 260-263]. Although such approaches can reliably measure the number of HSC clones contributing to haematopoiesis, it does not provide data about the individual, single HSC behaviour and its lineage choices, i.e., does not yield insights into functional and transcriptional heterogeneity. Coupling cellular barcoding and single-cell transcriptomic approaches enable the comparison of the shared and distinct barcodes between different progeny of the founder cells as well as linking cell lineage history to the HSC transcriptome. This approach allows a deeper analysis of the HSC cell identity and the reconstruction of a more comprehensive lineage

tree. Recently, some studies implemented joint read out of barcodes in lineages with single-cell transcriptomics. These studies used *in vivo* lineage tracing methodologies such as CRISPR-Cas9-based barcoding [276, 277], PolyloxExpress [243] and Sleeping Beauty transposon [247], or a DNA barcode construct referred to as LARRY [278, 279]. Combining LARRY lentiviral barcodes with single-cell transcriptomics, Fraticelli *et al* demonstrated murine Mk-biased HSCs with low-output repopulation activity and Multilineage HSCs with high repopulation activity. This study also revealed transcriptional differences among the HSCs with distinct fate patterns, which might be a possible explanation of functional heterogeneity in the murine HSC compartment [279].

Historically, the use of DNA barcodes to count stem cell clones that contribute to haematopoiesis relied on the identification of retroviral integration sites (as mentioned earlier in section 1.6). However, measuring clonal contribution based on the sequencing reads is inaccurate and amplification of different length fragments introduces biases [280]. On the other hand, retrieving the identity of the cell from the integrated sequence rather than the integration site overcomes many of these issues. The DNA barcode library contains repeated variable sequences (random nucleotides) that consist of interspersed with fixed sequences. Therefore, it is composed of unique sequences equal in length (~30 base pairs) that can be detected by a set of uniform primers [254]. In contrast to single-cell transplantations, it allows tracking multiple HSC clones in a single mouse, thus revealing the behaviour of HSC clones in a polyclonal environment. Recently, a number of transgenic reporter systems have emerged which allow the study of steady-state haematopoiesis. Polylox [242, 243] and CRISPR-based barcoding [244] have an

advantage due to the generation of barcodes *in vivo* through HSC-specific inducible reporter systems, allowing the interrogation of haematopoiesis in a native setting; whereas, lentiviral barcoding relies on *in vitro* manipulation of cells and this is followed by their transplantation. However, these *in vivo* barcoding techniques are only available in mouse models and are therefore not suitable for studying human haematopoiesis. The effect of lentiviral barcoding becomes less of a problem when studying transplantation haematopoiesis because of its similarity in the design of gene therapy approaches, one of the widely used treatment in the clinic [217]. Overall, cellular barcoding offers many advantages over other techniques due to its simplicity in barcode design, applicability to different vector systems and tissues, high-throughput capacity, unbiased analysis, and suitability to combine with single-cell transcriptomics (see below section 1.7).

1.7. Single-cell transcriptomic approaches for studying haematopoiesis and advanced haematopoiesis models

To determine the role of the genome in cellular functions and cell fate decisions, researchers sought to explore the products of the genome: RNA and proteins. However, proteins are dynamic, can change their shape, as well as contain several post-translational modifications that make proteins challenging to study [281]. On the other hand, RNA serves as an intermediate molecule between gene and proteins serve as a molecular linkage, allowing the study of the cell functions during development as well as in health and disease. All cells of an individual have the same genome, but not all genes are transcriptionally active in all cells, and they vary in their expression levels (stochastic); therefore, unique patterns of gene expression are observed in every cell type and that determines cellular phenotype and function

[282]. Additionally, several messenger RNA (mRNA; constitutes codons for amino acids) variants are formed from each gene as a result of alternative splicing, multiple transcription-start sites, and polyadenylation sites [283, 284]. Thus, the transcriptome conveys more information than the genome and is indicative of gene activity.

RNA-sequencing (RNA-seq) is one of the latest approaches that enabled characterising and profiling of transcriptomes using deep-sequencing technologies. In the last decade, the advances in single-cell transcriptomics and computational approaches for analysis of the vast amount of sequencing data have enabled scientists to address questions that could not be shown using bulk-cell RNA-seq or microarray technologies [285-287] Although traditional bulk RNA-seq analysis has shown a great transcriptional diversity across different haematopoietic populations [288, 289], this sort of analysis averages out the gene expression of all cells in the sample, therefore, inherently masking the biological differences between every single cell within a population or tissue. There is now a large body of evidence that seemingly homogenous populations demonstrate differences in their transcriptional expression [285, 287, 290]. Single-cell RNA sequencing (scRNA-seq) has allowed the profiling of thousands of individual cells based on their gene expression signatures. In this way, single-cell transcriptomics gives an in-depth overview of the entire cell population, allowing characterisation of cell-type diversity within a phenotypically homogenous sample. Moreover, it enables the identification of new subpopulations/rare subsets based on the transcriptomic landscape and computational reconstruction of developmental trajectories. However, a single cell contains small number of mRNA transcripts, therefore it needs to be considered that

scRNA-seq data are often extremely noisy in nature and contains technical dropouts [291].

Different single-cell sequencing methods have been developed until now. These methods differ in their detection sensitivity, throughput, and scalability. Some high-throughput techniques (e.g. droplet-based sequencing) such as 10X Genomics Chromium [292], Drop-Seq [293], and InDrops [294] are able to profile up to 10^4 cells and measure 1000-3000 expressed genes per cell; however, they can capture only the 3'-terminal part of the transcripts [295]. Droplet-based methods encapsulate individual cells (from a cell suspension) in nanolitre droplets containing DNA-barcoded beads to identify each cell and apply unique molecular identifiers for correction of biases [292-294, 296]. On the other hand, low-throughput methods (e.g. Smart-seq2 [297], CelSeq2 [298], and mcSCRB-Seq [299]), can analyse up to 10^3 cells, but detect more than 5000 genes per cell [295] (various single-cell RNA sequencing methods are reviewed and compared in depth elsewhere [291, 300-302]). Among the latter, Smart-seq2 has emerged as a gold-standard technique as it allows analysis of single-cells and a low number of cell populations in an unbiased manner, at high sensitivity and less variability [297, 303, 304]. In addition, Smart-seq2 enables characterising the full-length transcripts of the polyadenylated mRNAs. Single-cells from the cell suspension can be purified by FACS directly into lysis buffer that contains oligo-dT nucleotides which capture the mRNAs from the poly(A) tail [297]. The Smart-seq2 protocol further improved by the Mead group for high sensitivity single-cell mutation detection (BCR-ABL1) along with the transcriptome in human chronic myeloid leukaemia samples (the method referred to as targeted Smart-seq2 and it has been used in this work). This method allowed

researchers to detect genes of interest which could not have been detected reliably using Smart-seq2 due to the technical dropouts and the stochastic nature of gene expression in each cell [305]. This technique linked the mutation information to the transcriptome of the cells. However, only expressed genes (mutations) could be identified with this technique, but not the lowly expressed genes or non-coding mutations. For this reason, Rodriguez-Meira *et al* developed TARGET-seq to detect both coding and non-coding mutations in parallel to the transcriptomes of cells [306].

The power of scRNA-seq techniques and advances in computational methods allowed researchers to purify single cells of different developmental cell states and study their transcriptomes. Single-cell transplantations and single-cell sequencing technologies have unveiled intrinsic heterogeneity (i.e., heterogeneity in self-renewal potential, lineage contribution etc.) in the haematopoietic system and proposed novel developmental trajectories which interrogate the conventional haematopoietic tree (some of these important findings was discussed in section 1.5). A growing body of evidence suggests that only a portion of the mouse HSC compartment contributes to the balanced production of lineage output; distinct classes of HSCs exist with biased and restricted lineage outputs [201-203].

The lineage commitment to haematopoietic cells is regulated by the expression of lineage-affiliated genes through the activation of lineage-affiliated transcriptional programs and downregulation of programs related to alternative lineages concomitantly [307]. Recently, single-cell assays, as well as large-scale single-cell transcriptomic studies that cluster cells computationally based on the coexpression of lineage-affiliated genes proposed that there were no discrete progenitor states, it

was rather gradual and continuous differentiation spectrum of progenitors [98, 195, 285, 308, 309]. These studies suggested that lineage commitment occurred earlier than anticipated in the haematopoietic tree, and distinct lineages (unipotent lineages) may emerge directly from HSC/HSPCs without passing through a series of oligopotent progenitors [98, 285]. Another study, from the Nerlov group, suggested early bifurcation of murine myeloid oligopotent progenitors based on *Gata1* expression [310] using scRNA-seq and single-cell quantitative real-time PCR. *Gata1* gene encodes the transcription factor GATA-1 which plays an important role in Mk and Ery development [311, 312], also expressed in eosinophils, basophils, and mast cells [310, 313-315]. In that study, by generating a *Gata1*-EGFP reporter mouse line, it was demonstrated that *Gata1*-EGFP^{-} progenitors exist with neutrophil/monocyte-restricted potential, whereas *Gata1*-EGFP^{+} cells have eosinophil/basophil/mast cell-restricted potential [310]. Moreover, eosinophil/basophil/mast cell and megakaryocyte/erythroid cells were distinct from neutrophil/monocyte and lymphoid populations. These observations indicated that *Gata1*-positive myeloid cells (eosinophil, basophil, and mast cells) separated from *Gata1*-negative myeloid lineages (neutrophil and monocytes) at even the most primitive stages [310, 316]. Similarly, the Nerlov group also demonstrated the segregation of human myeloid progenitors into neutrophil/monocytes and eosinophil/basophil/mast cells based on the *Gata1* expression (Figure 1.7.1) [317]. The data generated during these studies led the authors to propose CD114 and CD131 as surface markers to define CMP subpopulations that have neutrophil/monocytes (CD114^{+} CMP in human and *Gata1*$^{-}$ GMP in murine) and eosinophil/basophil/mast cell potential (CD131^{+} CMP in human and *Gata1*$^{+}$ GMP in murine), respectively [317]. Together, these findings demonstrate that myeloid cells

can derive from different upstream cells both in mouse and human, enforcing a revised model of myelopoiesis (Figure 1.7.1). All these studies that challenge the classical haematopoietic hierarchy accept the general broad lines and directions of the classical haematopoietic tree. However, they propose that the classical tree represents an oversimplification and haematopoiesis is more complex than anticipated as haematopoietic populations seem to be heterogeneous, and thus transitions and subpopulations in the haematopoietic tree need to be identified and functionally characterised [285, 308, 310, 317-319].

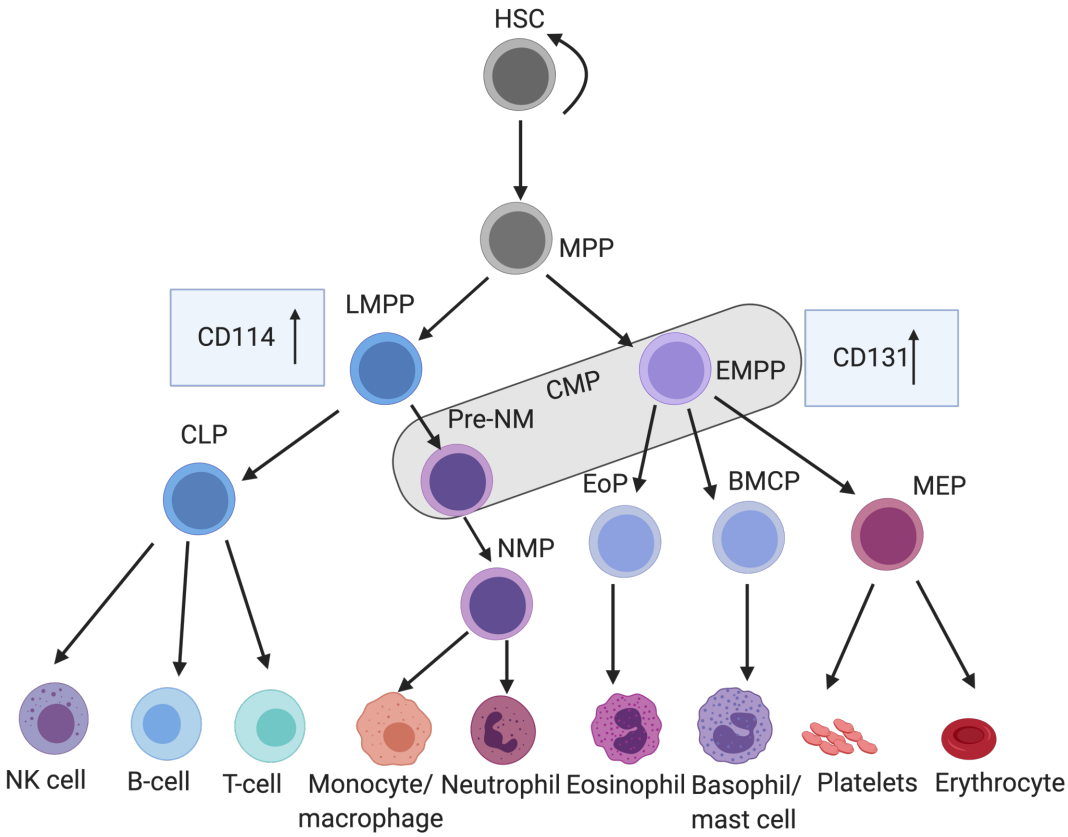


Figure 1.7.1. Advanced human haematopoietic hierarchy
The myelopoiesis occurs via erythroid/megakaryocyte-primed MPP (EMPP) which corresponds to CMP $CD131^{+}$ and pre-neutrophil-monocyte progenitor (Pre-NM) corresponds to $CD114^{+}$ CMP. This model suggests that eosinophil/basophil and neutrophil/monocyte are generated by different progenitors (based on Drissen's model [317]). NMP, neutrophil-monocyte progenitor; EoP, eosinophil progenitor; BMCP, basophil-mast cell progenitor. Created with Biorender.com

Due to the explosion of single-cell technologies, there is a wealth of transcriptomic data helping researchers to evolve the understanding of murine and human haematopoiesis. However, there is also a growing need to understand the strengths and limitations of the current techniques used to study heterogeneity in haematopoiesis. Diverging research findings may be the result of different scRNA-seq methods, library depths as well as computational analysis pipelines used [316]. Therefore, it is critical to understand how to interpret and integrate such data. After the demonstration of HSC heterogeneity in the murine system, several studies have emerged with haematopoietic lineage trajectory predictions for human, based solely on single-cell transcriptomic data [285-287, 320]. Even though these studies provide detailed transcriptional profiling of many haematopoietic populations in different stages of differentiation, they cannot shed light on the true genetic lineage relationships between different populations. As discussed earlier, combinatorial techniques, particularly transcriptomics and lineage tracing, will be crucial in the haematological research enabling refinement of the lineage relationships based on genetic labels along with detailed phenotyping. Besides, they will allow researchers to assess changes in gene expression dynamics at different developmental cell states, with potential to revolutionise haematopoietic research. Although recently few research groups combined lineage tracing tool with scRNA-seq for the identification of murine HSC heterogeneity [243, 244, 247, 278], yet scRNA-seq has not been applied in combination with a lineage tracing method for the elucidation of heterogeneity in human adult HSCs.

1.8. Aims and objectives

It has long been assumed that all HSCs had the capacity to produce all blood lineages. Single cell and serial transplantation studies in mice demonstrated that there is a profound degree of heterogeneity in murine HSCs with the majority of them displaying a restricted pattern of lineage output [201-204]. Such studies have challenged the idea of functionally homogeneous HSC subsets and instead suggest that individual HSCs differ in their properties. Considering the high level of heterogeneity identified in murine BM, a similar extent of heterogeneity may also exist in human HSCs. This has recently been shown in cell cycling properties [14], serial regenerative activity [214], and in lineage output of UCB-derived ST-HSCs and/or LT-HSCs [205]. However, human adult HSCs remain poorly characterised. Ideally, the clonal dynamics of human HSCs should be analysed by tracking the fate of progenitors from individual HSCs *in vivo*, thereby unveiling the haematopoietic repopulation dynamics of HSCs post-transplantation. The importance of characterising human adult HSCs is apparent from their frequent use in the clinic, e.g., BM transplantations in patients with haematological malignancies.

The overall aim of the thesis is to characterise HSC heterogeneity in human adult BM and to identify transcriptional differences between functionally distinct human HSCs. To achieve this, we utilised cellular barcoding for *in vivo* lineage tracking of human adult HSCs and their lineage output in combination with single-cell transcriptomics. The workflow includes different steps: first, I establish the FACS purification of human adult BM HSCs, their lentiviral barcoding, and their engraftment in non-irradiated immunodeficient NBSGW mouse BM as well as the

purification of barcoded HSCs and their lineage output (B-Ly, My, Mk, and Ery) from the xenografts. Additionally, I perform single-cell RNA-seq to characterise the human HSC transcriptome and to identify the barcode of each HSC. Moreover, to identify the barcode representation of the mature cell output of HSCs, I optimise and perform bulk RNA-seq. Together, these allow us to correlate HSC lineage output with HSC transcriptome simultaneously and, in particular, to identify transcriptomic signatures that may be associated with specific haematopoietic lineages.

Chapter 2: Materials and Methods

2.1. Animals

The NBSGW (NOD.Cg-Kit^{W-41J}Tyr^{+}Prkdc^{scid}Il2rg^{tm1Wjl}/ThomJ) mouse strain was obtained from the Jackson Laboratory (Stock #026622) and bred in the animal facility at the University of Oxford (Biomedical Services). This mouse strain was generated by intercrossing NSG strain with mice homozygous for the Kit^{W41} allele, particularly with C57BL/6J-Kit^{W-41J}/J (C57BL/6. KitW41) strain [148]. All experimental procedures, mouse breeding, and maintenance were in accordance with the UK Home Office regulations and approved by the relevant authorities (University of Oxford).

2.2. Construction and validation of barcoded lentiviral vector libraries

The peGZ2-linkerBC322 barcoding library consists of 725 defined barcodes and it was kindly provided by Leonid Bystrykh (University of Groningen, Netherlands) (the plasmid map is shown in Figure 2.2.1). The generation and the validation of plasmid library stock were performed by Ryan Beveridge (Virus Screening Facility, WIMM). Briefly, 300 ng of the barcoding library was electroporated into 50 µl of one-shot Stbl3 chemically competent *E. coli* bacteria (Invitrogen) grown at 37°C for an hour, and then plated onto 50 x 15 cm Luria broth (LB) plates supplemented with Ampicillin. Four plates were counted with an average colony count of 1420±122.5 colonies indicating ~100-fold coverage of the library. Plates were incubated for 18 hours (hrs) at 30°C before the bacteria were harvested. Plasmid DNA extracted using a Qiagen Plasmid Maxi Kit (Qiagen, 12162) following the manufacturer's protocol.

The lentiviral barcode library was produced by transfecting human embryonic kidney (HEK293T) cells with 5 µg of BC322, 3.75 µg psPAX2, and 1.25 µg of pMD2.G in complex with PEI Pro (PolyPlus, 1 µl per µg), supernatants were collected 48 hrs and 72 hrs post-transfection. The virus was concentrated by ultracentrifugation at 98,000 g for 2 hrs at 4°C (SW41 rotor and Optima XPN80). To validate the library, HEK293T cells were transduced with the barcoding library at a Multiplicity of Infection (MOI) of 0.3, selected with puromycin, and their genomic DNA was harvested using Qiagen Blood Maxi kit (Qiagen, 51194) following the manufacturer's instructions. The barcodes were amplified using custom primers around the barcoding region (Figure 2.2.1) with NexteraXT compatible overhangs allowing for indexing. Indexed samples were sequenced on an Illumina MiSeq. Custom scripts to identify barcode sequences were written by Simon McGowan (CBRG, WIMM). All reads were searched for a sequence matching the following barcode pattern: NNNACNNNGTNNNTANNNCANNNTGNNN ("N" refers to random nucleotides that generate barcode diversity, flanked by fixed nucleotides).

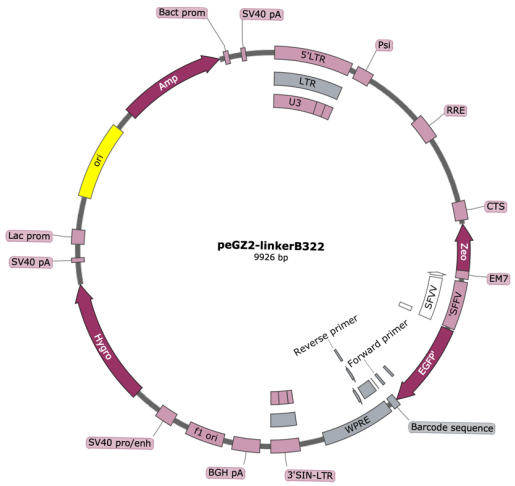


Figure 2.2.1.The map of peGZ2-linkerB322 vector.
The barcode is situated by the EGFP reporter gene and their expression is controlled by a spleen focus forming virus (SFFV) promoter, therefore, it is expressed as mRNA and polyadenylated. The barcode can be amplified and detected using a set of custom primers which binds to the end region of EGFP and Woodchuck hepatitis virus regulatory element (WPRE).

2.3. Processing of human bone marrow samples

Fresh unprocessed BM samples of anonymous healthy young donors (at the age of 20 to 30 years old) were purchased from AllCells or Lonza. BM cells were filtered with 70 μm cell strainers (Corning) to remove any cell clumps and divided into 50 ml Falcon tubes (Corning). Mononuclear cells (MNCs) were isolated by Ficoll (GE Healthcare, 17-1440-02) density gradient centrifugation. 13 ml of BM were diluted with FACS buffer in 1:1 proportion (Iscove's Modified Dulbecco's Medium (IMDM, Thermo Scientific) + 10% Foetal Bovine Serum (FBS, Thermo Scientific) + 20 μg/ml DNAse). 26 ml of BM dilution were layered gently on top of 15 ml of Ficoll and centrifuged at 400xg for 25 minutes without break at room temperature. Next, the upper phase was removed and the interphase of MNCs was transferred to a new 50 ml falcon tube. MNCs were washed with 2x volume of FACS buffer and spun down for 5 minutes at 500xg. The number of MNCs was determined using both the Sysmex and haemocytometer.

2.4. CD34 enrichment

Following MNC isolation, HSPCs were enriched using the CD34 Microbead Kit (Miltenyi Biotech, 130-046-702) and LS MACS separation columns (Miltenyi Biotech, 130-042-401). CD34 enrichment was carried out following the manufacturer's protocol, and the volume of reagents was adjusted according to the number of cells. The MNC suspension was centrifuged at 500xg for 5 minutes and the supernatant was removed completely. The cell pellet was resuspended in 300 μl of PBE buffer (Phosphate-buffered Saline (PBS) + 2% FBS + 2 mM Ethylenediamine tetra acetic acid (EDTA)) and incubated with 100 μl of FcR blocking reagent, and 100 μL of CD34 Microbeads (for up to $1x10^8$ cells) for 30

minutes on ice. After incubation, the cell suspension was washed with PBE buffer and spun down at 500xg for 5 minutes. The cell pellet was resuspended in 500 μL of PBE buffer for up to $1x10^8$ cells.

LS MACS column and a 30 μm filter (Sysmex) were placed in the MACS separator (Miltenyi Biotech) and pre-washed with 3 ml of PBE buffer. The cell suspension was applied onto the column and the flow-through (unlabelled cells) collected in a 50 ml falcon tube. This was followed by washing the column three times with 3 ml of PBE buffer. After washing steps, the column was removed from the MACS magnetic separator. The labelled $CD34^+$ cells were eluted in 5 ml of PBE buffer in a 15 ml falcon tube by pushing the plunger into the column. The number of $CD34^+$ cells was determined by counting on a haemocytometer.

2.5. Fluorescence-activated cell sorting staining and HSC purification

All antibodies and their dilutions used for FACS staining were listed in Table 2.5.1. All antibody incubations for FACS were carried out in FACS buffer for 20 minutes on ice, in the dark. After the staining, cells were washed with 1 ml of FACS buffer, centrifuged at 500xg for 5 minutes, resuspended in FACS buffer, and transferred into polypropylene tubes (BD Falcon). In this study, a BD FACS ARIA Fusion I was used for the cell sorting and the flow cytometry analysis, unless otherwise stated. The experiments included unstained, single colour stained (CompBeads, BD Biosciences) and Fluorescence Minus One controls to check the background

staining and compensation setup. Before sorting, the sorter alignment was checked by sorting fluorescent beads into a respective collection tube.

Table 2.5.1 List of anti-human antibodies used for FACS staining

Antibody	Fluorochrome	Clone	Dilution	Company	Catalogue no.
CD2	PE-Cy5	RPA-2.10	1/80	Biolegend	300210
CD3	PE-Cy5	HIT3a	1/80	Biolegend	300310
CD4	PE-Cy5	RPA-T4	1/80	Biolegend	300510
CD7	PE-Cy5	CD7-6B7	1/80	Biolegend	343110
CD8	PE-Cy5	RPA-T8	1/80	Biolegend	301010
CD10	PE-Cy5	HI10a	1/80	Biolegend	312206
CD11b	PE-Cy5	ICRF44	1/80	Biolegend	301308
CD14	PE-Cy5	61D3	1/80	Biolegend	15-0149-42
CD19	PE-Cy5	HIB19	1/80	Biolegend	302210
CD20	PE-Cy5	2H7	1/160	Biolegend	302308
CD56	PE-Cy5	MEM-188	1/40	Biolegend	304608
CD235ab	PE-Cy5	HIR2	1/500	Biolegend	306606
CD34	APC	581	1/160	Biolegend	343510
CD38	PE-TXRED	HIT2	1/40	Invitrogen	MHCD3817
CD45RA	FITC	HI100	1/160	Biolegend	304148
CD45RA	AF700	HI100	1/50	Biolegend	304120
CD90	PE	5E_10	1/40	Biolegend	328110
CD10	BV421	HI10a	1/20	Biolegend	312218
CD123	PE-Cy7	6H6	1/40	Biolegend	306010
CD71	APC-Cy7	CY1G4	1/50	Biolegend	334110
CD33	PE	WM53	1/100	Biolegend	303404
CD19	PE-Cy7	HIB19	1/100	Biolegend	302216
CD45	BV650	HI30	1/200	Biolegend	304044
CD41	AF700	HIP8	1/50	Biolegend	303728
CD45 (anti-mouse)	BV605	30-F11	1/40	Biolegend	103155

The human BM samples enriched for $CD34^+$ cells were stained with an antibody cocktail against the following human markers: lineage cocktail PE-Cy5 (CD2, CD3, CD7, CD8, CD11b, CD14, CD19, CD20, CD56, and CD235a), CD34 APC, CD38 PE-Texas Red, CD90 PE, CD45RA FITC, CD10 BV650 and CD123 PE-Cy7. Cells were kept on ice until the analysis, in the dark. 7-amino-actinomycin D (7AAD, Sigma Aldrich) viability dye was added to the samples prior to sorting to exclude

non-viable cells. HSCs ($Lin^-CD34^+CD38^-CD90^+CD45RA^-$) were sorted as a bulk into an Eppendorf tube containing StemSpan SFEM serum-free media (Stem Cell Technologies, 09650) supplemented with human cytokines Flt3 ligand (100 ng/ml), SCF (100 ng/ml), Thpo (50 ng/ml), and IL-6 (50 ng/ml) (all from Peprotech) [55].

2.6. Transduction of HSCs with a lentiviral barcoding library

Following FACS purification, HSCs were washed, spun down, and resuspended in StemSpan serum-free media supplemented with human cytokines Flt3 ligand, SCF, Thpo, and IL-6 at indicated concentrations. HSCs were transduced with the concentrated EGFP-expressing pe-GZ2-linkerBC322 lentiviral barcoding library ($2x10^9$ transduction units/ml (TU/ml) or $7x10^8$ TU/ml) in a well of 96-well flat bottom plate (Corning) at a MOI of ranging 40 to 125. MOI represents the number of infectious lentiviral particles per cell. In the earlier experiments performed (Exp 1-3, see Results Chapter 3), HSCs were pre-stimulated overnight in StemSpan media supplemented with cytokines prior to transduction. Our transduction protocol used for the earlier experiments and cytokine concentrations in cell media is very similar to the one used by the Gentner group for genetic engineering of HSPCs for gene therapy [55]. The number of cells transduced varied depending on the number of HSCs sorted in each experiment (see Results Chapter 3 for detailed statistics). The lentiviral transduction was performed overnight (~18 hrs) at 37^{o}C in a humidified incubator with an atmosphere of 5% CO_2.

After an overnight transduction, cells were collected, and the virus was washed away in IMDM media with 1% FBS. Cells were spun down at 500xg for 5 minutes

and resuspended in IMDM media with 1% FBS for *in vivo* transplantation (30 μl volume/mouse). A small aliquot of cells was maintained in culture for three days to check the transduction efficiencies as only few of the transduced cells would express EGFP at the time of the transplantation.

2.7. Intraosseous transplantation assay

Injections were performed under general anaesthesia with isoflurane (1-3%) and maintained throughout the procedure, using an anaesthetic machine. The depth of anaesthesia was checked before starting the procedure by observing respiratory rate and measuring pedal and tail pinch reflexes. The animal to inject was kept on a thermostatically controlled surface to prevent hypothermia throughout the procedure.

HSCs were transplanted into 8-11 weeks old female NBSGW murine recipients using an intraosseous injection because this technique has been reported to give higher engraftment levels compared to the intravenous route when the number of donor cells is limiting [50, 321]. The fur of the left tibia of the mouse was shaved and the injection area was revealed by cleaning the knee with ethanol wipes. Before injections, insulin syringes (volume: 0.5 ml gauge: 30G, BD micro-fine, 324893) were wetted with a buffer and the air inside was expelled. The cell suspension was distributed to insulin syringes (30 μl volume/mouse). The left knee was bent by 90° and the needle was inserted through the white ligament (patellar tendon). The needle was swivelled clockwise and counter clockwise while moving it slowly into the BM cavity, and the cells were injected into the left tibia. After the procedure, the

injected mice were transferred to a new clean cage and their recovery was monitored.

2.8. Xenograft peripheral blood analysis

The human cell engraftment in PB of transplanted recipients was analysed by flow cytometry at week 6, 10, and 16 post-transplantation. The PB samples were collected from the tail vein of the mice into lithium heparin-coated microvettes (Sarstedt). 100 μl of PB sample was diluted with 1 ml of PBE buffer. Samples were centrifuged at 500xg for 5 minutes and the supernatant discarded. Erythrocytes were lysed with 1 ml of red blood cell lysis buffer (0.15 M ammonium chloride,10 mM potassium bicarbonate, and 0.1 mM EDTA) and incubated for 5 minutes at room temperature. After incubation, samples were washed with 1 ml of FACS buffer and centrifuged at 500xg for 5 minutes. The supernatant was removed, and the cells were resuspended in FACS buffer.

PB samples were stained with the following WBC antibody panel: anti-human CD71 APC-Cy7, anti-human CD235a PE-Cy5, anti-human CD19 PE-Cy7, anti-human CD33 PE, anti-human CD45 BV650, anti-human CD41 AF700, and anti-mouse CD45 BV605. Samples were analysed by flow cytometry (LSR Fortessa X-20, BD Biosciences).

2.9. Xenograft bone marrow analysis

The human engraftment was first analysed 12 weeks post-transplantation by BM sampling from the live mice. BM aspiration was performed as described in section 2.7 but pulling the plunger back to create a negative pressure to aspirate BM. BM samples (~5 μl) were obtained from the injected (at week 12 and 16) or uninjected (at week 22) leg of the mice by aspiration using insulin syringes. Samples were collected in lithium heparin-coated microvettes (Sarstedt) containing 200 μl of PBE buffer. They were then washed with 1 ml of PBE buffer, spun down at 500xg for 5 minutes, stained, and analysed by flow cytometry (staining panel for this analysis is described below). Mice with reconstituted human cells were kept for further analysis up to 27 weeks post-transplantation for the earlier experiments performed (Exp 1-5) and 16 weeks for the later experiments (Exp 6-10) (see Results Chapter 3 for detailed statistics).

The engrafted mice were culled at week 27 (Exp 1-5) or week 16 (Exp 6-10) post-transplantations. BM cells were harvested from femurs, tibias, hips, sternum, and spine. BM cell suspensions were prepared by crushing bones using a pestle and a mortar in a small volume of FACS buffer and filtered using a 70 μm cell strainer (Corning). Cells were spun down at 500xg for 5 minutes and then the pellet was resuspended in PBE buffer. The cell number was determined by counting on a haemocytometer. BM cells were then subjected to human CD34 magnetic enrichment (as described in Chapter 2.4). The flow-through from this step (including human $CD34^-$ and mouse cells) was kept and this fraction was used to purify human

lineage populations (mature cell output of HSCs), while the $CD34^+$ population was used to sort HSCs.

Staining for FACS sorting was carried out as described in section 2.5. $CD34^+$ cells were stained with the following markers for single-cell index-sorting of $EGFP^+$ HSCs (HSCs defined as $Lin^-CD34^+CD38^-CD90^+CD45RA^-$) : anti-human lineage cocktail PE-Cy5 (CD2, CD3, CD7, CD8, CD10, CD11b, CD14, CD19, CD20, CD56, and CD235a), anti-human CD34 APC, anti-human CD38 PE-Texas Red, anti-human CD90 PE, anti-human CD45RA AF700, anti-human CD45 BV650, and anti-mouse CD45 BV605. 7AAD was used for dead cell exclusion.

BM aspirates and $CD34^-$ cells were stained with the following WBC panel: anti-human CD71 APC-Cy7, anti-human CD235a PE-Cy5, anti-human CD19 PE-Cy7, anti-human CD33 PE, anti-human CD45 BV650, anti-human CD41 AF700, anti-human CD34 APC, and anti-mouse CD45 BV605. $CD34^-$ fraction was used to sort $CD235^+CD71^+EGFP^+$ erythroblasts, $CD19^+EGFP^+$ B-lymphoid, $CD33^+EGFP^+$ myeloid cells, and $CD41^+EGFP^+$ megakaryocytes for bulk RNA-seq (50 or 100 cells/biological replicate). Hoechst 33342 was used as a viability marker in this panel.

Flow cytometry analysis and cell sorting were performed on BD FACS ARIA Fusion I as stated previously. $CD34^-$ lineage populations were sorted using a 130 μm nozzle and HSCs were purified using an 85 μm nozzle. The alignment of the sorter was checked by sorting single fluorescent beads into the corner wells of a 96-well PCR

plate (Thermo Scientific) and 50-100 fluorescent beads into 0.2 ml thin-walled PCR tubes. Cell sorting purity was also tested before the experiment. FACS data analysis was performed using FlowJo software.

2.10. Single-cell RNA sequencing of HSCs

As described previously, 725 unique barcodes in our lentiviral library are present at the 5' end of the EGFP transcript controlled by an SFFV promoter, and the barcode is therefore expressed as mRNA and polyadenylated. This allows the barcode transcript to be captured by the oligo-dT primer employed in the initial stage of scRNA-seq. ScRNA-seq is one of the latest approaches that enabled the profiling of single-cell transcriptomes using deep-sequencing technologies.

Single-cell libraries were generated by following the Smart-seq+ mRNA targeting protocol described by Rodriguez-Meira et al., 2019 [306]. Smart-seq+ is an optimised enzymatic version of the Smart-seq2 protocol [297], which was reported with an increased number of genes detected, reduced library bias, and allelic dropout rates [306]. A set of custom primers to amplify the barcoding sequences were included during the initial cDNA preparation and pre-amplification steps. The details of the primers used in single-cell and bulk RNA-seq is provided in Table 2.10.1).

Table 2.10.1 List of primers used in single-cell and bulk RNA sequencing

Primer ID	Primer sequence (5'- 3')
Reverse custom primer	GAAAGCCATACGGGAAGCAATAGC
Forward custom primer	CGGCATGGACGAGCTGTACCAG
Reverse custom primer with ISPCR adapters	AAGCAGTGGTATCAACGCAGAGTGAAA GCCATACGGGAAGCAATAGC
Forward custom primer with ISPCR adapters	AAGCAGTGGTATCAACGCAGAGTCGGC ATGGACGAGCTGTACCAG
ISPCR Primers	AAGCAGTGGTATCAACGCAGAGT
TSO	AAGCAGTGGTATCAACGCAGAGTACATr GrG+G

2.10.1. Sample collection and lysis

Single cells ($EGFP^+$ HSCs) were FACS sorted through an 85 μm nozzle directly into 96-well PCR plates (Thermo Scientific) with freshly prepared lysis buffer (4 μl/well) consisting of 0.4% Triton X-100 (Sigma Aldrich), 4 units (1 U/μl) recombinant RNase inhibitor (Clontech), 2.5 μM dNTPs (Thermo Scientific), and 2.5 μM oligo-dT_{30}VN primer (Biomers). Samples were vortexed, spun down upon collection, and kept on dry ice. They were stored at -80°C until further processing for a maximum of 4 weeks before being further processed.

2.10.2. First-strand synthesis

Frozen PCR plates were thawed, heated for 3 minutes at 72°C in a thermal cycler, and then returned to a cold block for oligo-dT hybridisation. The first-strand synthesis reaction (also referred to as reverse transcription, RT) mix was prepared as described in Table 2.10.2. 5.6 μl of the mix was added to each sample. A set of primers to target the barcode region was included in the RT step. The barcode is a short unique DNA sequence in between EGFP and the WPRE (a plasmid map can be found in Figure 2.2.1). The forward and reverse primers bind to the EGFP and

WPRE sequences, respectively. The reaction was performed in a thermal cycler with a heated lid using the RT program provided below Table 2.10.3.

Table 2.10.2 First-strand synthesis reaction mix

Supplier	Catalogue no.	Component	Volume/sample
Takara	639537	5x SmartScribe buffer	2 μl
Clontech	2313B	RNase Inhibitor (40U/μl)	0.25 μl
Qiagen	N/A	LNA-TSO (100 μM)	0.1 μl
Takara	639537	SmartScribe enzyme (100U/μl)	0.5 μl
Ambion	AM9935	Nuclease free water	2.68 μl
Biomers	N/A	Reverse custom primer (200 μM)	0.035 μl
Biomers	N/A	Forward custom primer (200 μM)	0.035 μl

Table 2.10.3 First-strand synthesis reaction conditions

Temperature	Duration	Cycles
42°C	90 min	1
50°C	2 min	10
42°C	2 min	
72°C	15 min	1
4°C	∞	∞

2.10.3. Pre-amplification

The PCR mix was prepared as described below (Table 2.10.4) and 15 μl of the mix was added to each of the RT reaction product. The reaction was incubated in a thermal cycler with a heated lid using the pre-amplification program detailed in Table 2.10.5.

Table 2.10.4 PCR reagents

Supplier	Catalogue no.	PCR	Volume/sample
Takara	638504	2x Sep Amp Buffer	12.5 μl
Biomers	-	ISPCR Primers (10 μM)	0.125 μl
Takara	638504	SeqAMP (100 U/μl)	0.5 μl
Ambion	AM9935	Nuclease free water	1.8 μl
Biomers	-	Reverse custom primer with ISPCR adaptor (20 μM)	0.035 μl
Biomers	-	Forward custom primer with ISPCR adaptor (20 μM)	0.035 μl

Table 2.10.5 Pre-amplification reaction conditions

Temperature	Duration	Cycles
98°C	3 min	1
98°C	20 sec	22
67°C	15 sec	
72°C	6 min	
72°C	5 min	1
4°C	∞	∞

2.10.4. Magnetic purification and quality control of cDNA libraries

After pre-amplification, cDNA libraries were purified using Agencourt Ampure XP (Beckman Coulter, A63882) magnetic beads according to the manufacturer's instructions, in a ratio of 1.2 to 1 with cDNA. The ratio was decided based on the manufacturer's guidelines to remove the transcripts smaller than 200 bp (the amplified barcodes are expected to be around 300 bp). Beads were equilibrated to room temperature for 30 minutes before the purification, and vortexed. 30 μl of the bead suspension was added to each well of the V-bottom assay plate (Axygen, P-96-450V-C-S), mixed with 25 μl of PCR products, and incubated for 5 minutes. After the bead-cDNA incubation, the plate was transferred to the magnet (Ambion, AM10027) and incubated for another 5 minutes. After this, the supernatant was discarded, and samples were washed with 200 μl of freshly prepared 80% ethanol

solution twice. The pellet was let dry for 2 minutes and then 17.5 μl of Elution buffer (Qiagen, 19086) was added to each sample. The plate was removed from the magnet, mixed by pipetting, and returned to the magnet for 2 minutes. 15 μl of the eluted products was transferred to a new clean 96-well PCR plate (Thermo Scientific, AB0800) and stored at -20°C.

After the magnetic purification, all samples were quality checked using an Agilent high-sensitivity chip (5067-4626) and an Agilent 2100 Bioanalyzer (G2938C) or using an Agilent fragment analyser high-sensitivity kit (DNF-474-500) and an Agilent fragment analyser (M5311AA) following the manufacturer's instructions. cDNA concentrations were assessed using a Qubit dsDNA HS kit (Invitrogen, Q32851) and a Qubit fluorometer (Invitrogen, Q32866).

2.10.5. Tagmentation and library preparation

A Nextera XT DNA Sample Preparation kit (Illumina, FC-131-1096) and Index kit (Illumina) was used for cDNA tagmentation and indexing. The tagmentation reaction was carried out using 0.7 ng of cDNA according to the manufacturer's instructions using one-fourth of the recommended volume. The tagmentation reaction mixture was prepared on an ice-cold PCR block (Table 2.10.6). After the addition of the reaction mixture, the tagmentation was performed at 55°C for 6 minutes in a thermal cycler. After the reaction, samples returned to a cold PCR block. 1.25 μl of neutralisation buffer was added to each sample, vortexed, spun down, and incubated for 5 minutes at room temperature. This is for stripping the Tn5 transposase from the tagmented DNA. After this, an index PCR was carried out by following the Illumina protocol that allows pooling of up to 384 cells with ligation of

dual indexes (i7 and i5) (Table 2.10.7). The index PCR program used for this is detailed in Table 2.10.8.

Table 2.10.6 Tagmentation reaction mix

Component	Volume
Tagment DNA buffer	2.5 μl
Single-cell DNA sample	1.25 μl
Amplicon tagment mix	1.25 μl

Table 2.10.7 Index PCR reagent mix

Component	Volume
Index 1 primers (i5)	1.25 μl
Index 2 primers (i7)	1.25 μl
Nextera PCR Master mix	3.75 μl
DNA Sample in each well/plate	6.25 μl

Table 2.10.8 Index PCR program

Temperature	Duration	Cycles
72°C	3 min	1
95°C	30 sec	1
95°C	10 sec	12
55°C	30 sec	
72°C	30 sec	
72°C	5 min	1
4°C	∞	∞

The tagmented libraries were bead-cleaned up after the reaction using a 1:1 ratio of AMPure XP beads (the method described in section 2.10.4), and they were resuspended in Resuspension buffer provided in the Nextera kit (Illumina, FC-131-1096). The size, quality, and concentration of the libraries were checked as described above (see section 2.10.4). Dual-indexed libraries were pooled to a final concentration of 2 nM and sequenced on an Illumina NextSeq instrument (WIMM Sequencing facility, University of Oxford) with 75-bp single-end reads using a NextSeq 500/550 High Output kit v2.5 (Illumina, 20024906).

2.11. Bulk RNA-seq of lineage populations

5 to 10 biological replicates for each of the EGFP-expressing Mk, Ery, My, and B-Ly populations (50 or 100 cells/biological replicate) were FACS sorted through a 130 μm nozzle directly into 0.2 ml thin-walled PCR tubes containing 4 μl of lysis buffer for bulk RNA-seq. Because the human Mk population in murine BM was rare and the cell sorting, therefore, took longer in comparison to the other cell types, which were typically less than a minute, 50 cells per replicate were sorted for the Mk population. This was to avoid sample degradation caused by longer sorting durations. After cDNA amplification and magnetic bead purification, each two of the 50 cells replicates were pooled together to maintain a similar library complexity with the other bulk libraries.

The lysis buffer was prepared as described in sub-section 2.10.1. Samples were vortexed and spun down upon collection, then kept in -80°C until further processing. The bulk RNA-seq libraries were prepared using the targeted Smart-seq2 protocol described by the Mead group for the targeted amplification of transcribed sequences using custom primers [305]. A pair of custom primers recognising the barcode region were included in the initial cDNA preparation and pre-amplification steps to increase the sensitivity of barcode detection (the primer sequences was provided in Table 2.10.1). cDNA libraries were prepared with SuperScript II (Invitrogen) (see Table 2.11.1) and pre-amplified using Kapa HiFi Hot Start Ready Mix (Kapa Biosystems) (see Table 2.11.2). The RT and PCR reaction conditions were detailed before in Table 2.10.3 and

Table 2.10.5, respectively. The magnetic bead purification and quality check were performed as described in sub-section 2.10.4. 0.7 ng of cDNA were used for each tagmentation reaction and the sequencing libraries were prepared and sequenced as described in section 2.10.5.

Table 2.11.1 Reverse transcription mix

Supplier	Catalogue no.	Component	Volume/sample
Invitrogen	18064-014	5x Superscript II first-strand buffer	2 μl
Invitrogen	18064-014	DTT (100 mM)	0.5 μl
Sigma	61962	Betaine (5 M)	2 μl
Sigma	M8266	$MgCl_2$ (1 M)	0.1 μl
Clontech	2313A	RNase Inhibitor (40 U/μl)	0.25 μl
Qiagen	N/A	TSO (100 μM)	0.1 μl
Biomers	N/A	Reverse custom primer (200 μM)	0.035 μl
Biomers	N/A	Forward custom primer (200 μM)	0.035 μl
Invitrogen	18064-014	Superscript II RT enzyme (200 U/μl)	0.25 μl
Ambion	AM9935	Nuclease free water	0.33 μl

Table 2.11.2 Pre-amplification mix

Supplier	Catalogue no.	Component	Volume/sample
Kapa Biosystems	KK2601	2x Kapa HiFi HS Ready Mix	12.5 μl
Biomers	N/A	ISPCR primers (10 μM)	0.125 μl
Biomers	N/A	Reverse custom primer with ISPCR adaptor (20 μM)	0.035 μl
Biomers	N/A	Forward custom primer with ISPCR adaptor (20 μM)	0.035 μl
Ambion	AM9935	Nuclease free water	2.305 μl

2.12. Bioinformatic analysis

2.12.1. Pre-processing of RNA-seq data and quality control assessment

The dual-indexed single-cell and bulk RNA-seq samples were pooled and sequenced on four lanes on an Illumina NextSeq instrument, using 75-bp single-end reads. Demultiplexed FastQ files (four FastQ files for each sample) were obtained at the end of the run. FastQ files belonging to the same sample were merged and then the Nextera adapter sequences and low quality reads from the ends of reads were removed using Trim Galore (v0.5.0) [322] in UNIX environment. Reads were aligned to the hg38 human reference genome using STAR (v2.6.1d) [323]. Unique read counts were quantified using featureCounts (subread/v1.6.2) [324]. Quality control (QC) metrics were generated using MultiQC (v0.9) [325].

2.12.2. Barcode detection and analysis

The merged FastQ files were used for the barcode detection in single-cells and bulk samples. A custom-written script in the R statistical computing environment (www.r-project.org) was used to detect the barcodes (prepared by Simon McGowan, CBRG, WIMM). All reads were searched for sequence matching the following barcode pattern: NNNACNNNGTNNNCGNNNTANNNCANNNTGNNN. At the end of the run, unique barcodes found in single-cells (HSCs) and bulk cells (their lineage output) and the number of barcode counts were written in a .csv file.

The barcode normalisations were performed in R with the help of Prof Claus Nerlov. The biological replicates of B-Ly, My, Mk, and Ery lineages were pooled: the total

barcode reads for each barcode and for each lineage were calculated. Barcodes with less than 10 total reads were filtered out. The barcode frequencies were calculated by dividing the barcode sum values to the total aligned reads of their respective lineage. This normalises the barcode counts to the sequencing depth. Next, for a reliable comparison of the lineages, the barcode detection level must be normalised, as barcodes make up different fractions of the transcriptome of different lineages. For this reason, scaling factors were generated using the multilineage clones (clones with barcode reads for all lineages, where average contribution to the four lineages was assumed to be even, as observed in mouse multilineage clones) as a reference and used to normalise the clone barcode counts across lineages. Clone sizes were calculated as the sum of the normalised barcode read counts divided by the total aligned reads for the individual lineages for the experiment in question. To classify clones, the normalised barcode reads were divided by their sum for each clone, and the resulting values were used for hierarchical clustering.

2.12.3. Hierarchical clustering

HSC clones (daughter HSCs with the same barcode ID) were hierarchically clustered based on their lineage contribution using the barcode data and functional HSC subtypes were identified. We used the dist function (R stats package) to calculate Euclidean distance and hclust function (R stats package) to use wardD2 agglomeration method for the hierarchical clustering. The hierarchical clusters were identified by cutting the dendrogram to six clusters using the cutree function (R stats package). The lineage contribution of HSC clones was plotted on a heatmap using the pheatmap package in the R environment. This will be presented in Chapter 4.

2.12.4. Processing single-cell RNA-seq data from HSCs

ScRNA-seq analysis was performed using the Seurat package (v4.0.0) in R [326]. The scRNA-seq count matrix was imported into R Studio and used to create a Seurat object using the Seurat::CreateSeuratObject function. The normalised barcode data showing the lineage contribution of every single cell added to the Seurat object as metadata. Single cells with duplicate barcode detection were filtered out.

The initial pre-filtering step was performed to remove the genes that were not expressed by any of the cells. Quality control metrics were explored using Seurat and the number of genes detected in each cell and the total number of reads detected within a cell can be found in Figure 2.12.1. After this, additional filtering was performed and cells with < 3,000 and >10,000 expressed genes were removed from the dataset (see Figure 2.12.2). Because a low nFeature_RNA might be an indication of a dead/dying cell or an empty droplet and a high nFeature_RNA may indicate doublets. Next, the data were normalised to the sequencing depth (the total number of gene counts for each cell) and multiplied with a scale factor of 10,000 and log transformed (natural log). The variable features that show high cell-to-cell variation were calculated using the Seurat::FindVariableFeatures function and 2000 features were selected by default. As a standard pre-processing step, the data were scaled using the Seurat::ScaleData function. Using the variable features identified in previous steps, linear dimensional reduction analysis was performed on the scaled data using the Seurat::RunPCA function. The appropriate number of principal components (PCs) was determined using an elbow plot that represents the

PC number on the X-axis and the standard deviation of PCs on the Y-axis (Figure 2.12.3).

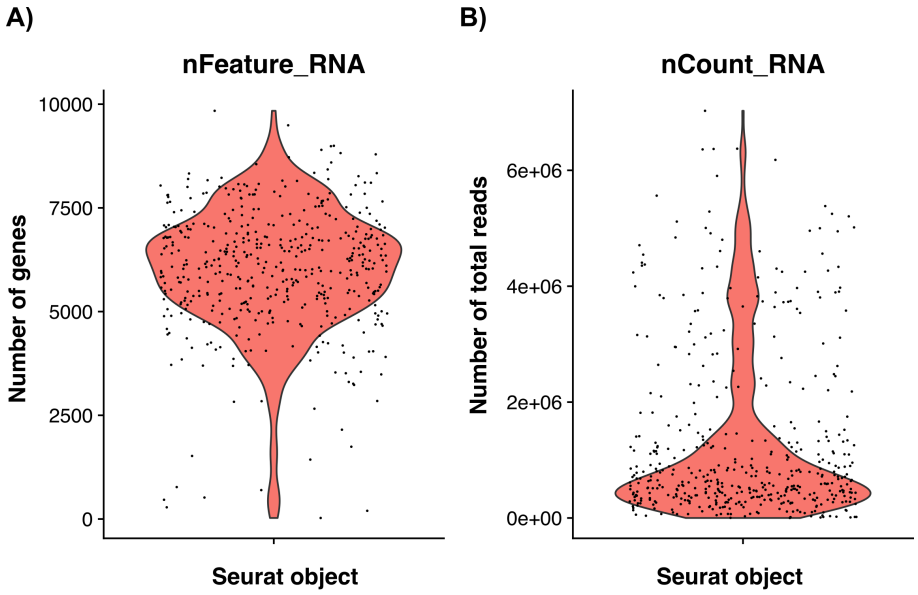


Figure 2.12.1. Quality control metrics of scRNA-seq data
A) The number of genes per cell and **B)** the total number of reads that each cell got. Each circle represents a cell.

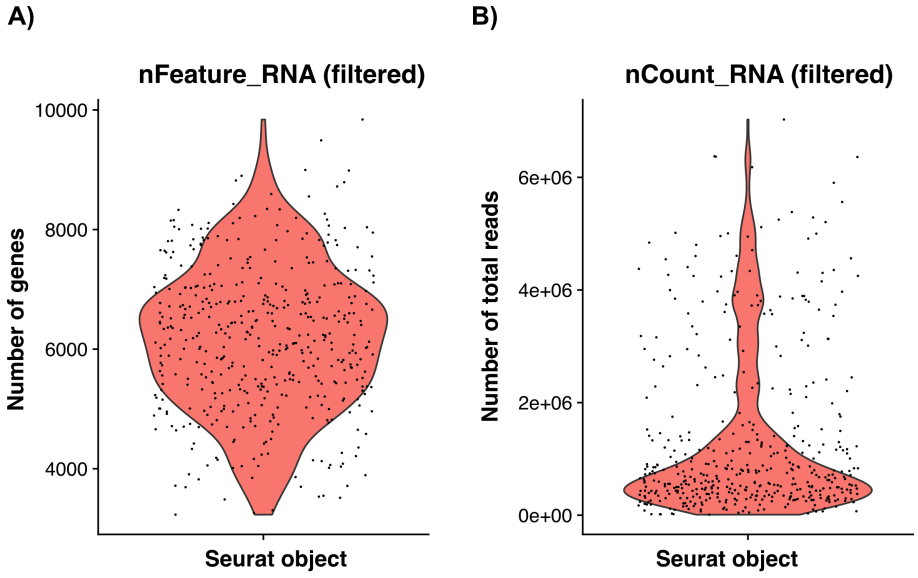


Figure 2.12.2. Filtering cells on number of expressed genes
A) Cells with gene counts less than 3,000 or over than 10,000 were filtered. **B)** nCount_RNA (number of reads per cell). Each circle represents a cell.

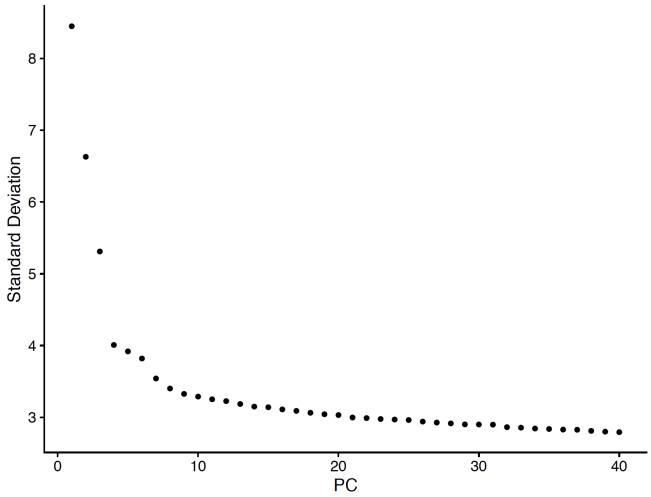


Figure 2.12.3. Linear dimensionality reduction.
Elbow plot helping to choose the optimal numbers of components for dimensionality reduction

Based on the elbow plot in Figure 2.12.3, we selected the first 10 PCs and this was passed to the UMAP. Using the Seurat::Idents function, we set the clusters to the hierarchical clusters identified using the barcode data (see section 2.12.2). Next, these HSC subtypes were statistically compared to identify differentially expressed genes (DEGs). Differential gene expression (DGE) analysis is applied as a routine analysis task to identify genes that are differentially expressed between cell populations. The DEGs were detected using the Seurat::FindAllMarkers function to compare a single HSC subtype to all the other cells (in all other clusters/subtypes) and this process was automatised for all HSC subtypes. In addition, where relevant, we compared a HSC subtype to another using the Seurat::FindMarkers function. For the DGE analysis, we used a non-parametric Wilcoxon rank-sum test, and the p-values were corrected for multiple tests using Benjamini-Hochberg method [327, 328] (employing a significance level of $p<0.05$, $logFC \geq 0.25$, and $min.pct \geq 0.25$ [a threshold for a gene to be detected at a minimum percentage of cells]). The significant DEGs per cluster were plotted on a heatmap with the Seurat::DoHeatmap function. These will be presented in Chapter 4.

2.12.5. Gene set enrichment analysis

Gene-set enrichment analysis is a standard analysis task used to compare cell populations at the level of gene-sets using prior biological knowledge or published information to facilitate the biological interpretation of DGE results [329]. Gene set enrichment analysis (GSEA) was performed using the Broad Institute's GSEA software (v4.1) [329, 330]. The normalised single-cell HSC gene expression matrix as well as human cell-type-specific gene-sets [194, 205, 279, 285, 287, 288, 317, 331] and cell cycle gene-sets [293] from the literature were used for GSEA. Mouse gene-sets [194, 279, 332] were converted into human genes using BioMart R package [333, 334]. Gene ontology analysis was performed using molecular function and cellular component GO terms and hallmark gene-sets obtained from the Broad Institute's molecular signature database [329, 335]. The number of permutation and permutation type was set to 1000 and phenotype, respectively. A complete summary of the computational analyses pipeline used in this study is provided in Figure 2.12.4. These will be presented in Chapter 4.

Computational analysis pipeline

1-Pre-processing of scRNA-seq (n=749 cells) and bulk RNA-seq fastq files and getting the feature counts table ready for the gene expression analysis
2- MultiQC quality control assessment
3- Barcode detection using a script that searches for the barcode pattern in fastq files
4- HSCs with double barcodes were filtered out (remaining number of single cells n= 472)
5- Barcode normalisations of bulk RNA-seq data (Barcodes with less than 10 reads were filtered out)
6- Barcodes in HSCs and their lineage output were linked, and a metadata file generated to be used in identification of transcriptional differences in HSCs with distinct lineage outputs
7- Visualise QC metrics and filter cells based on number of detected genes >3,000 <10,000
(remaining number of single cells n= 457)
8- Log normalization to the sequencing depth
9- Identify variable features
10- Scale expression matrix
11- Linear dimensional reduction, PCA
12- Non-linear dimensional reduction, UMAP
13- Cluster cells based on their *in vivo* lineage output using hierarchical clustering and identify HSC subtypes
14- Filter out cells that have no lineage output (remaining number of single cells n=377)
14- DGE analysis among the HSC subtypes
15- GSEA analysis to compare the HSC subtypes using prior biological knowledge

Figure 2.12.4. A summary of performed bioinformatic analysis in this study .

Chapter 3: Optimisation of methodologies for cellular barcoding of human adult BM-derived HSCs

3.1. Introduction

For decades, it has been assumed that the HSC pool is homogeneous, and a pure population of HSCs could be isolated based on the expression of cell surface markers [170]. Commonly, the engraftment and differentiation capacity of HSCs is analysed based on the quantification of donor-derived cells in PB or BM of the host using flow cytometry [102]. However, single cell and serial transplantation studies in mice demonstrated that there is a profound degree of heterogeneity in murine HSCs with the majority of cells displaying a restricted pattern of lineage outputs [200-204]. This has challenged the traditional concept of a homogenous HSC pool where individual HSCs produce all blood lineages in equal proportions. Despite its importance and major clinical implications, heterogeneity in human HSCs has not yet been fully elucidated.

Most of the research on human HSCs relies on the xenotransplantation of UCB-derived HSCs into murine recipients [55, 98, 106-112]. The reason for this is their easy accessibility as a source of HSCs, higher proliferative capacity, and susceptibility to viral transductions as well as better repopulation ability upon transplantation into irradiated mice as compared to adult BM [112-115]. More recent studies, using UCB-derived short-term and/or long-term HSCs, have demonstrated heterogeneity in their cell cycling properties [14], serial regenerative activity [214], as well as in their lineage outputs [205]. However, human adult BM-derived HSCs remain poorly characterised. To our knowledge, there are no studies addressing heterogeneity in human BM-derived HSCs that is associated with their lineage output yet. This has particular relevance for clinical procedures involving the transplant of HSCs. Therapies must target the correct cell type: in gene therapy for

instance, long-term repopulating HSCs must be stably transduced to deliver the therapeutic gene and pre-selection of HSCs, that rapidly reconstitute platelets, granulocytes, and erythroid cells might reduce morbidity and mortality associated with high dose chemotherapy in transplantations [336]. Although UCB-derived HSCs are an alternative source to BM transplantation, it is important to note that there are functional differences between UCB and adult BM HSCs in terms of composition and properties [117] and that the adult BM remains more clinically relevant and thus also requires further studies to be fully characterised.

Considering the high level of heterogeneity identified in murine BM, it is important to understand the behaviour of individual human BM HSCs in a polyclonal environment and the number of HSC clones that contribute to haematopoiesis. This work aims to characterise human BM-derived HSC heterogeneity. It focuses on the identification of HSC subtypes with specific lineage outputs and their transcriptional differences, using cellular barcoding in conjunction with single-cell transcriptome sequencing. This Results chapter describes and presents the optimisation of methodologies crucial to achieve this research aim: first, transduction culture conditions, transduction duration, and the viral Multiplicity of Infection (MOI) for lentiviral barcoding of human BM-derived HSCs were optimised. Transduced cells were then transplanted into NBSGW, non-irradiated immunodeficient mice with loss-of-function *Kit* receptor mutations. Following, at the end of the experiments, barcoded HSCs and their mature cell output (B-Ly, My, Mk, and Ery) were purified by FACS and sequencing libraries were generated using Smart-seq2 based protocols. For a successful cDNA library preparation of mature cell output of HSCs,

RNase inhibitor concentration, FACS sorting duration, and different diameter of FACS nozzles were tested.

3.2. Results

The results in this chapter will be presented in two parts: The first section describes establishment of the protocols for the lentiviral barcoding of human HSCs and their engraftment in murine xenografts. The following one focuses on optimisation of conditions for successful RNA-seq library preparations using targeted Smart-seq2 to analyse the barcodes in the HSC lineage output, i.e., B-Ly, Ery, Mk, and My.

3.2.1. Optimisation of lentiviral transduction conditions and the engraftment of human adult BM-derived HSCs in murine BM

Human HSCs ($Lin^-CD34^+CD38^-CD90^+CD45RA^-$) were purified from healthy, young human BM samples (donors aged between 20 to 30 years) as described in the Materials and Methods chapter, section 2.3, 2.4, 2.5. A representative FACS sorting plot is given in Figure 3.2.1. HSCs were genetically labelled with short unique DNA barcodes (33-bp) encoded by lentiviruses (total of 725 different barcodes) [274] in order to perform detailed clonal tracking of human HSCs and their mature cell output (see Materials and Methods chapter, section 2.6 for lentiviral transduction). The barcode is present at the 3' end of the EGFP transcript controlled by an SFFV promoter, and thus the barcode is therefore expressed as mRNA and polyadenylated, allowing the barcode transcript to be captured by the oligo-dT primer employed in the initial stage of scRNA-seq. Consequently, barcode expression can be straightforwardly determined by EGFP fluorescence in live cells.

First, to achieve unique lentiviral barcoding, transduction conditions were optimised for adult BM-derived HSCs (Figure 3.2.2A). Purified HSCs were transduced with the concentrated EGFP-expressing pe-GZ2-linkerBC322 lentiviral barcoding library overnight in one well of 96-well flat bottom plate in StemSpan serum-free cell culture media supplemented with SCF, TPO, FL, and IL-6 human cytokines [55]. In the earlier experiments performed (Exp 1-3), HSCs were pre-stimulated overnight in StemSpan media supplemented with cytokines prior to transduction as pre-stimulation leads to increased cell cycling and improved transduction. While lentiviral vectors can infect quiescent cells, the efficacy is far higher in cycling cells [162]. Different MOIs and transduction durations were used to optimise the transduction efficiency (MOI represents the number of infectious lentiviral particles per cell). The number of cells transduced with lentiviruses and transplanted into each mouse varied in each experiment due to the variation in the number of HSCs purified from each donor BM sample (Figure 3.2.2 A). After transduction, the virus was washed away in fresh media to minimise the risk of viral carryover and the cells were resuspended in IMDM media with 1% FBS (30 μl/mouse) for transplantations. Cells were shared between female 8 to 10 weeks old NBSGW (NOD.Cg-Kit^{W-41J}Tyr^{+}Prkdc^{scid}Il2rg^{tm1Wjl}/ThomJ) immunodeficient murine recipients and transplanted into the left tibia under general anaesthesia. A small sample was maintained *in vitro* in order to determine gene transfer efficiencies by EGFP fluorescence 72 hrs post-transduction (Figure 3.2.2 B). Because at the time of the transplantation (~18 hrs post-transduction), only a few of the transduced cells would express EGFP due to the dynamics of lentiviral integration/expression. Barcoding protocols aim for low transduction efficiency to decrease the likelihood of multiple integrations within each cell (as viral infection follows the Poisson distribution);

therefore, we here aimed for a maximum of ~25% of transduction efficiency [106, 249]. The MOI was slightly adjusted in some of the experiments to increase the transduction efficiency, hence, to be able to purify enough barcoded cells at the end of the experiments (Figure 3.2.2A). Although the cells were transduced at a high MOI (between 40 and 125), obtained transduction efficiency never exceeded our target (Figure 3.2.2). This may be due to the low cell cycling activity making BM-derived HSCs difficult to infect with lentiviruses. For the later experiments (Exp 4-10), the pre-stimulation step was combined with the transduction in order to shorten the total cell culture time as the media conditions used in the transduction step is the same (Figure 3.2.2A). Collectively, using different transduction conditions, an optimised protocol was developed to accommodate our lentiviral barcoding strategy.

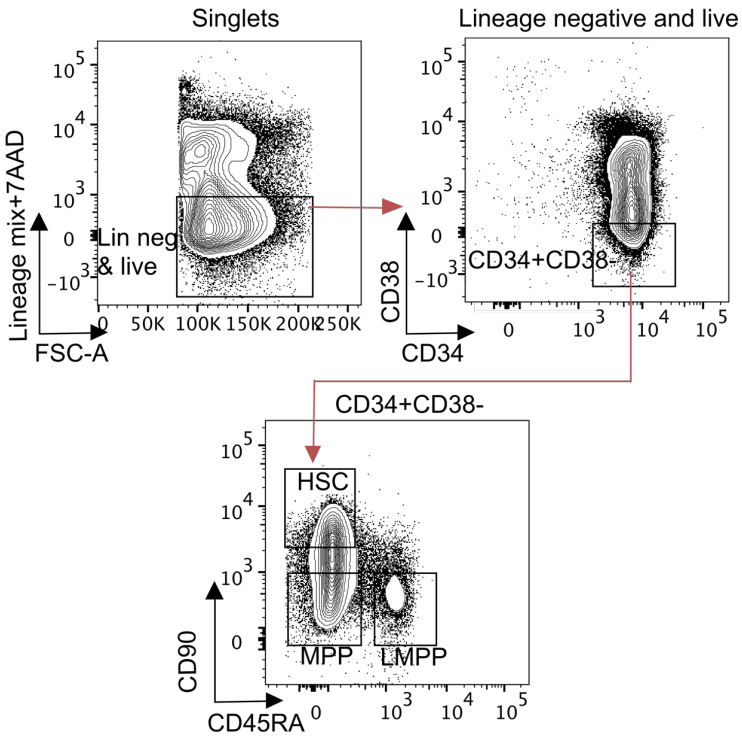


Figure 3.2.1. Representative FACS plots used to define and purify human HSCs.
Human HSCs were gated from the lineage negative and live gate. This was followed by their identification based on CD34 and CD38 expression: they are known as $CD34^{+}$ and $CD38^{-}$. HSCs were distinguished from the MPP and LMPP populations based on CD90 and CD45RA expression. Together, HSCs were defined and purified as $Lin^{-}CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}$.

A)

Experiment	Pre-stimulation	MOI	Transduction duration/Number of cells/Volume of media	Virus preparation	Number of cells/mice
1	16 hrs	50	6 hrs, 11,000 cells in a Retronectin coated 48-well plate, 300 µl volume	Prep 1 ($2x10^9$ TU/ml)	2,500 HSCs/mouse x3
2	22 hrs	40	15 hrs, 20,000 cells in a flat-bottom 96-well plate, 200 µl volume	Prep 1 ($2x10^9$ TU/ml)	2,500 HSCs/mouse x2 5,000 HSCs/mouse x2
3	22 hrs	80	15 hrs, 30,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 1 ($2x10^9$ TU/ml)	5,000 HSCs/mouse x2 7,500 HSCs/mouse x2
4	N/A	80	18 hrs, 93,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 1 ($2x10^9$ TU/ml)	22,000 HSCs&MPPs /mouse x4
5	N/A	80	18 hrs, 18,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 1 ($2x10^9$ TU/ml)	5,000 HSCs&MPPs /mouse x3
6	N/A	100	18 hrs, 18,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 2 ($7x10^8$ TU/ml)	5,000 HSCs/mouse x3
7	N/A	100	18 hrs, 29,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 2 ($7x10^8$ TU/ml)	4,000 HSCs/mouse x7
8	N/A	100	18 hrs, 21,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 2 ($7x10^8$ TU/ml)	5,000 HSCs/mouse x4
9	N/A	125	18 hrs, 74,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 2 ($7x10^8$ TU/ml)	14,000 HSCs/mouse x5
10	N/A	125	18 hrs, 67,000 cells in a flat-bottom 96-well plate, 100 µl volume	Prep 2 ($7x10^8$ TU/ml)	13,000 HSCs/mouse x5

B)

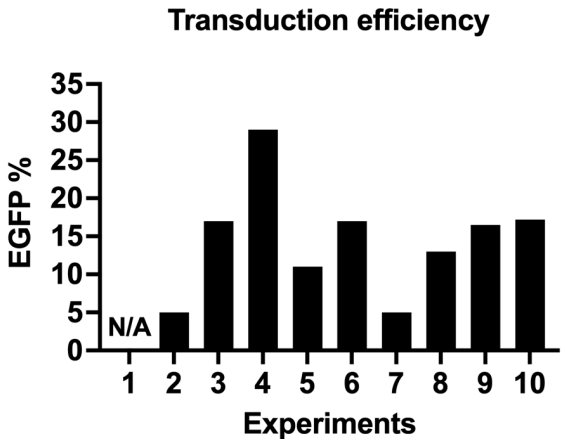


Figure 3.2.2. Summary of lentiviral barcoding of primary human HSCs.
A) Detailed summary of transduction conditions for each experiment **B)** Lentiviral barcoding efficiencies were determined by EGFP expression in viable cells three days post-transduction by FACS (N/A: in the first experiment, the cells maintained *in vitro* was not sufficient to check the transduction efficiency).

Human cells in the mouse BM were identified by human CD45 (hCD45) cell surface marker expression. A flow-cytometry gating strategy on detection of $hCD45^+$ cells in xenograft BM and the $EGFP^+$ content is given in Figure 3.2.3. In earlier experiments (labelled 1-5 in Figure 3.2.5), engraftment/chimerism levels were analysed by FACS at week 12, 16, and 22 and for the later experiments (labelled 6-10 in Figure 3.2.5) 12 weeks post-transplantation using BM aspirates obtained from left or right tibia of the xenografts (Figure 3.2.5). At the end of the experiments (for Experiment 1-5 at

week 27, Experiment 6-10 at week 16), BM cells were harvested from the leg bones (femurs, tibias, and hips), spine, and sternum, CD34 magnetic enrichment performed (both $CD34^-$ and $CD34^+$ cell suspensions were kept from this step), and the chimerism levels were analysed from the BM pool by FACS. Mice without human cell engraftment/with low engraftment (i.e., less than 1% of hCD45 in murine BM) and/or with low barcoding efficiency (determined by EGFP% in the hCD45 compartment) detected at any time point of engraftment analysis were eliminated from the study. Human cell engraftments in murine BM are detailed in Table 3.2.2.

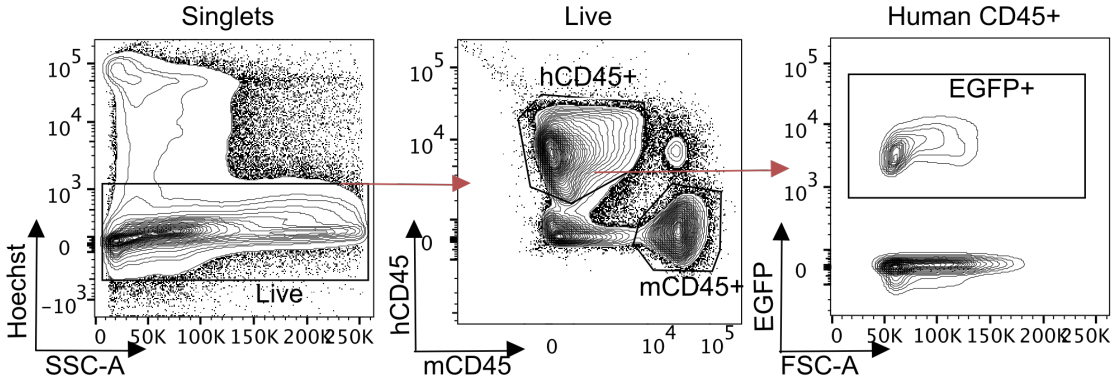


Figure 3.2.3. Human cell identification in murine BM in NBSGW recipients.
Live cells were detected using Hoechst 33342 viability dye. Human cells were distinguished from the mouse cells ($mCD45^+$) by the expression of hCD45 pan-leukocyte marker ($hCD45^+$). Barcoded human cells in mouse were identified by the expression of EGFP.

Peripheral blood of transplanted mice was also analysed for human cell engraftment at week 6, week 10 and week 16 post-transplantation using the flow-cytometry panel given above (Figure 3.2.3). However, there was either poor or no detectable human chimerism in the murine blood (hCD45% statistics in murine blood are given in Table 3.2.1). For instance, we achieved 93% human cell engraftment in the mouse BM at week 16, however, the engraftment efficiency was significantly lower in murine blood than that of in BM in the same xenograft (Figure 3.2.4 A and B). Due to the human engraftment detection inefficiency in murine peripheral blood, only the

xenografts of Experiments 1 to 5 were analysed for that parameter (Table 3.2.1) and this was discontinued for the later experiments (Experiment 6-10).

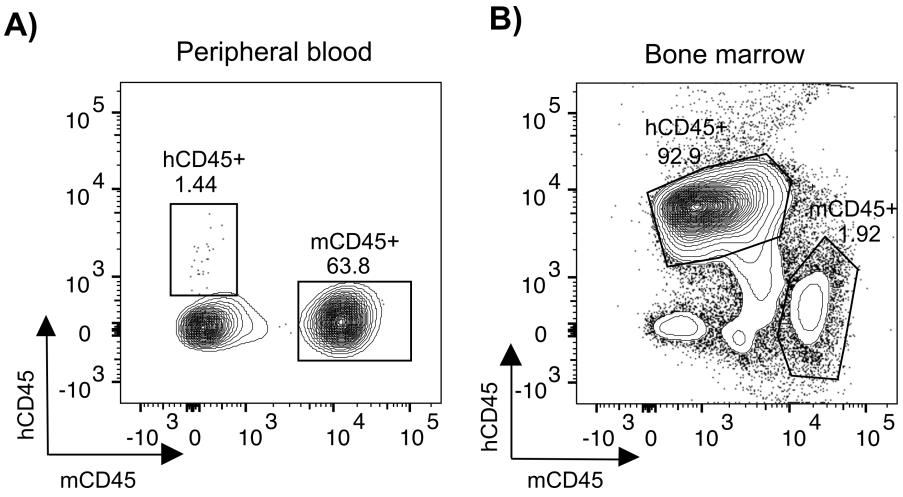


Figure 3.2.4. Engraftment of human cells in peripheral blood and bone marrow of a murine xenograft 16 weeks post-transplantation.
Representative flow cytometry plots of human cell engraftment in murine **A)** Peripheral blood and **B)** Bone marrow. Mouse and human cells were identified by the expression of mCD45 and hCD45 cell surface markers, respectively (Plots belong to the same xenograft, analysed at week 16).

Table 3.2.1. hCD45% in murine peripheral blood for Exp 1-5 analysed at different time points post-transplantations.
NA refers to the not available data time point analysis.

Mouse ID	6 weeks	10 weeks	16 weeks
1_Exp1	0	1.88	0.73
2_Exp1	0	0.69	0.16
3_Exp1	0	2.14	0
4_Exp2	0	0	0
5_Exp2	0	0	0
6_Exp2	0	0	0
7_Exp2	0	0	0
8_Exp3	0	0	1.44
9_Exp3	0	0	0.13
10_Exp3	0	0	0.29
11_Exp3	0	1	3.47
12_Exp4	0	0.11	NA
13_Exp4	0	0.35	NA
14_Exp4	0	0.13	NA
15_Exp4	0	0	NA
16_Exp4	0	0.034	NA
17_Exp5	0.32	NA	NA
18_Exp5	0.12	NA	NA
19_Exp5	0	NA	NA

In experiment 1, two mice out of three transplanted were found to be above the engraftment threshold when analysed by FACS at week 16 from the BM pool: hCD45% levels were 47% and 3% at week 16 (Figure 3.2.5 A) and the EGFP% in the hCD45 compartment were 10% and 1.5%, respectively Figure 3.2.6 A). This experiment had only one data point analysis. In experiment 2 and 3, there were five mice engrafted with human cells (only one mouse in Experiment 2 and four mice in Experiment 3) when a BM aspirate was taken from the left tibia 12 weeks post-transplantation (Figure 3.2.5 B and C). Engraftments sustained up to week 16 with observed hCD45% levels of above 70% (aspirate from the left tibia). However, when a BM aspirate was taken at week 22 from the right tibia, a reduction in the chimerism levels was observed: they decreased to 2%, 80%, 0.2%, and 40% (Figure 3.2.5 B and C). In parallel the EGFP% content of the hCD45 compartment also reduced (Figure 3.2.6 B and C). The hCD45 levels decreased even further by the end of the experiments, week 27, and the highest engraftment level observed was 30% (Figure 3.2.5 B and C). This could be due to the effect of longer *in vitro* culture time of HSCs (the pre-stimulation and transduction steps, total of ~37 hrs), performed prior to transplantation. The timing between HSCs harvest and transplantation may affect the yield or engraftment duration. Therefore, for the subsequent experiments (Experiment 4-10), overnight transduction was performed without a separate pre-stimulation step (total of ~18 hrs cell culture time) in order to reduce the overall *in vitro* culture time and to maintain stemness of the cells (Figure 3.2.2 A). In experiment 4 and 5, we used a mix of human MPPs ($Lin^-CD34^+CD38^-CD90^-CD45RA^-$) and HSCs ($Lin^-CD34^+CD38^-CD90^+CD45RA^-$) due to the low number of HSCs recovered from these donor BM samples to perform transplantations, during the FACS purification step (Figure 3.2.2 A). In experiment 4, only two mice out of

four transplanted showed human chimerism at week 12 (51% and 81%, Figure 3.2.5 D) when analysed using a BM aspirate from left tibia. However, only one of them was kept for the further analysis as the other one did not contain $EGFP^+$ cells. The engraftment level of the remaining mouse reduced dramatically from 51% to 1% when analysed at week 22 (aspirate from the right tibia) and it was found to be at 9% by week 27 (analysed from the BM pool, Figure 3.2.5 D). The EGFP level in the hCD45 compartment was also lost by week 27 (Figure 3.2.6 D). In experiment 5, the engraftment levels of three transplanted mice were below 1% threshold when analysed at week 12 using a BM aspirate (left tibia), therefore no further analysis was carried out (Figure 3.2.5 E). Low and loss of engraftment obtained from Experiment 4 and 5 might be due to the fact that the transplanted cells were mix of human MPPs and HSCs, meaning that this may have resulted in a transient engraftment of HSPCs. Due to the observation of a decrease in the human engraftment levels in murine BM after week 16 in the experiments presented above (Experiment 1-5, performed in parallel), it was decided to change the final timepoint, to ensure a sufficient number of barcoded cells could be recovered at the end of the experiments for the downstream analysis. Therefore, for the next set of experiments (Experiment 6-10), the endpoint analysis was performed at week 16 (Table 3.2.2).

In experiment 6, all three transplanted mice engrafted and displayed chimerism levels of 76%, 88%, and 6% at week 12 (aspirate from left tibia, Figure 3.2.5 F). They had EGFP levels of 1.6%, 57%, and 0.2%, respectively (Figure 3.2.6 F). Two mice with high engraftment levels and EGFP content were kept for the further analysis and their engraftment levels were found stable when analysed at week 16 from the BM pool. In experiment 8, three out of four mice were found with human

chimerism with the hCD45 levels of 63%, 54%, and 2.7% (Figure 3.2.5 H). By week 16, the analysed two mice showed a trend of decrease in the engraftment levels. In experiments 7,9, and 10 most mice displayed an increase in observed engraftment from starting points (week 12, aspirate from left tibia), generally below 40%, with many mice returning two-fold higher engraftment at week 16, by the end of experiments (Figure 3.2.5 G, I and J).

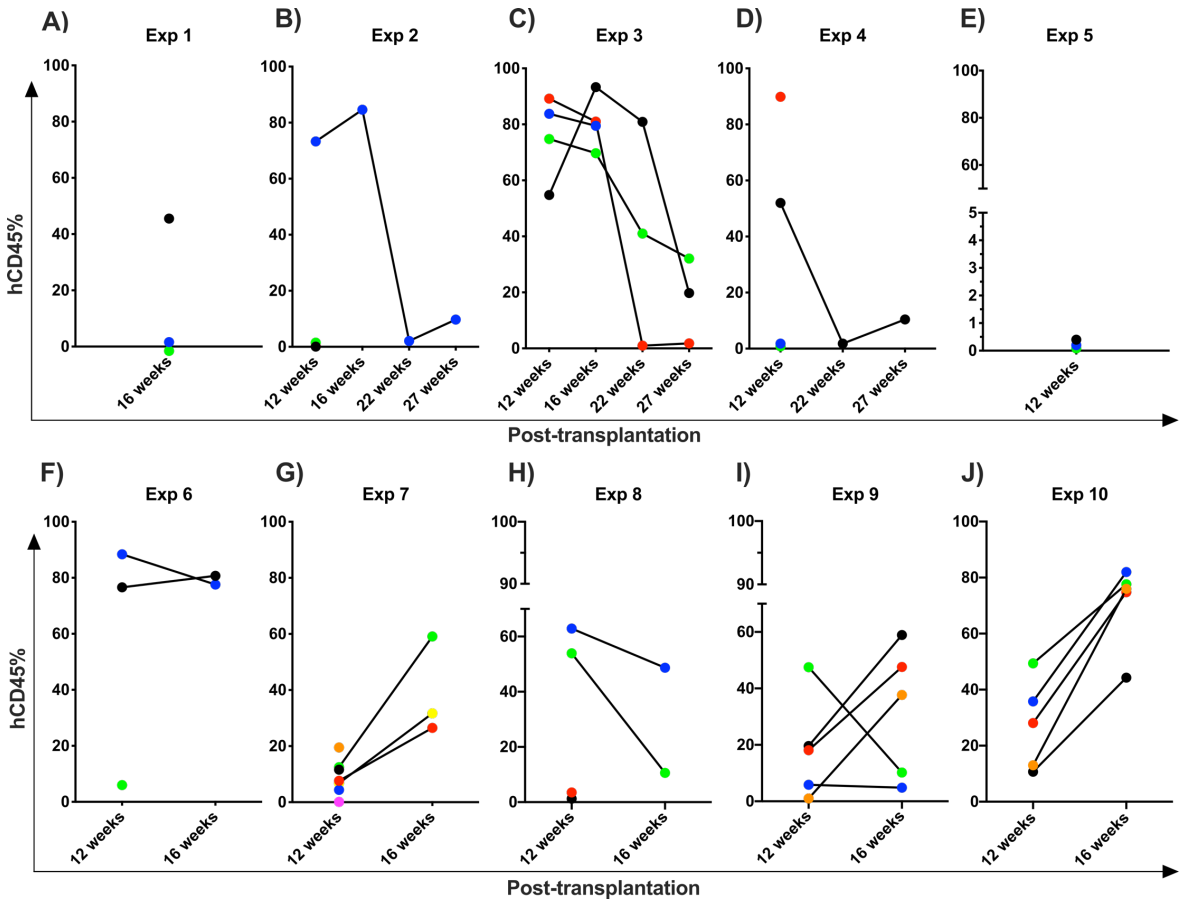


Figure 3.2.5. Human cell engraftment in the BM of NBSGW murine recipients detected by the expression of hCD45 marker by FACS.
A) Experiment 1, **B)** Experiment 2, **C)** Experiment 3, **D)** Experiment 4, **E)** Experiment 5, **F)** Experiment 6, **G)** Experiment 7, **H)** Experiment 8, **I)** Experiment 9, **J)** Experiment 10 (Each different coloured circle represents one mouse, n=3-7 for each experiment).

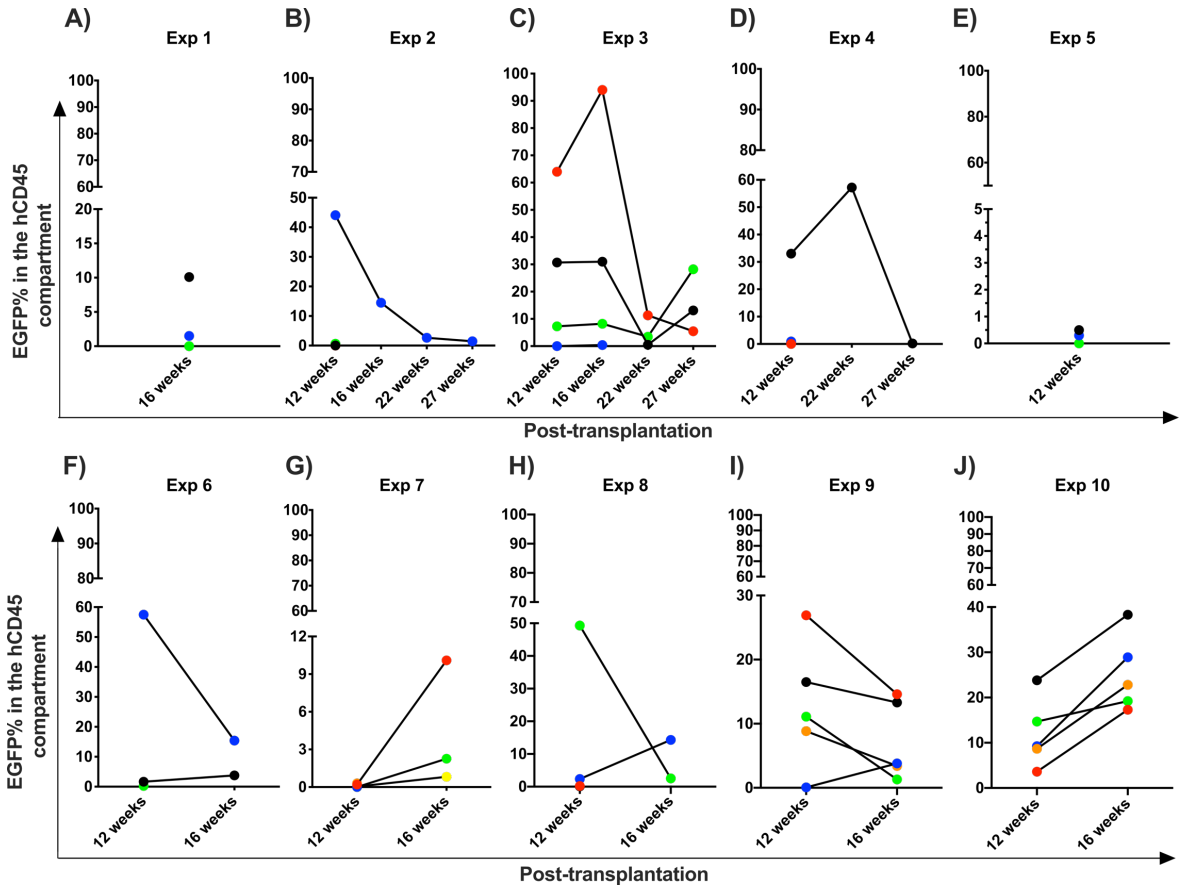


Figure 3.2.6. EGFP percentages in the human cell compartment in the BM of NBSGW murine recipients detected by FACS.
A) Experiment 1, **B)** Experiment 2, **C)** Experiment 3, **D)** Experiment 4, **E)** Experiment 5, **F)** Experiment 6, **G)** Experiment 7, **H)** Experiment 8, **I)** Experiment 9, **J)** Experiment 10 (Each different coloured circle represents one mouse and corresponds to the colours in Figure 3.2.5, n=3-7 for each experiment).

Table 3.2.2. hCD45% in murine bone marrow for Exp 1-10 analysed at different time points post-transplantations.
NA refers to the not available data time point analysis.

Mouse ID	12 weeks	16 weeks	22 weeks	27 weeks
1_Exp1	NA	47.1	NA	NA
2_Exp1	NA	3.16	NA	NA
3_Exp1	NA	0	NA	NA
4_Exp2	73.3	84.6	2.06	9
5_Exp2	0	NA	NA	NA
6_Exp2	0	NA	NA	NA
7_Exp2	0	NA	NA	NA
8_Exp3	54	92.5	80.1	19
9_Exp3	83	78.7	NA	NA
10_Exp3	88.4	80.2	0.21	0.99
11_Exp3	74	68.9	40.2	31.3
12_Exp4	51.2	NA	1	9.64
13_Exp4	66.3	NA	37.9	NA
14_Exp4	89.1	NA	NA	NA
15_Exp4	0.4	NA	NA	NA
16_Exp4	0.045	NA	NA	NA
17_Exp5	0.21	NA	NA	NA
18_Exp5	0.063	NA	NA	NA
19_Exp5	20.9	NA	NA	NA
20_Exp6	76.6	80.7	NA	NA
21_Exp6	6	NA	NA	NA
22_Exp6	88.4	77.6	NA	NA
23_Exp7	7.64	26.5	NA	NA
24_Exp7	4.39	NA	NA	NA
25_Exp7	11.6	NA	NA	NA
26_Exp7	0.11	NA	NA	NA
27_Exp7	19.5	NA	NA	NA
28_Exp7	12.5	59.1	NA	NA
29_Exp7	6.8	31.7	NA	NA
30_Exp8	2.7	NA	NA	NA
31_Exp8	62.9	48.7	NA	NA
32_Exp8	53.9	10.6	NA	NA
33_Exp8	1.36	NA	NA	NA
34_Exp9	47.5	10.2	NA	NA
35_Exp9	5.83	4.81	NA	NA
36_Exp9	18.1	47.6	NA	NA
37_Exp9	19.6	58.9	NA	NA
38_Exp9	0.59	37.7	NA	NA
39_Exp10	10.4	76	NA	NA
40_Exp10	35.8	82	NA	NA
41_Exp10	28.1	74.8	NA	NA
42_Exp10	49.4	77.6	NA	NA
43_Exp10	10.7	44.3	NA	NA

As mentioned earlier, xenotransplants were culled at the end of the experiment (for Exp 1-5 at week 27, Exp 6-10 at week 16), BM cells were harvested, and human CD34 enrichment was performed (see Materials and Methods chapter section 2.4). The CD34 enriched population was used to purify viable single $EGFP^+$ (barcoded) human HSCs for scRNA-seq. A representative FACS gating strategy on the identification of barcoded human HSCs in mouse is given in Figure 3.2.7. We used Smart-seq+ mRNA targeting method [306] (modified chemistry of Smart-seq2 protocol [297], outlined in the Materials and Methods Chapter section 2.10) that allows combining whole-transcriptome analysis with the detection of barcode transcripts from the same single-cell, using specific primers designed to amplify barcoding region. In addition, viable barcoded ($EGFP^+$) lineage output of HSCs such as Mk ($CD41^+$), My ($CD33^+$), Ery ($CD235^+CD71^+$), and B-Ly ($CD19^+$) populations were purified as bulk (50 or 100 cells/biological replicate) from the mouse BM (using the $CD34^-$ cell fraction which was kept after the CD34 enrichment) for RNA-seq using targeted Smart-seq2 [305]. A detailed gating strategy for the purification of barcoded bulk populations is given in Figure 3.2.8. Exploiting this strategy, the barcode of each HSC and their transcriptome, as well as the barcode representation in their mature cell output can be identified, followed, and linked.

Our initial barcode retrieval strategy from the bulk populations was based on DNA sequencing. However, the total number of cells that were purified for DNA isolation by FACS varied between populations. In particular, the Mk population was rare and fewer cells were able to be sorted; for this reason, the quantity of DNA recovered was below the detection level of the Qubit. Therefore, to analyse the barcodes in all four populations, we changed our barcode read out strategy to RNA-seq (after Exp

5) utilising a targeted Smart-seq2 method, to use protocols generated for the transcriptome analysis of single cells and low-cell number populations, which also allows barcode and transcriptome detection from the same sample [297]. Amplifying barcodes from the mRNA is advantageous and sensitive as there are thousands of copies in a cell. This way, we could purify 100 cells/biological replicate (collecting 5-10 of the replicates to provide enough population complexity) for each population directly into lysis buffer without any prior need for nucleic acid extraction. The experimental workflow for the final experiments (Exp 6-10) using optimised conditions and timings is given in Figure 3.2.9.

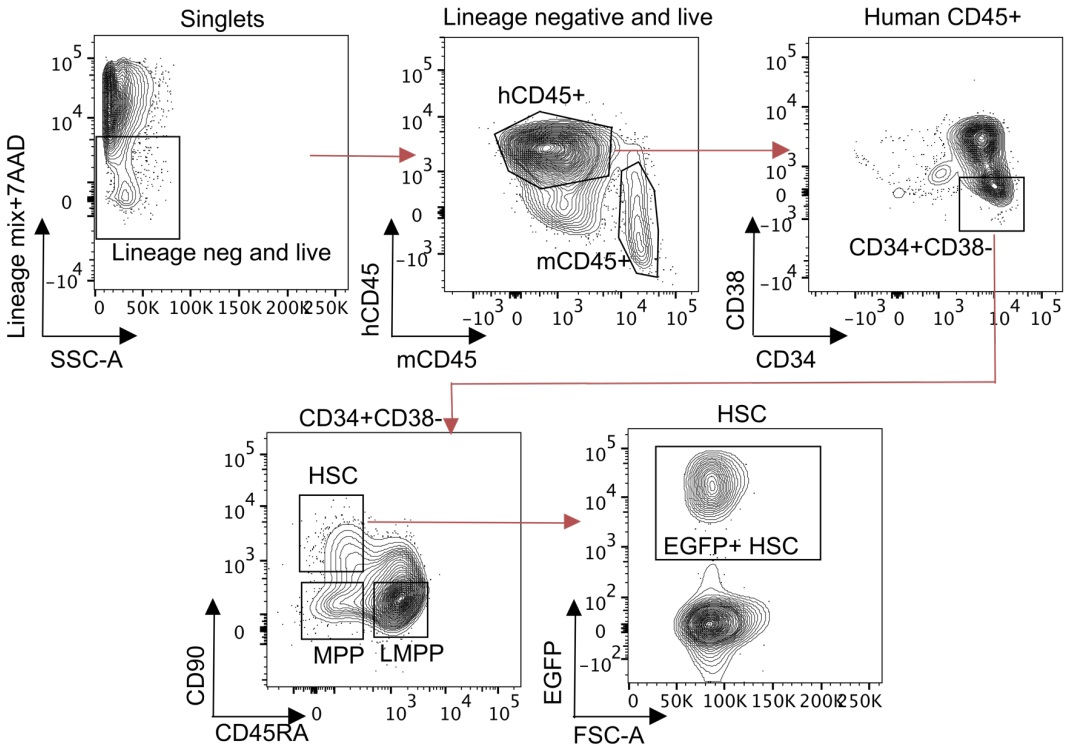


Figure 3.2.7. FACS gating strategy used to identify and sort barcoded ($EGFP^+$) human HSCs in the murine BM.
Human cells in murine BM were distinguished as $hCD45^+$ and they were gated from the lineage negative and viable cell compartment to enrich HSCs. Next, HSCs were gated based on their CD34 and CD38 expression ($CD34^+CD38^-$). HSCs were distinguished from the MPP and LMPP populations based on CD90 and CD45RA expression. Barcoded HSCs were defined as $EGFP^+$ and sorted in single-cell purification mode by FACS. Together, human barcoded HSCs were defined and purified as $Lin^-CD34^+CD38^-CD90^+CD45RA^-EGFP^+$.

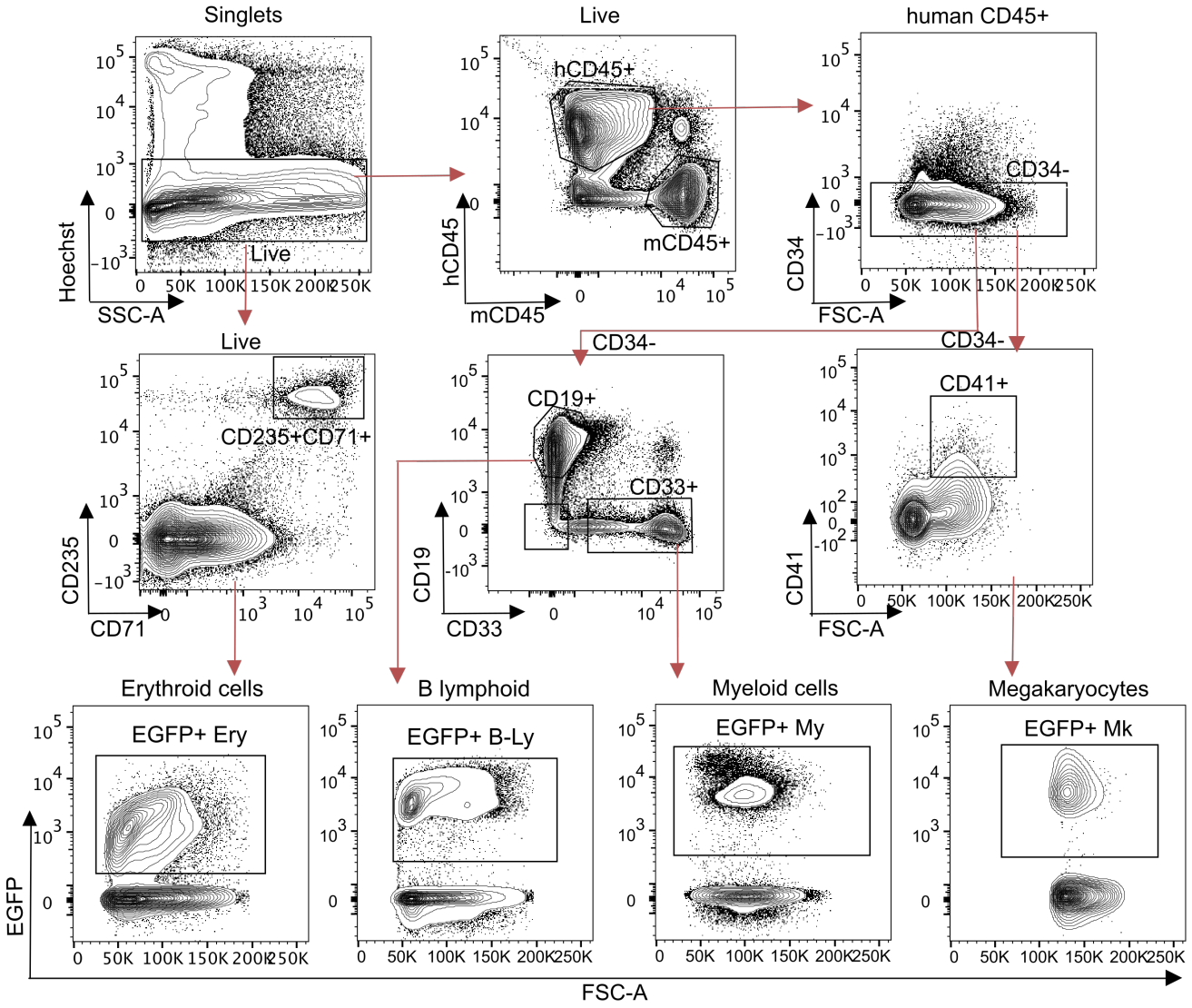

Singlets
Hoechst
SSC-A
Live
Live
hCD45+
mCD45+
hCD45
mCD45
human CD45+
CD34-
CD34
FSC-A
Live
CD235+CD71+
CD235
CD71
CD34-
CD19+
CD33+
CD19
CD33
CD34-
CD41+
CD41
FSC-A
Erythroid cells
EGFP+ Ery
B lymphoid
EGFP+ B-Ly
Myeloid cells
EGFP+ My
Megakaryocytes
EGFP+ Mk
EGFP
FSC-A

Figure 3.2.8. FACS gating strategy used for the identification and purification of barcoded (EGFP$^+$) human B-Ly, My, Ery, and Mk populations.
Viable human cells from mouse BM were identified as human CD45$^+$ (hCD45$^+$). hCD45$^+$ cells were separated according to their CD34 expression to analyse mature cell populations. The following populations were gated from hCD45$^+$CD34$^-$ population: CD19$^+$ (B-Ly), CD33$^+$ (myeloid cells; My), and CD41$^+$ (megakaryocytes; Mk). CD235$^+$CD71$^+$ Ery cells are defined as hCD45$^{low/neg}$; therefore, they were analysed directly from the viable gate. Described B-Ly, My, Mk, and Ery populations were sorted by FACS for their EGFP$^+$ content.

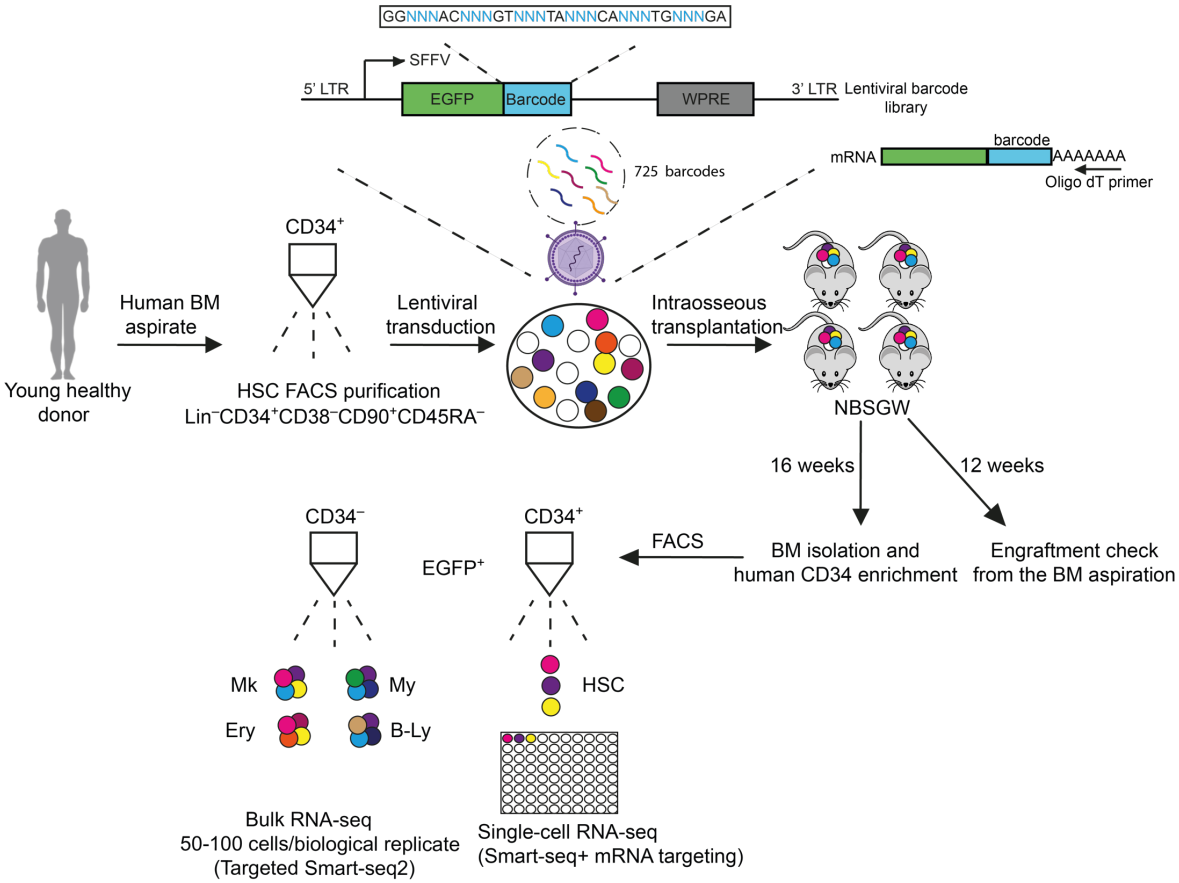


Figure 3.2.9. Established experimental workflow for final experiments using optimised conditions and timings.
Human HSCs were isolated from healthy adult BM, transduced with the lentiviral barcoding library carrying an EGFP reporter, and transplanted into NBSGW Kit mutant immunocompromised mice. Engraftment analysis was performed at week 12 from a BM aspirate and at week 16 from the total BM. The barcoded human HSCs in xenografts were FACS sorted at week 16 for scRNA-seq using Smart-seq+ mRNA targeting protocol. In addition, the barcoded lineage output of HSCs (Mk, Ery, My, and B-Ly) were FACS sorted at week 16 for bulk RNA-seq (50 or 100 cells/replicate) using targeted Smart-seq2 method. Exploiting this strategy, the barcode of each HSC, as well as the barcode representation in their mature cell output can be identified, followed and linked.

3.2.2. Optimisation of sequencing library preparations

We utilised a targeted Smart-seq2 protocol [305] in the preparation of sequencing libraries for the analysis of barcodes in the lineage output of barcoded HSCs using specific primers designed to amplify barcoding region. The primary information for retrieval from the bulk populations are the barcodes, thus allowing us to attribute lineage contributions to individual HSCs. For this, EGFP-expressing B-Ly, Ery, Mk, and My populations were purified by FACS as depicted in Figure 3.2.8. 50 or 100 cells per biological replicate were sorted for each of these populations directly into lysis buffer for bulk RNA-seq. After cDNA pre-amplification and magnetic bead purification, the overall size and quality of cDNA libraries was determined using high sensitivity DNA chips on a Bioanalyzer (described in Materials and Methods chapter, section 2.11 and 2.10.4). A good cDNA library should be composed of an average of 1500-2000 bp fragment-lengths, e.g., B-Ly cDNA libraries obtained after successful cDNA pre-amplification are given in Figure 3.2.10 A and Figure 3.2.10 B. In our experiments, the cDNA pre-amplification was successful for B-Ly and Ery populations (Figure 3.2.10 A and G), however cDNA libraries prepared using the Mk and My populations were mainly dominated by short fragments, indicative of library prepared with degraded RNA [297] (Figure 3.2.10 C and E). To improve the quality of the Mk and My libraries, we hypothesized that a slight increase in concentration of the recombinant Ribonuclease inhibitor (referred to as RNase inhibitor) in lysis buffer may overcome this issue. The addition of Ribonuclease inhibitor in lysis buffer directly prevents cells from secreted internal RNases (i.e., pancreatic ribonucleases) that may cause RNA degradation [337]. Therefore, we proposed to increase amount of RNase inhibitor in lysis buffer as some of the cell populations may have contained high levels of RNase.

The lysis buffer in Smart-seq2 protocols contains approximate of 4 units (U) of RNase inhibitor per sample in 4 μl of lysis buffer (1 U/μl) [297, 305] (one unit by definition equivalents to the amount of RNase inhibitor required to block the activity of 5 ng RNase A by 50%). To test this hypothesis, the concentration of RNase inhibitor in lysis buffer was increased to approximate of 3U /μl or 4U /μl. B-Ly were used as a control as B-Ly cells proved good substrates for successful cDNA library amplification (Figure 3.2.10A). At the end of the test experiment, however, we found that increasing the concentration of RNase inhibitor in lysis buffer did not improve the quality of the Mk and My cDNA libraries, as the signs of RNA degradation remained (Figure 3.2.10 D and Figure 3.2.10 F). This indicated that RNases were not the cause of degradation before library preparation. Next, we sought to investigate if RNA degradation occurred during the cell sorting step i.e., result of longer sorting duration or utilisation of a small FACS nozzle size. Because human Mk population in murine BM are rare and the cell sorting (50 cells/replicate) therefore took up to 10 minutes in comparison to the other cell types which were typically less than a minute. We sought to investigate whether the Mk samples were being degraded due to the increased incubation in lysis buffer while the cell sorting was performed. For this, B-Ly controls were sorted by FACS into the lysis buffer which was prepared according to the original Smart-seq2 protocol [297, 305]. These samples were kept at room temperature for 10 minutes until Mk cells were sorted. After cDNA pre-amplification and purification, the quality of the B-Ly cDNA libraries was checked using a high sensitivity Bioanalyzer chip. Despite the longer cell sorting duration, the B-Ly libraries still yielded good quality cDNA preparation. This demonstrated that the RNA degradation in Mk samples was not due to the longer cell sorting durations. The sorting duration may not be the cause of the degradation

in My cDNA libraries, because purification of each of the replicates for this population usually took maximum of a minute. After this, different diameters of FACS nozzle (that the cells were sorted through) were tested. That is an important consideration due to the reasons relating to pressure. Utilization of smaller nozzle sizes to sort larger cells applies high pressure and intense shear forces that results in inappropriate deposition of the cells into the collection tubes, more importantly, causes cell stress and damage on cell membrane, hence decreasing viability. The size of the FACS nozzle that we initially used to sort the HSCs and their lineage output (B-Ly, Ery, My, and Mk) was 85 µm in diameter (Figure 3.2.10). The estimated cell sizes are 6 μm for human BM HSPCs [338, 339], 7 to 10 μm for B-Ly [175, 181], 10 to 15 μm for Ery [187], 15 to 20 μm for My [175, 176], and 50 to 100 μm in diameter for Mk cells [184]. The Mk population displays the largest cell diameter in the BM due to the multiple rounds of DNA replications without cell division (endomitosis) [183]. Therefore, to investigate whether shear forces damaged the Mk and My cells as they were deposited to the collection tubes and this could be improved using a larger diameter FACS nozzle, we sorted the Mk and My populations as well as B-Ly controls through a 130 µm diameter nozzle. Using a larger diameter FACS nozzle was also suggested by colleagues within the institute who have encountered similar issues with the Mk cells (Bethan Psaila and Rong Li, personal communications). At the end of the test experiment, we found that the Mk libraries were improved and resulted in better quality library preparations (average peak sizes of 1500-2000 bp, Figure 3.2.11 D), suggesting using a larger diameter FACS nozzle for the purification of the Mk cells increased cell viability. However, cDNA libraries of the My population still yielded a library displaying a wide distribution of short fragments, suggesting the use of the 130 µm nozzle did not

improve those libraries (Figure 3.2.11 F). Despite the various tested conditions, it is not clear why My cDNA libraries were in poor quality, displaying signs of RNA degradation. However, this may be due to the nature of the myeloid cells i.e., high enzyme content due to their phagocytic role [340]. Although the quality of My cDNA libraries was not the optimal, since we are mainly interested in the barcodes, those libraries could potentially still yield enough information using targeted amplification of the barcode sequence.

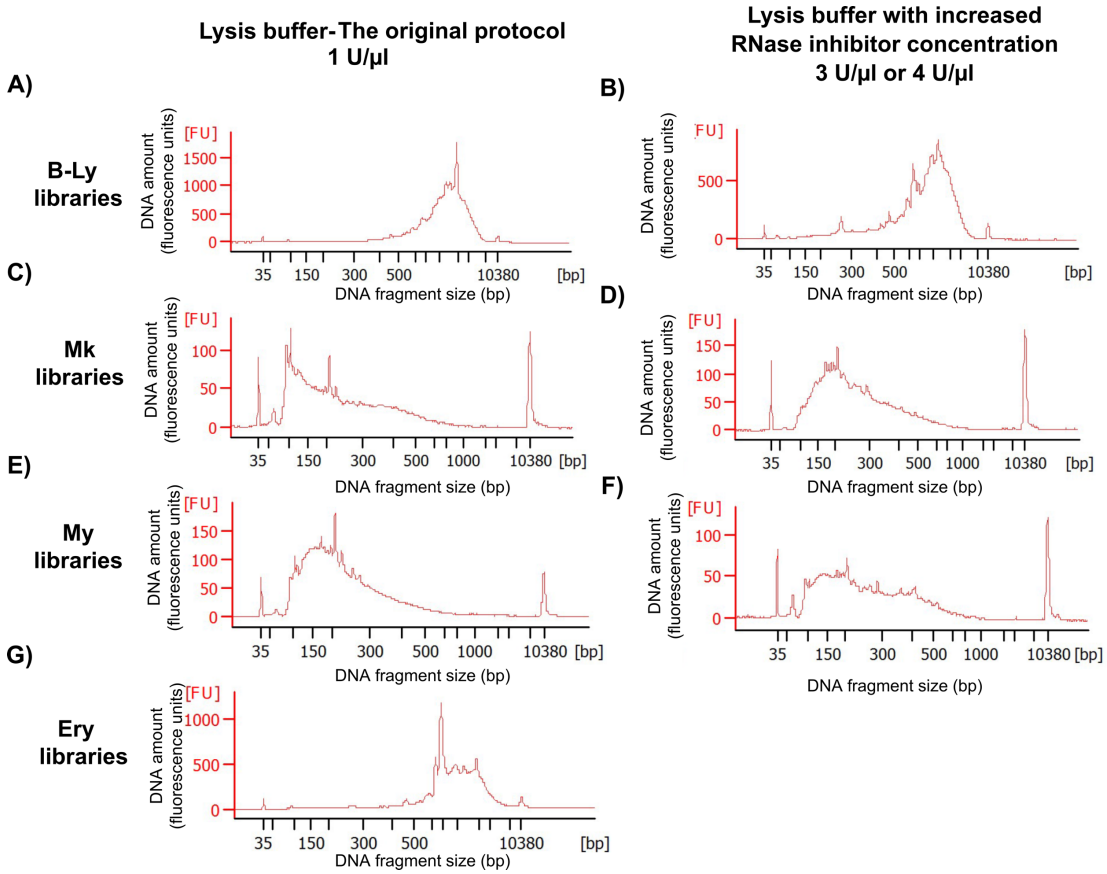


Figure 3.2.10. Bioanalyzer electropherograms of cDNA libraries: the optimisation of RNase inhibitor concentration in lysis buffer.
Representative cDNA library bioanalyzer traces for **A)** B-lymphoid, **C)** Megakaryocyte, **E)** Myeloid cell, and **G)** Erythroid samples that were FACS sorted directly into lysis buffer that were prepared according to the original Smart-seq2 protocol (RNase inhibitor in lysis buffer was 1U /µl). Representative cDNA library bioanalyzer traces for **B)** B-lymphoid, **D)** Megakaryocyte, and **F)** Myeloid cell samples that were FACS sorted directly into lysis buffer with increased RNase inhibitor condition (RNase inhibitor in lysis buffer was 3U /µl or 4U /µl). cDNA libraries prepared using B-Ly and Ery samples show examples of good-quality library with an average peak at a fragment length of 1500-2000 bp **A), B)** and **G)**. The patterns seen in the Mk and My libraries, in both conditions, were the signs of RNA degradation **C)**, **D)**, **E)**, and **F)**.

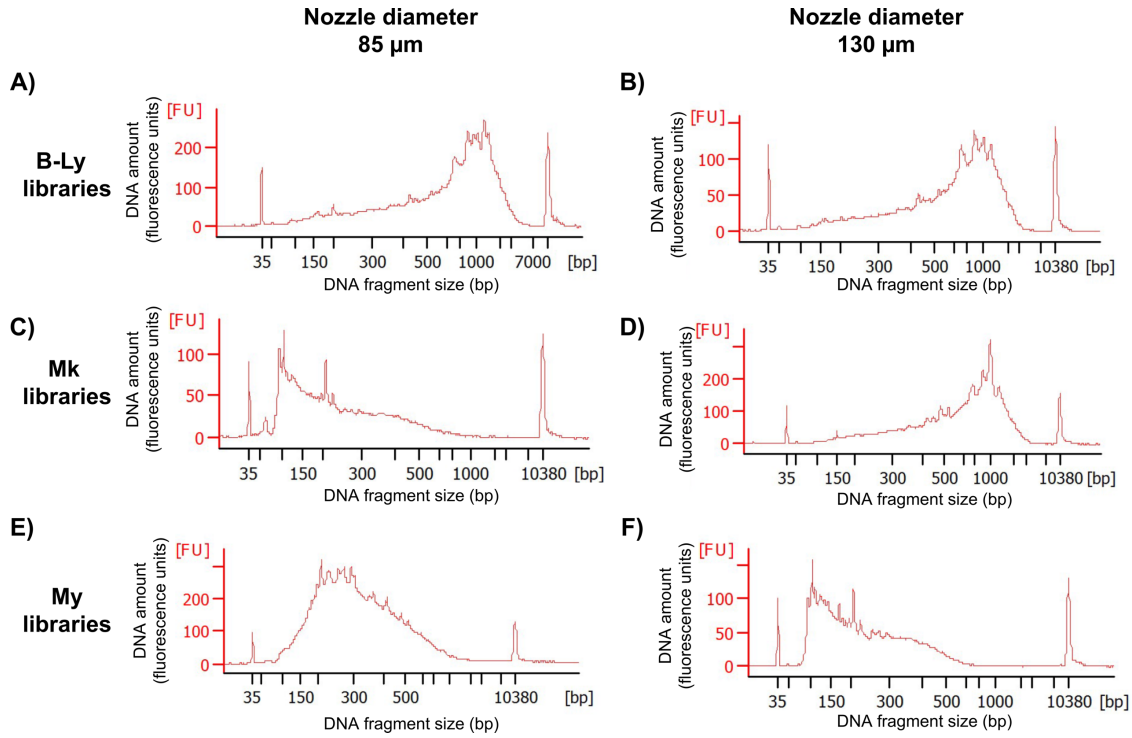


Figure 3.2.11. Bioanalyzer electropherograms of cDNA libraries: using different diameters of FACS nozzles.
Representative cDNA library traces for **A)** B-lymphoid **C)** Megakaryocyte, and **E)** Myeloid cell samples that were purified by FACS through an 85 µm diameter nozzle. Representative cDNA library traces for **B)** B-lymphoid, **D)** Megakaryocyte, and **F)** Myeloid cell samples that were purified by FACS through a 130 µm diameter nozzle. The B-lymphoid population was used as a control to test the different diameters of FACS nozzle, which still yielded a good-quality library in both conditions **A)** and **B)**. The patterns of RNA degradation in the Mk libraries, purified through an 85 µm nozzle **C).** Good-quality cDNA library preparation in the Mk libraries purified through a 130 µm nozzle **D)**. Poor-quality My libraries in both conditions **E)**, and **F)**.

3.3. Discussion

To study the heterogeneity in human adult HSCs, we combined cellular barcoding and single-cell transcriptomic profiling (the latter will be discussed in detail in Chapter 4). All experiments were performed using fresh human BM samples. This project presented several challenges: first, for each experiment, HSCs were isolated by FACS based on their immunophenotype: these were $Lin^{-}CD34^{+}CD38^{-}CD90^{+}CD45RA^{-}$ (Figure 3.2.1). This was one of the challenges of the project, because HSCs are rare cells, and they are estimated as 1 in 10,000 BM cells [341].

The number of HSCs purified and transduced with the lentiviral barcoding library varied, between 10,000 and 75,000 out of hundreds of millions of BM cells in each donor BM sample (Figure 3.2.2). In addition, these samples are costly, as it requires donors to go under surgery to donate their BM. Therefore, most of the research on human HSCs relies on the enrichment or purification of $CD34^+$ cells from UCB or mobilised PB, and their subsequent *in vitro* culture and/or xenotransplantation into irradiated immunocompromised hosts [55, 98, 106-112]. However, there are functional differences between the different sources of HSCs in terms of composition, number, and properties: as mentioned before, UCB is more widely utilised due to its easy accessibility as a source of HSCs, higher proliferative capacity, and susceptibility to viral transductions as alongside a better repopulation ability upon transplantation into irradiated mice when compared to adult BM [118, 147, 153]. Most research using HSCs and clinical gene therapy protocols also use all $CD34^+$ cells rather than highly purified HSCs, this gives the advantage of a larger starting number of cells with which to work with [58, 98, 109, 251]. However, $CD34^+$ cells comprise not only stem cells, but also the progenitor cells, which lack self-renewal ability and long-term engraftment potential. Therefore, the source and immunophenotypic definition of human HSCs is crucial for the downstream functional and molecular analysis which might have important implications on a design of experimental conditions for *in vitro* culture, transduction, xenotransplantation, and experiment duration.

Secondly, this work required optimisation to improve lentiviral transduction of human primary BM-derived HSCs by minimising the negative impact of *in vitro* culture to ensure long-term reconstitution, and to avoid multiple vector integrations per cell for

unique barcoding. Multiple integrations might not be a big of an issue as we are performing scRNA-seq of HSCs, allowing detection of barcodes in every single cell. Our transduction protocol for earlier experiments (Exp 1-3, up to 37 hrs of *in vitro* culture time) and the cytokine concentrations in cell media was very similar to one used by the Gentner group for lentiviral-based gene therapy using mobilised PB, UCB, and BM HSPCs [55]. HSC culture *in vitro,* particularly minimising differentiation, has been a focus of intense research due to its utilisation in a variety biological/biomedical and clinical applications [55]. This is because HSCs are in a quiescent state in the BM and outside the natural environment, in particular *in vitro*, they rapidly proliferate and differentiate. As mentioned in Chapter 1, many transduction protocols, particularly using retroviral vectors to transfer a gene of interest, require several days of culture to recruit quiescent HSCs into the cell cycle [60, 146]. Although cytokine pre-stimulation enhances transduction efficiency through increasing cell proliferation, a longer *in vitro* culture with cytokines decreases their engraftment ability due to impaired homing and migration [336, 342]. This suggests that transduction efficiency and reconstitution ability of the cells must be balanced. Therefore, in order to maintain the biological properties of stem cells, for the following experiments (Exp 4-10), we reduced the *in vitro* cell culture time to a total of 18 hrs (overnight incubation) (Figure 3.2.2) and optimised a minimal *in vitro* manipulation lentiviral barcoding protocol of highly purified human BM-derived HSCs in a serum-free medium supplemented with human cytokines. HSCs were transduced in one well of a 96-well cell culture plate (100 μl cell culture medium/well) to improve the transduction efficiency by minimising the cell medium volume to increase the cell density and virus concentration. For some of the experiments, we increased the MOI (Figure 3.2.2 A) to be able to recover enough

barcoded cells at the end of the experiments (Figure 3.2.6). Using these optimised transduction conditions, we obtained sufficient level of barcoding in human HSCs, up to 25% barcoding efficiency (Figure 3.2.2), that could stably engraft in humanised murine xenografts and repopulate up to 4 months post-transplantation (Figure 3.2.5).

The choice of the specific murine model was critical for this study towards characterising human HSC heterogeneity because it affects the engraftment levels of the transplanted HSCs and the type of the cells that can be successfully engrafted. To date, various mouse xenograft models have been used to study the properties of human HSCs and their engraftment ability, as this for ethical reasons cannot be tested in humans. Most of these murine models rely on irradiation prior to transplantation to deplete the murine HSCs and open up niche space for the human cells (NSG mice remain as the most popular strain due to its high engraftment efficiency) [105, 110, 129, 130]. Bearing in mind that haematopoiesis is regulated by the BM niche and irradiation causes systemic side effects, engraftment without prior irradiation requirement is clearly advantageous. Therefore, we used NBSGW (*Kit* mutant) non-irradiated mouse model [148] for *in vivo* lineage tracing of human HSCs and their mature cell output under more physiological conditions. Due to the *Kit* mutation, as c-Kit is essential to maintain murine HSCs, NBSGW mice provide a competitive advantage for human cells to engraft in murine BM with no irradiation [148]. These mice also support improved differentiation of myeloid and erythroid lineages as compared to the irradiated NSG mice [107, 148]. Although the Thomson group did not study the human megakaryocyte engraftment in the NBSGW strain [148], a study from the Waskow

group using a similar *Kit* mutant strain (NSGW41), reported efficient differentiation of UCB HSPCs into megakaryocytes, erythroblasts, and myeloid cells in the murine recipients' BM after transplantations [107, 108]. By establishing a multicolour FACS panel for differentiation markers, we detected EGFP-expressing human B lymphoid, myeloid, erythroid, and megakaryocytic lineage engraftments together with the non-transduced compartments in the BM of murine NBSGW xenotransplants (Figure 3.2.8). This is in line with the reported multilineage engraftment in these mouse models [107, 148]. In addition, we set up a FACS panel, similar to the one that was used to purify human BM-derived HSCs and identified engrafting EGFP-expressing human HSCs in murine xenograft's BM (Figure 3.2.7). The detection of transgene expression in human HSCs and their multiple lineage outputs indicate that transduced HSCs can engraft, sustain themselves and repopulate the host BM, maintaining transgene expression throughout differentiation.

Although both the Thomson and Waskow groups reported human cell engraftment in murine PB (up to 70% and 80%, respectively) [107, 108, 148], in our xenografts, there was very low or no detectable chimerism observed in the blood when analysed at week 6, week 10, and week 16 post-transplantation using flow-cytometry (Figure 3.2.4). However, direct comparisons are difficult, as both previous studies transplanted unmanipulated (without any prior *in vitro* culture) UCB-derived $CD34^+$ cells, whereas we transplanted highly purified BM-derived HSCs transduced with lentiviruses *in vitro*. The route of cell transplantation to the mice might also affect the observations. Many HSC repopulation assays (including the aforementioned studies) rely on the intravenous injection of HSPCs which requires blood circulation, detection, and extravasation through BM vasculature to reach a suitable

microenvironment for engraftment, meaning that a big fraction of donor cells are lost in the circulation and trapped in non-haematopoietic organs [321, 343]. We used an intraosseous route for the HSC transplantation, direct injection of cells into the BM cavity of the tibia as this had been reported to give higher engraftment levels than the intravenous route when the number of donor cells is limiting [50, 321, 343, 344]. A few other research groups who employed intraosseous transplantation of UCB derived HSPCs into NSG or NOD/SCID mice also reported low or absence of human cell chimerism in murine PB [50, 343] and/or distal bones [50]. Futrega *et al* described that saturation of the BM of the injected leg with human cells did not lead to high levels of systemic engraftment in the distal BM or PB of mice [50]. A study from John Dick's laboratory also demonstrated that a substantial number of clones were retained within the injected leg, although many other clones were found in distal sites too [321]. Overall, poor PB engraftment in our study may be due to retention of donor cells in BM niches. Although it was not the scope of this project, to understand this further, the same experiments could be performed using an intravenous route too and, subsequently, PB chimerism levels could be compared.

BM aspiration from the xenograft enabled assessment of chimerism and contribution of transduced cells over time. Xenografts were followed up to 27-weeks post-transplantation in early experiments (Exp 1-5, performed in parallel). The endpoint analysis was then changed to week 16 for the following experiments (Exp 6-10), due to a reduction in engraftment and EGFP levels observed beyond 16 weeks in BM (Figure 3.2.5 and Figure 3.2.6). There could be several reasons for this observation: the longer HSC transduction protocol used for the Exp 1-3 could induce HSC differentiation and generate more committed repopulating cells that still

possess proliferative and differentiation potential but have lost homing or self-renewal properties [336]. Similarly, the transplantation of mix of HSCs and MPPs in Exp 4 and 5 due to the low number of cells recovered during the FACS sort, could only provide short-term engraftment, as MPPs have reduced and incomplete capacity for self-renewal [96]. Although the BM aspiration was performed using alternating mouse legs by sampling from the injected leg at week 12 and week 16, and from the uninjected leg at week 22 (due to the requirement by the Project licence to carry out animal research), the direct engraftment comparisons from different legs cannot be made as the samples were taken at different time points. Overall, the loss of engraftment by week 27 in Experiments 1-5 did not allow us to purify enough barcoded ($EGFP^+$) HSCs for the scRNA-seq experiments and their barcoded lineage outputs to be able to explore HSC heterogeneity. For the next set of experiments performed, using the optimised transduction conditions and performing the endpoint analysis at week 16, we were able to detect and purify barcoded human HSCs (Figure 3.2.7), as well as the mature cell output by FACS (Figure 3.2.8). As mentioned in Chapter 1, historically, long-term repopulation activity of HSCs has been considered as 16-24 weeks post-transplantation or beyond this period [97, 200]. The Thomson group, using this mouse strain, assessed the long-term reconstitution ability of UCB-derived HSCs up to week 16 after transplantations, and they successfully engrafted primary recipient' marrow into secondary recipients. However, the engraftment kinetics of the primary recipient was not studied post week 16 [148].

In this project, we utilised cellular barcoding to track human BM-derived HSCs and their megakaryocyte, erythroid, myeloid, and B-lymphoid *in vivo* lineage outputs to

identify lineage-biased/restricted and multilineage HSC clones that contribute to haematopoiesis. As described before, the barcode is expressed as polyadenylated mRNA under control of an SFFV promoter. Therefore, using an mRNA sequencing approach allows identifying the barcode of each HSC alongside their transcriptome, as well as the barcode representation of their mature cell output. Therefore, we use scRNA-seq of the barcoded HSCs and bulk RNA-seq of their *in vivo* cellular output using Smart-seq+ mRNA targeting and targeted Smart-seq2, respectively [305, 345]. This allows us to correlate HSC lineage output with HSC transcriptome simultaneously and to identify the transcriptional differences between HSC subsets with different functional outputs. To our knowledge, the combination of such an *in vivo* lineage tracing tool with scRNA-seq has not been applied for the characterisation of human HSC heterogeneity before. The targeted Smart-seq2 [305] and Smart-seq+ mRNA targeting [306, 346] methods were developed based on the original Smart-seq2 protocol by the Mead group for high sensitivity single cell mutation detection along with the single cell transcriptome, which could not have been detected reliably using original Smart-seq2 protocol due to the technical dropouts and stochastic nature of gene expression in single cells [297]. Using a targeted amplification of RNA sequences of interest, we can detect barcode and the transcriptome from the same single cell with improved sensitivity in comparison to the standard techniques [305, 306]. Smart-seq+ is an optimised enzymatic version of the Smart-seq2 protocol, which was reported to increase the number of genes detected, reduce library bias, and allelic dropout rates in single cells [306]. A major challenge of this project was the limited number of barcoded ($EGFP^+$) cells that could be retrieved from each xenograft. Particularly purifying rare cell types such as $EGFP^+$ HSCs for scRNA-seq and $EGFP^+$ Mk population for bulk RNA-seq have

been a challenge. As mentioned already, HSCs are extremely rare [341, 347], and the Mk population only represents ~0.01% of the total nucleated cells in human BM, therefore it is clear that the purification of xenograft derived $EGFP^+$ HSCs or Mk cells would be challenging. However, to explore the heterogeneity associated with HSC lineage output, we need a full set of lineage output samples collected from the xenografts: for this, barcoded HSCs needed to be purified as single-cells and their Mk, My, Ery, and B-Ly output as bulk (50 or 100 cells per population per biological replicate and 5-10 biological replicates to ensure sufficient power). Using the bulk population analysis, we could detect the barcodes generating the Mk, My, Ery, and B-Ly lineages and correlate these to the barcodes found in the single HSCs. Knowing each single HSC lineage output will allow us to explore the HSC transcriptome for molecular signatures that might indicate their lineage output (these will be presented in Results Chapter 4).

During the sequencing library preparation, we observed a degradation in the Mk and My cDNA libraries, however, the B-Ly and Ery libraries provided good substrates for a successful cDNA library preparation (Figure 3.2.10). To overcome the degradation observed in the Mk and My cell libraries, we tested a few conditions such as increased RNase inhibitor concentration, cell sorting duration, and different diameters of FACS nozzle (Figure 3.2.10 and Figure 3.2.11). At the end of the test experiments, using a larger diameter nozzle and subsequent library preparation according to the original protocol, good-quality Mk cDNA libraries were yielded. Mk cells are fragile, large (up to 100 μm in diameter) and polyploid cells [184], therefore using a 130 μm nozzle, reduced the shear forces on cells and possible cell damage that the cells were exposed to during the sorting. Changing the nozzle between the

HSC and bulk cell sorts is laborious and time consuming as it requires recalibration of the FACS machine but ensures less cellular degradation and improved sample quality. However, none of the tested conditions improved the My cDNA libraries. Surprisingly, the problems that we encountered during the cDNA library preparation using the Mk and My populations (following a Smart-seq2-based protocol) have not been reported in the literature before. This might be because human mature Mk and My population reconstituted in humanised xenografts have not been used in the same context and are not commonly purified to use for RNA-seq based analysis. Particularly, human megakaryocytes have been reported to be difficult to detect in murine models because of their rarity and dependency on a xenograft model itself [107, 108].

There could be many other reasons which lead to degradation of the cDNA libraries, some of which have been reported: this includes poor sample viability or poorly defined FACS viability gates, sorter alignment, reagent or sample contamination with RNases when handling tubes/plates, evaporation during the reverse transcription step, and problems during the snap-freezing or thawing the samples [297, 346]. However, these potential reasons do not seem to be the cause of degradation observed in the My cDNA libraries, as all the bulk populations were purified by FACS in parallel, snap-frozen in dry ice, and the libraries thawed and prepared simultaneously. A similar degradation during the My library preparation was also observed across our different preparations. Although, it is still not clear why degradation was observed in the My cDNA libraries, a few possibilities include the high enzyme content and phagocytic role of myeloid cells. Myeloid cells comprise granulocytes (eosinophils, basophils, and neutrophils) and monocytes that

are responsible for initiating immune responses against foreign antigens through secretion of inflammatory cytokines [173, 348]. Neutrophils are the most abundant myeloid cell type in blood [340]. Mature neutrophils have granules and secretory vesicles that accumulate variety of enzymes such as myeloperoxidase, elastase, metalloproteinases, and defensins to destroy pathogens through phagocytosis [340]. It could be hypothesised that when damage occurs on cell membrane during myeloid cell sort, consequently these granules may release the enzymes, and this may cause RNA degradation. However, on the bulk populations, we are mainly interested in detecting the barcodes rather than their whole transcriptome and, this partial degradation observed in the My cDNA libraries did still allow us to detect the barcodes (amplified barcode expected to be ~ 300 bp) by targeted amplification from the cDNA library from that lineage.

Chapter 4: Identification of human HSC fates and the underlying transcriptional differences

4.1. Introduction

The classical haematopoiesis model assumed that every cell within each isolated population had the same fate, underestimating functional heterogeneity within the identified populations. A growing body of evidence suggests that not all HSCs are alike, instead proposing that they differ in their functional and molecular properties such as self-renewal capacity [198, 199], cell cycle kinetics [14, 16], reconstitution kinetics [199], repopulating activity [198, 200], clonal lifespan [197], and lineage contributions [200-205]. The lineage commitment and self-renewal decisions are made at the single-cell level and thus all daughter HSCs from one clone follow a predetermined fate [201]. Therefore, analyses on the clonal level are required.

Earlier studies classified murine HSCs into myeloid-biased, lymphoid-biased, and balanced subtypes based on their lineage contribution [200, 201], where the measurement of platelet lineage was not possible with the mouse model used. By generating a transgenic reporter mouse model (vWF-EGFP) that allows tracking platelets together with myeloid, erythroid, and lymphoid lineages, the Nerlov and Jacobsen groups demonstrated the molecularly and functionally distinct self-renewing platelet-biased HSCs at the top of the haematopoietic hierarchical tree in mice [203]. The most recent study from the Nerlov and Jacobsen groups using Vwf-tdTomato/Gata1-EGFP double reporter mice demonstrated four distinct types of self-renewing murine multipotent HSCs that adopted restricted fates towards the generation of platelet, platelet-erythroid, platelet-erythroid-myeloid, or platelet-erythroid-myeloid-B cell lineages [202] (HSC heterogeneity described in detail in Chapter 1 section 1.5).

To date, a large body of evidence derived from studies in mice supports the idea of fate-restricted HSCs. However, little is known for potentially equivalent human HSC subpopulations. Belluschi *et al* recently demonstrated heterogeneity in lineage potential of human UCB single-cell FACS index-sorted HSC-MPPs and CD49f^{+} LT-HSCs using an optimised single cell colony-forming assay and scRNA-seq [205]. They reported two subsets in the HSC-MPP pool and LT-HSCs based on the differential expression of CLEC9A and CD34 surface markers: CLEC9A^{hi}CD34^{lo} HSC-MPP and CD49f^{+} were described as Subset 1 HSC-MPP and Subset 1 CD49f+, respectively. Subset 1 HSC-MPP and Subset 1 CD49f^{+} were enriched for single cells with My/Ery and Mk potential *in vitro* and upregulated for HSC-related and Ery/Mk transcriptional lineage priming programmes, showed slow kinetics of quiescence exit, and yielded long-term multipotency. Besides, CLEC9A^{lo}CD34^{hi} HSC-MPP and CD49f^{+} were referred to as Subset 2 HSC-MPP and Subset 2 CD49f^{+}, respectively. They consisted of single cells with My/Ly potential *in vitro* and upregulated for Ly differentiation-related gene-sets and cell cycle genes, indicating the marks of HSC activation and faster quiescence exit [205]. This work demonstrated that the lineage restriction already occurred within the CD49f^{+} HSCs [205].

Due to the increased availability of single-cell technologies in recent years, there is a wealth of scRNA-seq data that provides snapshots of HSC transcriptomes, however, these lack HSC fate information [285-287, 320]. These studies assume that cells with similar transcriptional profiles are closer to each other developmentally but the true genetic lineage relationships among the profiled populations are unknown. Analogously, most of the lineage-tracing studies lack

transcriptome information and focus on cell fate using DNA amplification of barcodes [106, 246, 248, 249]. In this study, we combined cellular barcoding and scRNA-seq to jointly read out barcodes and the transcriptomes, for the characterisation of human HSC heterogeneity in adult BM. Together, this enables us, for the first time, to determine human adult HSC subsets with specific haematopoietic fate restriction *in vivo* and identify the transcriptional differences among functionally distinct HSC subtypes.

4.2. Results

Chapter 3 presented the results of lentiviral barcoding of human adult BM-derived HSCs, their engraftment kinetics in murine xenografts (Exp 1-10, up to 7 months), and the retrieval of barcoded HSCs as well as their cellular *in vivo* output four months post-transplantation. To obtain information on HSC fate along with transcriptome in the same cell, we applied single-cell sequencing of HSC transcriptomes. This chapter contains the computational analyses of barcoded HSCs and their barcoded B-Ly, My, Mk, and Ery outputs that were successfully purified and sequenced from the xenografts of Exp 7 (one xenograft), Exp 9 (one xenograft), and Exp 10 (five xenografts) (three experiments indicate three independent human donor samples). The results in this chapter particularly focus on the classification of HSC subtypes in human adult BM based on their *in vivo* lineage output and the characterisation of transcriptional differences between these functionally distinct HSC subtypes.

4.2.1. Identification of human HSC subtypes by their fates

Human BM HSCs from three adult donors were barcoded with unique DNA sequences encoded in lentiviruses (725 unique barcodes) and transplanted into *Kit* mutant NBSGW immunodeficient mice (seven recipients in total). Barcoded HSCs and their progenies (B-Ly, My, Mk, and Ery lineages) were isolated from xenografts 16-weeks post-transplantation and processed for RNA-sequencing. Barcodes in HSCs (scRNA-seq data) and their lineage outputs (bulk RNA-seq data) were identified using a custom-written script in R that searches for the barcode pattern and combines sequences that match to the same barcode to quantify the count number of each barcode. We pooled the barcode counts of biological replicates of B-Ly, My, Ery, and Mk lineages in each experiment. Barcodes with less than 10 total reads (sum of the barcode counts in each lineage) were filtered out. The barcode counts were normalised to the sequencing depth to correct for the differences between sequencing runs and then, for a reliable comparison of the lineages, the barcode detection level was normalised, as barcodes make up different fractions of the transcriptome of different lineages (see Chapter 2 section 2.12.2 for the detailed description of the computational methods). In total, 749 single HSCs were sequenced and 277 of them were filtered out as they were detected with double barcodes (the remaining 472 cells to carry only one barcode). That might be due to the contamination in some samples during library preparation as the achieved transduction efficiencies were ~15%, highlighting a very low risk of multiple integrations in a cell (see Chapter 3 Figure 3.2.2). Moreover, the quality control metrics were explored using the Seurat package in R, and that the 457 HSCs out of 472 passed the quality controls (described in Chapter 2 subsection 2.12.4). A step-by-step computational analysis pipeline is provided in Chapter 2 Figure 2.12.4.

When a barcode was recovered from a labelled HSC, progenies derived from HSC were identified by the same barcode, thus allowing to correlate the functional lineage output to the HSC transcriptional profile (Figure 4.2.1 B). Daughter HSCs carrying the same barcode were referred to as HSC clones, meaning that each barcode corresponds to single HSC clone giving rise to mature lineages *in vivo* (Figure 4.2.1 A and B). It needs to be noted that HSCs identified by the same barcode were not the cells independently transduced with the same barcode: In experiments 9 and 10, about 70,000 human HSCs were purified by FACS and transduced with the barcode library. Here, the approximate number of barcoded cells was 10,000 due to the 15% achieved transduction efficiencies (see Chapter 3 Figure 3.2.2 for transduction efficiencies). After transductions, 70,000 cells were equally split and transplanted into five murine xenografts in each of those experiments (~14,000 cells contains a theoretical ~2000 barcoded cells). Although we transduced the cells with a relatively small barcode library that consisted of 725 unique DNA tags, the likelihood of more than one cell carrying the same barcode transplanted and successfully engrafted into the same xenograft was extremely low. At the end of the experiments (16-weeks), the number of retrieved barcoded human HSCs from each murine xenograft were well under the number of transplanted 2000 barcoded cells (Table 4.2.1), highlighting that the cells here identified by the same barcode were most likely to be the daughter HSCs rather than more than one cell transduced with the same barcode reconstituted in the same xenograft.

Table 4.2.1. The number of human HSCs and HSC clones (per donor) detected in murine xenografts 16 weeks post- transplantation

Donor	Xenograft	Number of HSCs	Number of clones
D1	Xenograft 1	38	2
D2	Xenograft 2	55	2
D3	Xenograft 3	58	7
D3	Xenograft 4	45	10
D3	Xenograft 5	105	18
D3	Xenograft 6	40	11
D3	Xenograft 7	37	7

We detected a total of 159 unique barcodes across all four analysed lineages (Figure 4.2.1 A), and 57 of them corresponded to barcodes in analysed HSCs (n=377 cells) (Figure 4.2.1 B) (Detailed statistics of the distribution of 57 HSC clones per donor and per xenograft and the number of constituting HSCs is provided in Table 4.2.1). This observation is perhaps not surprising as some HSCs might not be captured during the single-cell purification by FACS or were lost during processing. Another possibility is that some HSCs might be differentiated while generating progenitor cells and no longer exist, thus some barcodes might be present in the mature progeny but not in HSCs (e.g., lymphoid cells represent a long-lived cell population). We found that the HSC clones were diverse in their cellular *in vivo* outputs: some HSC clones gave rise to all lineages while a substantial proportion displayed restricted and biased patterns of lineage outputs towards the generation of erythroid/myeloid/megakaryocytic lineages (Figure 4.2.1 B). We also detected 29 HSC clones (n=80 cells) without contribution to any of the analysed lineages (Figure 4.2.1 C).

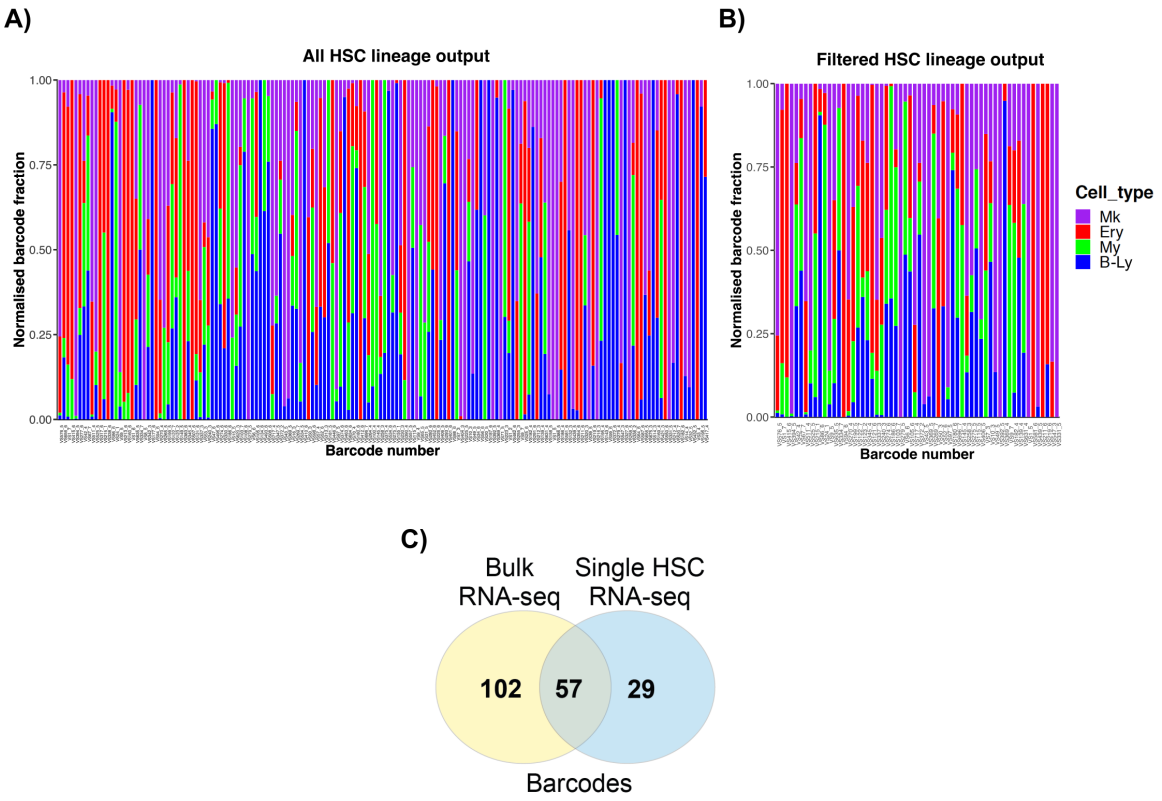


Figure 4.2.1. The cellular *in vivo* output of HSC clones with unique barcode IDs.
The X-axis shows the barcode IDs (HSC clones), and the Y-axis shows proportion of the barcode contribution to the B-Ly (blue), My (green), Mk (purple), and Ery (red) cell generation. **A)** All barcodes detected in the measured lineage output of HSCs, however some of which were not detected in the analysed HSCs (number of clones n=159), **B)** Barcodes detected in both HSCs and in their measured lineage output, allowing to link the functional output to the HSC transcriptional profile (number of clones n=57) and **C)** Venn diagram shows the number of barcodes detected in the lineage output by bulk RNA-seq (n=159) and in the analysed HSCs (n=86) by scRNA-seq. (Human donors n=3, xenograft n=7).

To identify subpopulations with distinct functional outputs within the HSC compartment, we carried out hierarchical clustering of barcoded HSC clones based on their lineage contribution, as daughter HSCs derived from a clone exhibit similar pre-determined fate [201] (the filtered HSC lineage output that was shown in Figure 4.2.1 B was used for the clustering). Using the Euclidean distance and wardD2 method, HSCs were subdivided into six clusters (Figure 4.2.2). The heatmap in Figure 4.2.2 shows that some HSC clones harbouring barcodes exclusively produced either Mk or Ery lineage without any contribution to other lineages: these

were classified as Mk-restricted (n=71 cells) and Ery-restricted subtypes (n=33 cells), respectively (orange and red blocks, Figure 4.2.2). Some other HSC clones showed biased patterns of output, predominantly contributing to the Mk and Ery lineages or Ery and My lineages, but absent or only marginally present in remaining lineages. These were assigned as MkEry-biased (n=95 cells) and EryMy-biased subtypes (n=51 cells), respectively (green and purple blocks, Figure 4.2.2). We also detected three HSC clones with lymphoid-biased output, and these were referred to as B-Ly-biased subtype (n=13 cells, yellow block, Figure 4.2.2). Moreover, another subtype was classified as Multilineage HSC clones whose barcodes were present in B-Ly and at least two other or all analysed lineages, showing a balanced pattern of output (Multilineage subtype, n=114 cells, blue block, Figure 4.2.2). A scheme illustrating the mature cell output of HSC subtypes is given in Figure 4.2.3.

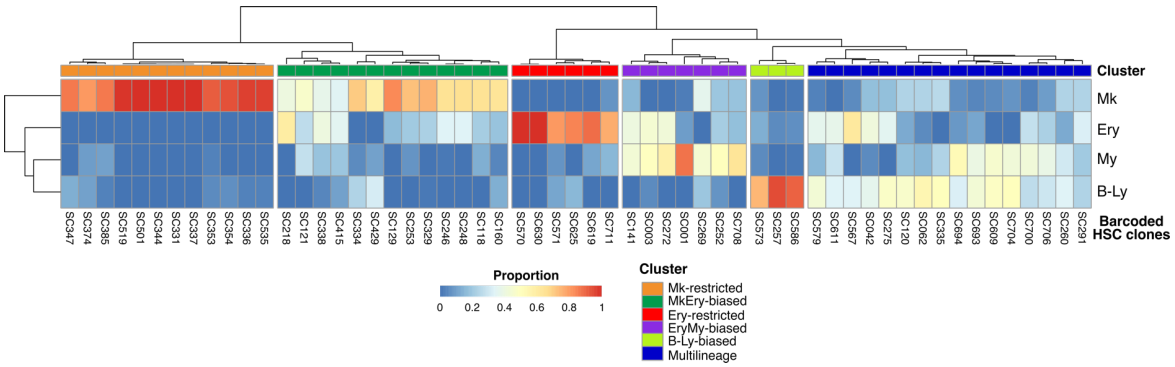


Figure 4.2.2. Hierarchical clustering of barcoded HSC clones based on their lineage output. The heatmap shows the measured cellular *in vivo* output (Mk, Ery, My, and B-Ly) of HSC clones in proportion (the same data that was shown in Figure 4.2.1 B). Daughter HSCs carrying the same barcode were referred to as HSC clones and they were hierarchically clustered using the Euclidean distance and wardD2 method (n=57 clones, 377 HSCs). Using this, phenotypically defined HSCs were subdivided into six clusters based on their functional output, and classified as Mk-restricted (orange), MkEry-biased (green), Ery-restricted (red), EryMy-biased (purple), B-Ly-biased (yellow), and Multilineage subtypes (blue) (from the left to right).

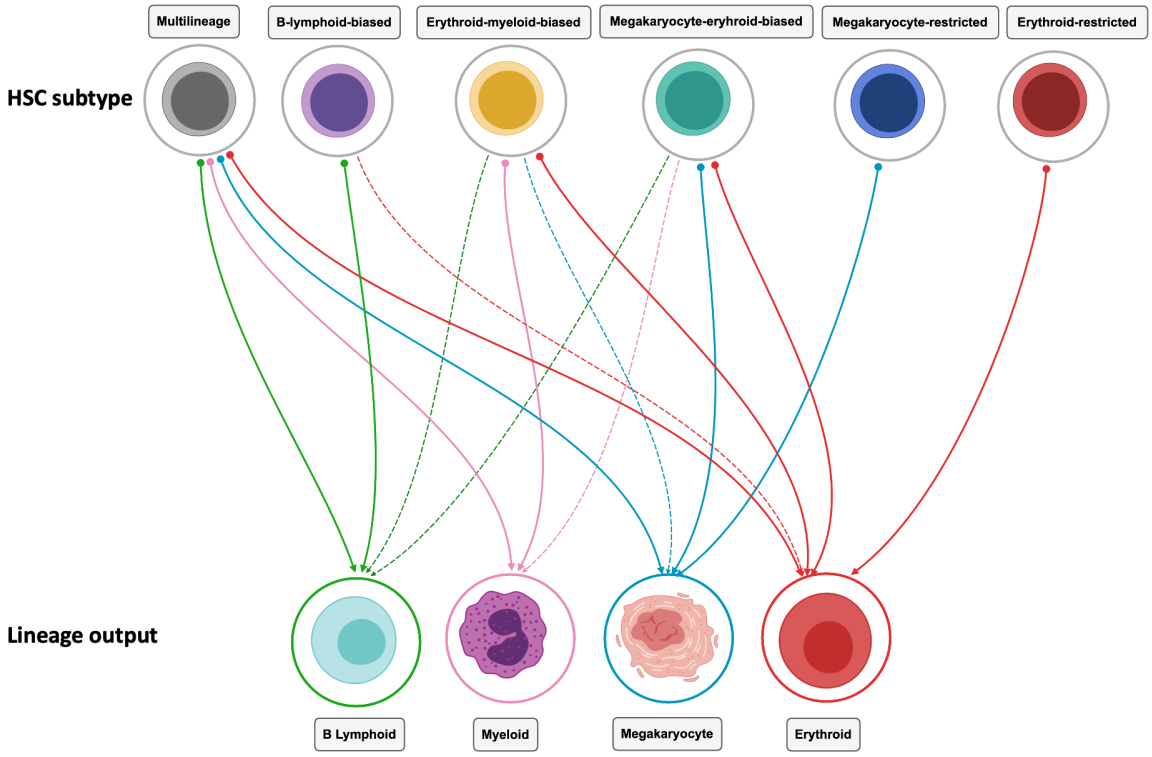


Figure 4.2.3. Fates of the 6 classes of human HSCs (as identified by hierarchical clustering) schematically illustrated.
Straight arrows depict that the lineages were produced by these HSC fate types, and the dashed arrows indicate the marginal contribution of biased subtypes to these lineages (the scheme is the illustration of HSC clone clustering, see Figure 4.2.2). Erythroid-restricted=Ery-restricted, Megakaryocyte-restricted=Mk-restricted, Megakaryocyte-erythroid-biased=MkEry-biased, Erythroid-myeloid-biased=EryMy-biased, B-lymphoid-biased=B-Ly-biased.

Next, we compared the HSC subtypes to investigate if there were any differences between their clone sizes. Clone sizes were calculated as the sum of the normalised barcode read counts divided by the total aligned reads for the individual lineages for the experiment in question. However, the mean clone sizes of HSC subtypes were similar to each other and did not show any significant difference (Wilcoxon rank-sum test, BH adjusted p-value>0.05) (Figure 4.2.4).

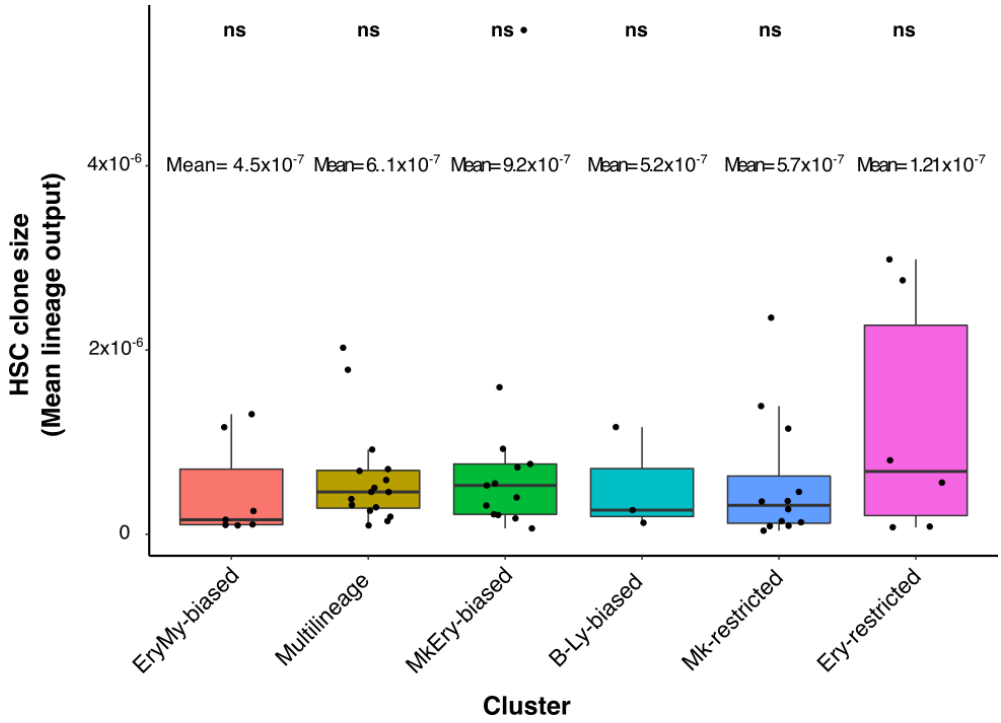


Figure 4.2.4. No significant difference between the clone sizes of the HSC subtypes. HSC subtypes were compared for their clone sizes to investigate their repopulation kinetics (ns=not significant, BH adjusted p-value >0.05, as determined by comparing the clones in one subtype to the remaining clones using Wilcoxon rank-sum test). X-axis shows the HSC subtypes, and the Y-axis shows the normalised values (normalised to the sequencing depth of the experiment in question. See Chapter 2 section 2.12.2 for the calculation)

4.2.2. Identification of transcriptional differences between human HSC subtypes

Cellular barcoding allowed us to cluster HSC subtypes based on their lineage outputs and identify heterogeneity in function. We next sought to explore if the molecular signature of HSCs could provide insight into the potential drivers of functional differences in HSCs. We first performed dimensionality reduction visualisation of HSC subtypes identified using hierarchical clustering. However, we found no separate clustering of functional HSC subtypes on low-dimensional UMAP space (Figure 4.2.5).

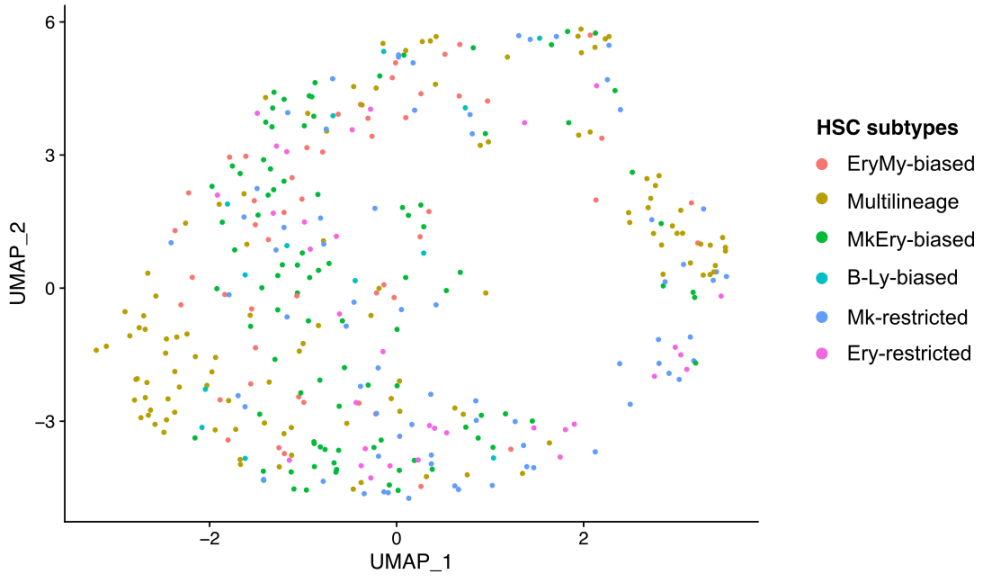


Figure 4.2.5. Dimensionality reduction visualisation of HSC subtypes on UMAP identified using the hierarchical clustering of HSC clones

Next, we carried out differential gene expression (DGE) analysis to compare HSC subtypes and reveal transcriptional differences among them. Differentially expressed genes (DEGs) were identified using the Wilcoxon rank-sum test and corrected for multiple tests using the Benjamini-Hochberg (BH) method [327, 328] (Figure 4.2.6). We detected a total of 222 genes differentially expressed between the HSC subtypes, based on the statistical comparison of one subtype to the remaining cells (BH-adjusted p-value < 0.05 was used as a significance level and logFC threshold≥0.25). The heatmap in Figure 4.2.6 illustrates the top 20 significant genes per cluster. Overall, hierarchically classified HSC subtypes based on their functional output showed distinct gene expression patterns although there was no distinct clusters on UMAP (Figure 4.2.6). This might be due the small transcriptional differences between the subtypes. Collectively, DGE analysis revealed transcriptional heterogeneity within the phenotypically defined human HSC

population as well as distinct transcriptional patterns between human HSC subtypes clustered based on their functional *in vivo* outputs.

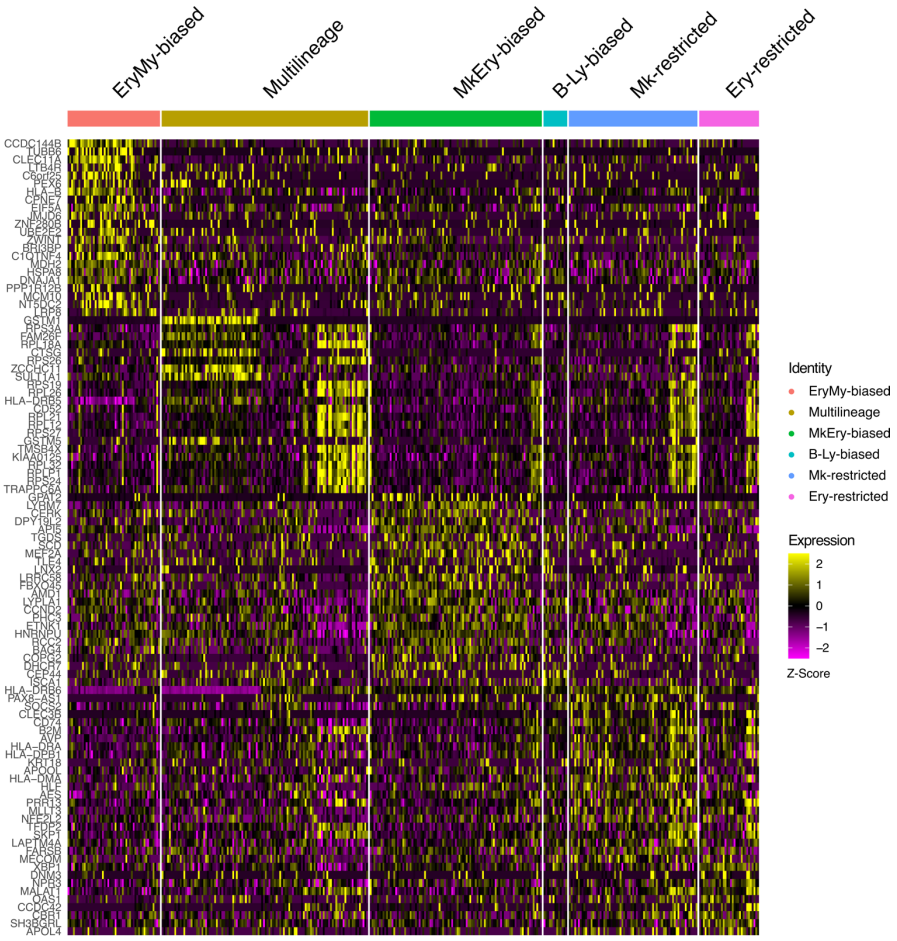


Figure 4.2.6. Differential gene expression analysis of functional human HSC subtypes revealed transcriptional heterogeneity.
A heatmap showing the top 20 genes per cluster identified by comparison of each subtype to all other cells using Wilcoxon rank-sum test (BH adjusted p-value<0.05). The heatmap displays the Z-score of log-normalised expression values. The rows show the genes, and the columns represent the cells grouped by using the hierarchical clusters identified in Figure 4.2.2.

Next, we performed Gene Set Enrichment Analysis (GSEA) [329, 330] to compare HSC subtypes using cell-type-specific gene signatures generated from human HSCs [317] with distinct lineage differentiation potential [205, 279] or repopulation kinetics [279], Mk/platelet lineage-affiliated [287, 288], and cell cycle gene signatures [293] (publicly available gene-sets). Compared to all other HSC subtypes, the Mk-restricted HSC subtype showed significant enrichment of Subset

1 CD49f$^+$ and Subset 1 HSC-MPP gene-sets (obtained from the statistical comparison of Subset 1 CD49f$^+$ to Subset 2 CD49f$^+$, and Subset1 HSC-MPP to Subset 2 HSC-MPP in [205], respectively) (Figure 4.2.7 A). Subset 2 gene-sets were downregulated in the Mk-restricted subtype and significantly enriched in the other HSC subtypes (False Discovery Rate (FDR) $\leq$0.25 for Subset 2 CD49f$^+$). In addition, the low-output HSC and Mk-biased gene-sets (obtained from the statistical comparison of low-output HSCs to high-output HSCs and Mk-biased HSCs to Multilineage HSCs in [349], respectively) were significantly upregulated in the Mk-restricted subtype (FDR$\leq$0.25) (Figure 4.2.7 A). Consistent with this, high-output HSC and S-phase cell cycle gene signatures were significantly enriched in other HSC subtypes. Moreover, two other HSC-related gene-sets were significantly upregulated in the Mk-restricted subtype (Figure 4.2.7 A). Although the Mk-restricted subtype had a positive normalised enrichment score (NES) for the Mk and platelet affiliated gene-sets; however, these were not significant (FDR>0.25) when compared to the remaining cells (Figure 4.2.7 A). This could be due to the presence of Mk-biased HSCs (e.g., MkEry-biased HSCs) on the other end of the comparison which might highly express Mk and platelet genes, too. Collectively, this GSEA indicates that the Mk-restricted HSC subtype might represent the least-proliferative subset within the HSC compartment with hallmarks for Mk-lineage priming, evidenced by the upregulation of Subset 1 gene signatures.

GSEA of Mk-restricted versus Multilineage subtypes yielded similar findings (Figure 4.2.7 B): while the Mk-restricted subtype was significantly enriched for Subset 1 CD49f$^+$ and Subset 1 HSC-MPP, the Multilineage subtype was significantly upregulated for Subset 2 CD49f$^+$ gene signatures. This indicates the activation of

Mk and Ly lineage-priming programmes in the Mk-restricted and Multilineage subtypes, respectively (due to the reported lineage differentiation potential of Subset 1 and Subset 2 cells, as described before (see section 4.1) [205]). In addition, we found significant upregulation of more HSC function related gene-sets as well as Mk and platelet differentiation-related gene-sets in the Mk-restricted subtype (Figure 4.2.7 B). Moreover, we observed significant enrichment of high-output HSC gene signatures in the Multilineage subtype. In accordance with this, low-output signatures had a significant positive enrichment in the Mk-restricted subtype (Figure 4.2.7 B). Taken together, this analysis suggests that the Mk-restricted HSC subtype might be transcriptionally primed for Mk lineage differentiation and have low repopulation kinetics, whereas the Multilineage HSCs might represent a less quiescent subset with higher repopulation activity as compared to the Mk-restricted HSCs and show active lymphoid differentiation programmes.

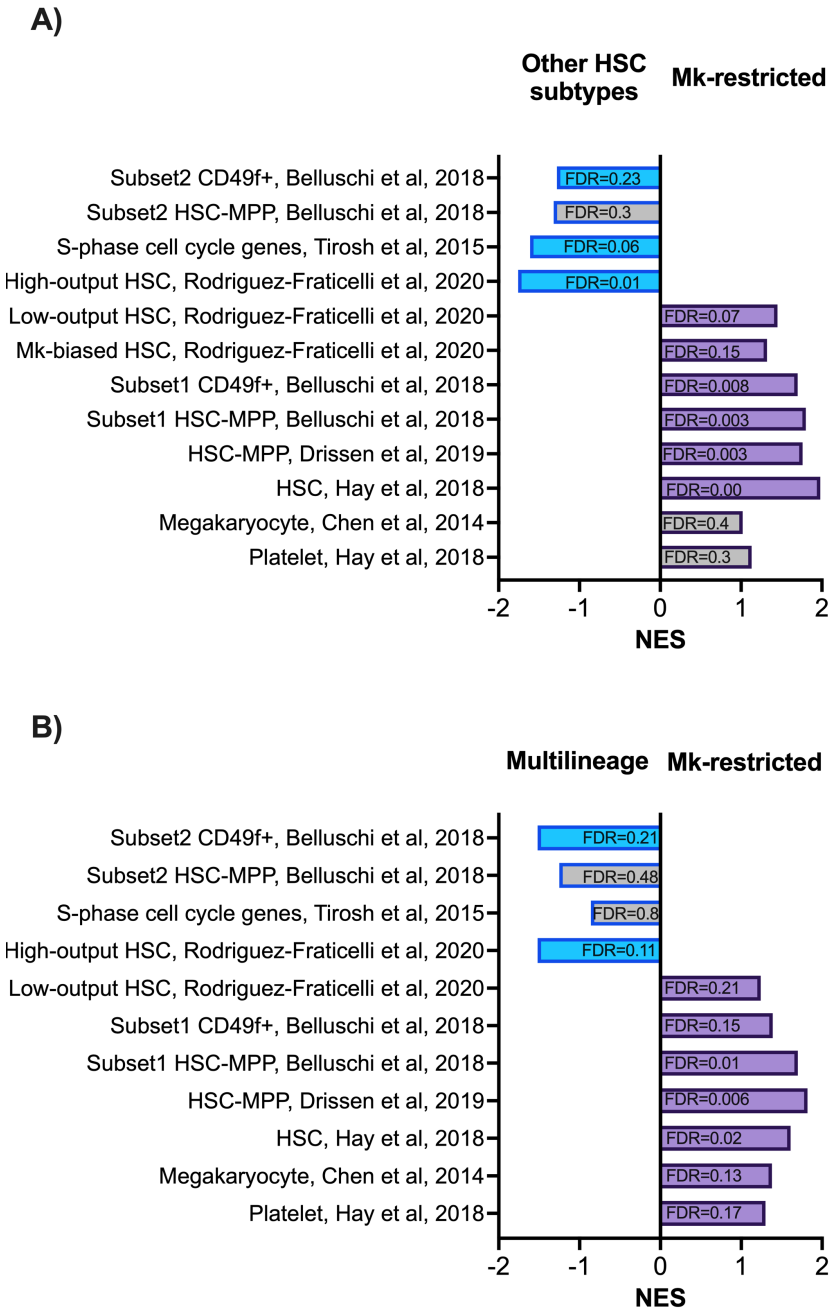


Figure 4.2.7. The Mk-restricted HSC subtype was upregulated for multiple HSC gene-sets some of which may predict their Mk output and less proliferative state.
GSEA analysis was performed to compare the Mk-restricted subtype to **A)** all other identified HSC subtypes (EryMy-biased, Multilineage, MkEry-biased, B-Ly biased, and Ery-restricted) as well as **B)** Multilineage subtype separately. The y-axis indicates the gene-sets from the literature. NES>0 indicates the positive enrichment of gene-sets and NES<0 indicates the downregulated gene-sets. FDR-values for each gene-set were indicated on the figure. FDR≤0.25 were considered as significant and the gene-sets above that threshold were indicated with the grey colour. NES=Normalised Enrichment Score, FDR=False Discovery Rate.

To identify genes that contributed the most to the enrichment score, leading-edge analysis was performed following GSEA [329], and key driver genes were plotted based on their log normalised expression values (Figure 4.2.8). *Socs2*, *Mecom*, *Mllt3*, *Hlf*, *Avp*, and *CD74* were some of the leading-edge genes that were driving the enrichment score of the analyses (Figure 4.2.7 A and B). These genes were also found significantly upregulated in the Mk-restricted subtype as compared to the remaining cells after performing DGE analysis (see Figure 4.2.6). *Socs2* [350], *Mecom* [351], *Mllt3* [352], *Hlf* [353, 354], *Avp* [320], and *CD74* [279, 355] were previously reported to play a role in HSC pool maintenance as well as their quiescence state. *Mllt3* was also previously identified as an early regulator of Mk and Ery lineage decisions in human UCB cells [356]. The *Avp* gene was reported to accelerate the differentiation of BM Ery precursors in anaemia [357]. As shown in Figure 4.2.8, the mean expression values of *Mllt3* and *Avp* genes were higher in the Mk-restricted and Ery-restricted subtypes than the other HSC subtypes and this could be relevant to their role in Mk and Ery lineage priming and differentiation (Figure 4.2.8).

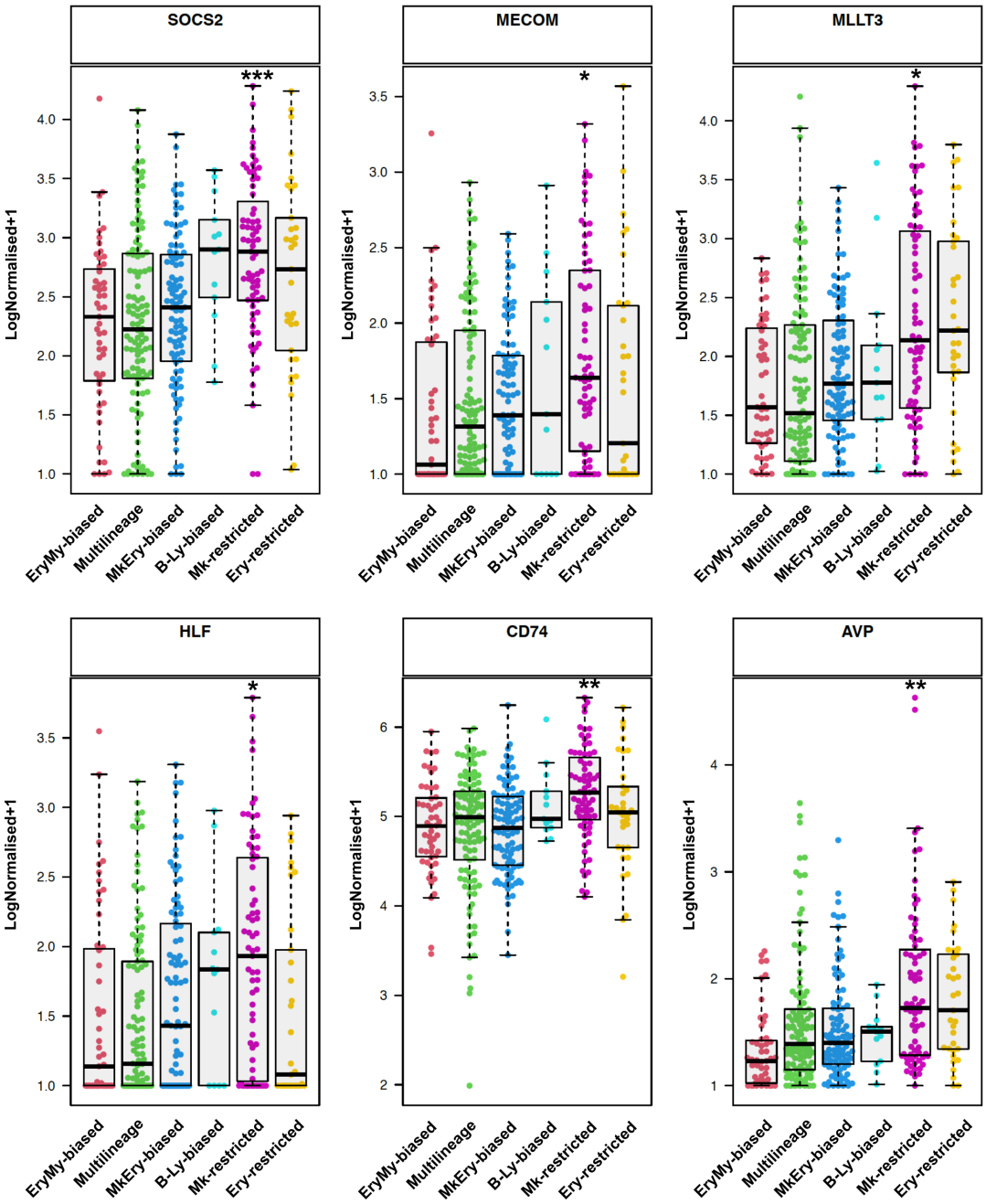


Figure 4.2.8. HSC maintenance related genes were significantly upregulated in the Mk-restricted subtype, as revealed by DEG and leading-edge analysis after the GSEA.
Log normalised expression values (natural log) of genes upregulated in the Mk-restricted subtype when compared to the remaining cells using Wilcoxon rank-sum test (selected genes: $\log_2FC \geq 0.5$ and * BH-adjusted p-value <0.05, ** <0.01, and *** <0.001). Every circle represents one cell, and the horizontal bar shows the mean values.

We also compared biased/restricted HSC subtypes (Mk-restricted, Ery-restricted, EryMy-biased, and MkEry-biased) to the Multilineage subtype in GSEA using the same gene-sets as before along with erythroid progenitor [194, 317] and HSC self-

renewal related gene-sets [331] (Figure 4.2.9 A). The analysis showed significant upregulation of Subset 1 HSC-MPP gene-set in the biased/restricted subtypes, while the Multilineage HSCs were enriched for Subset 2 CD49f^{+} gene-set. In addition, we detected significant enrichment of another HSC-MPP gene-set as well as protein-folding and Endoplasmic Reticulum (ER) stress/unfolded protein response (UPR) gene-sets in the biased/restricted subtypes (Figure 4.2.9 A). Protein-folding and ER-stress gene signatures were previously reported to be upregulated in human LT-HSCs when compared to the ST-HSCs and associated with LT-HSC self-renewal and/or quiescence exit [331]. Pre-CFUE, MEP, and Mk gene-sets were also enriched significantly in the biased/restricted subtypes as compared to the Multilineage (Figure 4.2.9 A). Collectively, this analysis suggests that the biased/restricted subtypes are activated for early Ery and Mk lineage priming programmes and they might have a higher self-renewing capacity as compared to the Multilineage HSCs (Figure 4.2.9 A).

To explore the transcriptional differences further, DGE analysis between biased/restricted subtypes and Multilineage subtype was performed, and this revealed different enrichment of gene signatures related to metabolic state (Figure 4.2.9 B). A few genes were significantly upregulated in the biased/restricted subtypes, including *Eno1* (involved in glycolysis [358]), *ERMP1* (Endoplasmic Reticulum Metallopeptidase 1, involved in UPR [359]), and *CD81* (associated with the re-entry of HSCs to the quiescent state following stress-induced proliferation [360]) (Figure 4.2.9 B). Genes significantly upregulated in the Multilineage subtype were mostly ribosomal protein-coding genes (RPS- and RPL- genes involved in ribosome biogenesis) (Figure 4.2.9 B). HSCs exiting dormancy and shifting to active

state show higher biosynthetic activity and protein synthesis than dormant HSCs [361]. Moreover, upregulation of ribosomal protein genes was previously demonstrated to be associated with ageing in HSCs [362]. Glutathione S-transferase genes, *Gstm1* and *Gstm5,* that play role in detoxification by conjugating to xenobiotics and defence against environmental stress, were also upregulated in the Multilineage HSC subtype. To understand the biological differences between the biased/restricted and the Multilineage subtypes further, gene ontology (GO) analysis was performed in GSEA using molecular function and cellular component GO terms and hallmark gene-sets obtained from the Broad Institute's molecular signature database [329, 335] (Figure 4.2.10 A and B). The biased/restricted subtypes showed significant enrichment of glycolysis and fatty acid metabolism gene-sets which are related to the HSC quiescence state [17] and self-renewal related asymmetric division in HSCs [363] (Figure 4.2.10 A). In contrast, the Multilineage subtype was enriched for multiple oxidative phosphorylation related gene-sets (two representative gene-sets were shown) and ribosomal gene-sets (Figure 4.2.10 B). Overall, comparison to the biased/restricted subtypes suggests that the Multilineage HSC subtype might be more proliferative/active, having exited from the quiescence state as indicated by the activation of oxidative phosphorylation and increased transcription of ribosomal genes. Ultimately, this increase in metabolic activity could impair HSCs, lead to their exhaustion and loss of self-renewal activity [364]. Biased/restricted subtypes might be still maintained in a quiescent state in hypoxic niches where they rely on glycolysis, given the upregulation of glycolytic gene signatures.

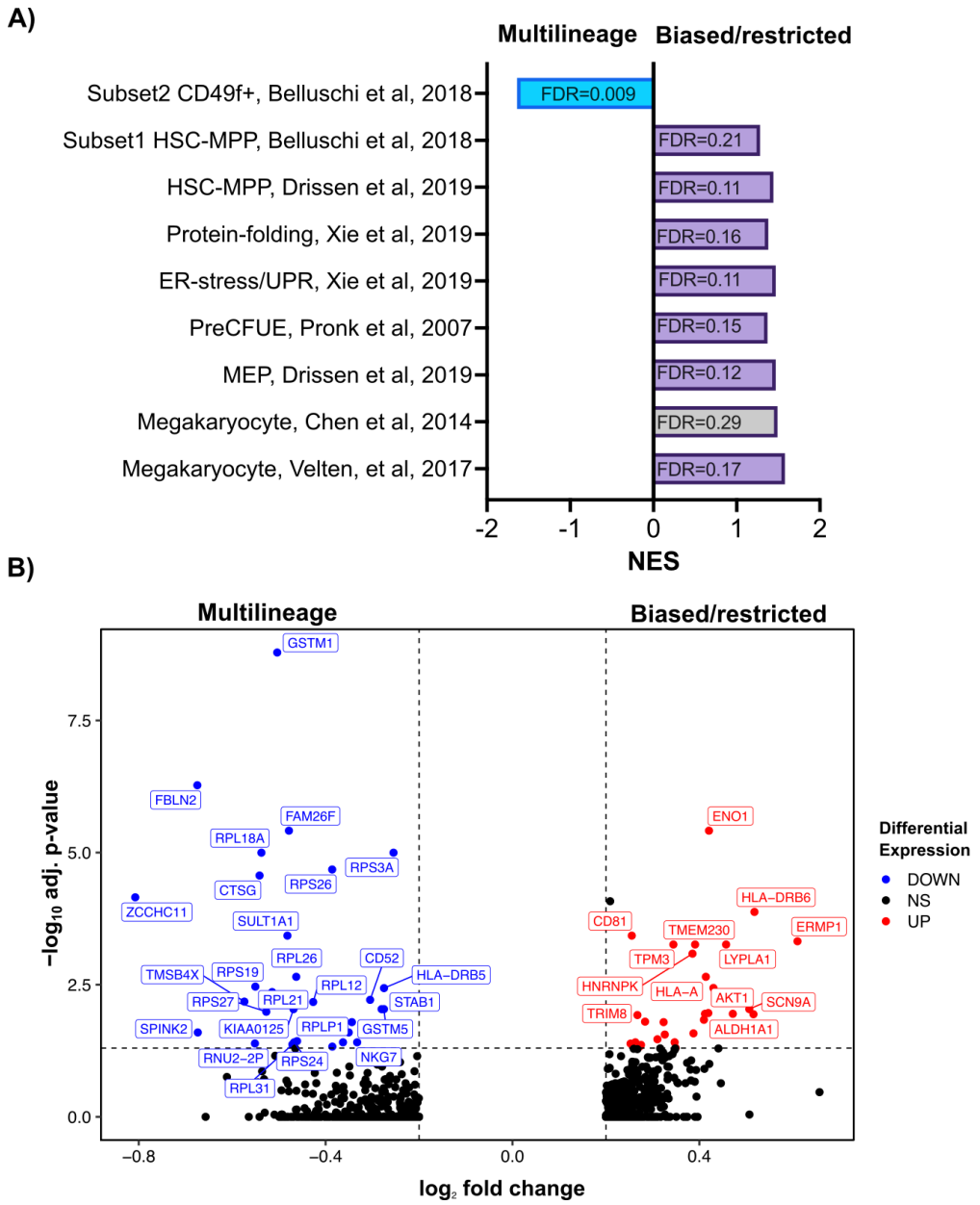


Figure 4.2.9. Biased/restricted subtypes (Mk-restricted, Ery-restricted, EryMy-biased, and MkEry-biased) might be more self-renewing and primed for Mk and Ery lineage differentiation as compared to the Multilineage subtype.
A) Positive enrichment: NES>0, downregulated gene-sets: NES<0. FDR-values for each gene-set were indicated on the figure. FDR≤0.25 were considered as significant and the gene-sets above threshold were indicated with the grey colour. NES=normalised enrichment score, FDR=False Discovery Rate. **B)** Volcano plot showing the differentially expressed genes between biased/restricted subtypes (red, n=250 cells) and the Multilineage subtype (blue, n= 114 cells). Genes were tested using Wilcoxon rank-sum test. BH-adjusted p-value <0.05 and fold change of 0.25 were highlighted. Black points show the not significant genes.

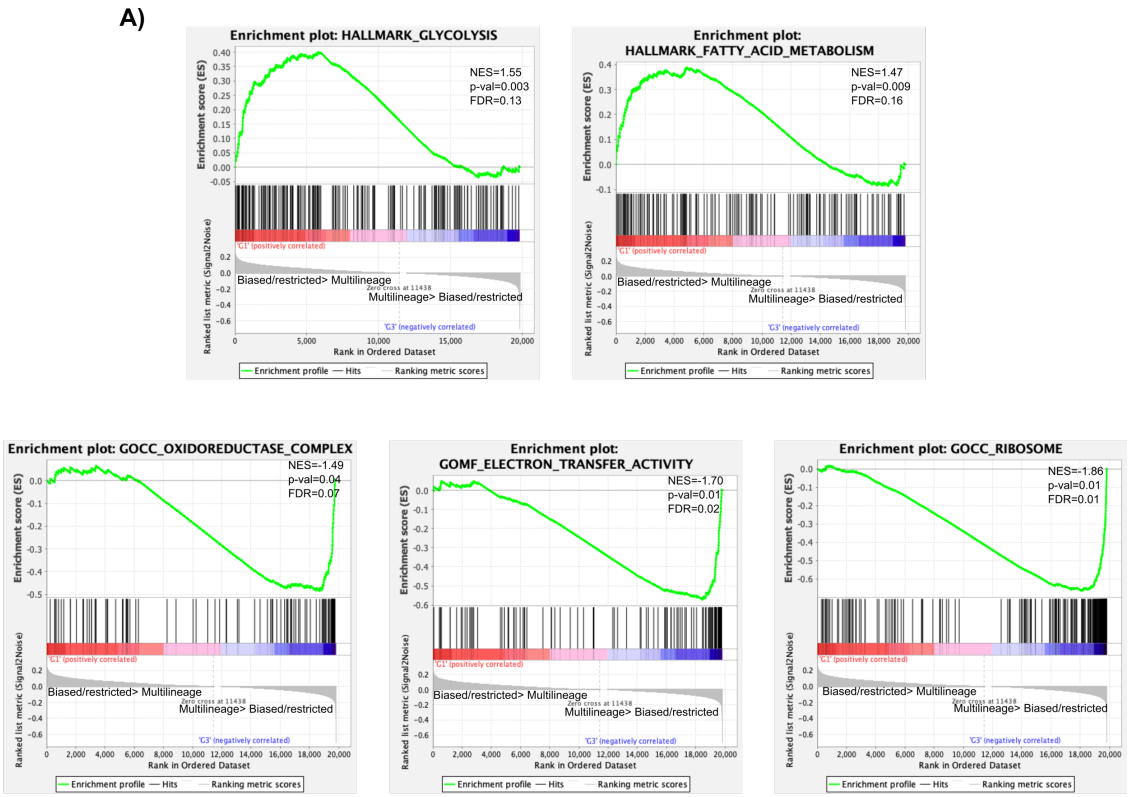


Figure 4.2.10. Representative GSEA plots of gene ontology terms from comparison of biased/restricted subtypes to the Multilineage subtype.
Gene-sets enriched in **A)** biased/restricted subtypes **B)** Multilineage. NES, nominal p-value, and FDR values for each gene-set were indicated on the enrichment plots. NES=Normalised Enrichment Score, FDR=False Discovery Rate. Gene ontology and hallmarks gene-sets were obtained from the Broad Institute's molecular signature database.

4.3. Discussion

In this study, we linked the human BM-derived HSC fates identified by clonal analysis to their single-cell transcriptomes to identify transcriptional signatures of HSC fates, which was not possible before using solely single-HSC transcriptomes [285, 288]. Our data provide the first demonstration of human adult BM HSC fates and associated transcriptional signature using combined barcode analysis and single-HSC transcriptomes. The barcoded human HSCs and their cellular *in vivo* outputs were purified 16-weeks post-transplantation for clonal and transcriptomic analysis. The clonal analysis showed that the HSC clones varied in their lineage

outputs with a significant proportion demonstrating restricted and biased patterns of mature cell outputs towards the generation of erythroid/myeloid/megakaryocytic lineages (Figure 4.2.1). Using hierarchical clustering of HSC clones based on their mature cell output, we identified six human HSC subtypes with distinct lineage fate and similar in mean clone size: Mk-restricted, Ery-restricted, MkEry-biased, EryMy-biased, B-Ly-biased, and Multilineage (Figure 4.2.2 and Figure 4.2.4). According to this, Mk cells were produced via a distinct pathway (independent of other lineages) by Mk-restricted HSC subtype, a shared pathway with the Ery/My lineages by MkEry-biased HSC subtype, and the Ery/My/Ly lineages by Multilineage HSC subtype (Figure 4.2.3). Carrelha *et al* using single-cell transplantations, previously described self-renewing murine HSCs with platelet-restricted fate replenishing only megakaryocyte/platelet lineages, but not any other cells, despite maintaining their multipotency [202]. Additionally, they also described simultaneous replenishment of Mk, Ery, and My lineages by multipotent self-renewing platelet-erythroid-myeloid fate-restricted HSC clones lacking Ly output, as well as multilineage HSC clones generating all lineages, including Ly [202]. Thus, our identified human HSC fate-restriction patterns (the Mk-restricted and MkEry-biased subtypes), as well as the Multilineage subtype are compatible with previous findings in murine HSCs. Moreover, Fraticelli *et al* utilising transposon tagging in mice yielded similar findings in unperturbed haematopoiesis. They demonstrated that half of the Mk lineage was replenished independently of other lineages [247]. Furthermore, Pei *et al* using Polylox barcoding, reported MyEry restricted and biased as well as Multilineage murine HSC subsets in unperturbed haematopoiesis; however, Mk or platelet lineage was not sampled in that study [256].

Cellular barcoding allowed us to cluster HSC subtypes by their lineage outputs and to identify functional heterogeneity within the HSC compartment. To understand the molecular hallmarks of these distinct differentiation behaviours, we investigated the transcriptional differences among functionally different human HSC subtypes. By performing DGE analysis, transcriptional heterogeneity within the phenotypically defined human HSC population was revealed. Most importantly, HSC subtypes with different *in vivo* lineage outputs showed distinct gene-expression patterns (Figure 4.2.6). Analysis of DEGs might allow identification of molecular markers to predict the HSC output.

Previous work from Rodriguez-Fraticelli *et al* using lentiviral barcoding and scRNA-seq demonstrated a Mk-biased murine HSC subset with low progeny output (described as “self-renewing significantly more than differentiating” [279]). The murine Mk-biased subset was also upregulated for markers associated with HSC quiescence and self-renewal. Belluschi *et al* reported two subsets in human UCB-derived HSC-MPP and CD49f$^+$: Subset 1 HSC-MPP and Subset 1 CD49f$^+$ were highlighted with My/Ery and Mk potential *in vitro* and Ery/Mk lineage priming programmes, while Subset 2 HSC-MPP and Subset 2 CD49f$^+$ emerged with My/Ly potential *in vitro* and they were upregulated for Ly differentiation-programs [205]. Using the above mentioned low-output, high-output and Mk-biased gene-sets [279], as well as Subset 1 and Subset 2 gene-sets [205], we compared the human HSC subtypes in GSEA. Compared to all other HSC subtypes, the Mk-restricted subtype was significantly upregulated for Subset 1 HSC-MPP, Subset 1 CD49f$^+$, Mk-biased HSC and low-output HSC gene signatures as well as two more tested HSC-related gene-sets [287, 317] (Figure 4.2.7 A). In line with this, Subset 2 CD49f$^+$, high-output

HSC, and S-phase cell cycle gene-sets were downregulated in the Mk-restricted subtype, showing significant enrichment in the rest of the HSC subtypes. Comparison of the Mk-restricted subtype to the Multilineage led to similar findings: we observed significant upregulation of megakaryocyte- and platelet- affiliated gene signatures in the Mk-restricted HSC subtype. The Multilineage subtype showed significant upregulation of Subset 2 CD49f^{+} and high-output HSC signatures. By performing DGE analysis and leading-edge analysis, we identified a set of genes linked to HSC function (*Socs2*, *Mecom*, *Mllt3*, *Hlf*, *Avp*, and *CD74*) significantly upregulated in the Mk-restricted subtype compared to other subtypes (Figure 4.2.8). Collectively, we identified a functionally and molecularly distinct human HSC subset within the HSC compartment, the Mk-restricted subtype: it represents a relatively quiescent subset with high-self-renewal and low-output activity as well as shows priming for megakaryocyte and platelet-specific gene expression (Figure 4.2.7 and Figure 4.2.8). This is only inferred from the gene signatures, and higher self-renewal capacity of Mk-restricted HSCs will need to be confirmed by secondary transplantations. These findings are in line with the previously demonstrated molecularly platelet/megakaryocyte-primed and functionally distinct murine platelet-biased HSC subpopulation described in [203]. This highlights that the lineage commitment and restriction occur at the top of the hierarchy, within the phenotypic HSCs, before differentiating into progenitors, in contrast to the classical haematopoiesis hierarchy where the lineage bifurcation thought to occur downstream of HSC-MPPs [189-192, 195].

The Multilineage HSCs seem to represent a subset with high repopulation activity and might be activated for lymphoid differentiation programs. The latter was suggested based on the observation that Subset 2 gene-sets (previously highlighted with myeloid-lymphoid potential) were upregulated in the Multilineage subtype as compared to the Mk-restricted or the biased/restricted subtypes in GSEA (Figure 4.2.7 A and Figure 4.2.9 A) since the lymphoid genes were previously reported as not being transcriptionally active in HSCs [212]. In addition, the biased/restricted subtypes demonstrated erythroid and megakaryocyte priming. They might be more self-renewing compared to the Multilineage subtype based on the significant enrichment of protein folding and ER-stress/UPR gene-sets (Figure 4.2.9 A), as previous data have demonstrated the enrichment of these gene-sets in LT-HSCs compared to the ST-HSCs [331]. This is concluded from the gene signatures, and higher self-renewal capacity of these HSCs will need to be confirmed by secondary transplantations. In addition, the enrichment of ER-stress/UPR might be associated with maintaining the integrity of the HSC pool under stress conditions, such as hypoxic stress or proliferation [365, 366]. HSCs reside in hypoxic niches within the BM in a quiescent state and rely on glycolysis [17]. This explains their adaptation to low oxygen levels and low energy requirements and displays a protection mechanism against cellular stress and accumulating DNA damage [17, 364, 367]. Cells alter their metabolism from glycolysis to oxidative phosphorylation during the transition from quiescence to an activated state, as they require more energy for proliferation [367]. The DGE and GO analysis showed the glycolytic phenotype of biased/restricted subtypes and upregulation of a fatty acid metabolism gene-set that was previously implicated to maintain HSC pool via asymmetric cell division [363] (Figure 4.2.9 B and Figure 4.2.10 A). In contrast, the Multilineage HSC subtype

showed the activation of oxidative phosphorylation as well as increased transcription of ribosomal genes (Figure 4.2.9 B and Figure 4.2.10 B). This suggests that the Multilineage HSC subtype may no longer be in a quiescent state, but rather they were activated for proliferation or aged, and switched to a metabolically active state (employing oxidative phosphorylation). This also explains the upregulation of ribosomal genes that are involved in protein translation and production, as the ribosomes need to meet the high energy demand [361, 362]. This metabolic change could cause cellular damage through an accumulation of reactive oxygen species, which may contribute to HSC ageing and/or HSC exhaustion [364].

As with any techniques, cellular barcoding also has some intrinsic limitations. The size of some clones might be small and close to the detection threshold, therefore the absence of some of the clones should be interpreted carefully. We filtered out barcodes with less than 10 reads in total as barcodes below the threshold may be subject to background noise. Although stricter thresholds could be considered, this may reduce the number of detected clones and lead to the failure of detection of small clones. Besides, our data do not provide information about the changes in the clonality overtime after the transplantation. That is simply limited by the amount of BM sample (up to 5 µl) that can be aspirated at a time from a live xenograft. The purification of all barcoded ($EGFP^+$) lineages, particularly megakaryocytic cells, with enough biological replicates from a BM aspirate was not possible. Furthermore, these experiments were performed by transplanting and retrieving phenotypically defined human HSCs based on the specific combination of cell surface markers. Tracking cells followed by transplantation is considered as a potential perturbation of native haematopoiesis and associated with stress-induced haematopoiesis [368].

It was proposed that the transplanted cells may not behave in the same way as they would in physiological conditions. Indeed, experiments performed using *in vivo* labelling to track murine HSCs (via generation of endogenous barcodes, such as Polylox, transposons, or CRISPR/Cas9) provide insight into native haematopoiesis [244, 246, 247, 255, 256], which allows simultaneous tracking of thousands of clones in a single mouse without a prior need for prospective isolation of cell populations by FACS and their transplantation. However, these lineage tracing methods are not applicable for the study of human HSCs for ethical reasons. Alternative approaches to study human haematopoiesis include tracking the endogenous randomly occurring somatic mutations such as single nucleotide variants, copy number variants, microsatellites, and mitochondrial DNA to infer fate decisions [228-231, 236]. These are useful methodologies, especially for retrospective analysis of human haematological malignancies [232, 233]. Although these methods do not require external manipulation to track the cells, the detection of nuclear somatic mutations by whole-genome sequencing in each cell is costly, contains high error rates, and it remains difficult to quantify the contributions of different clones [234, 235]. Even though the transplantation of human cells into mice might perturb the native haematopoiesis, we used the *Kit* mutant NBSGW strain allowing improved human cell engraftment without prior irradiation and read out of all blood lineages [148], therefore it provides a useful model to study human haematopoiesis in minimal disruption. Furthermore, in a clinical setting, patients with haematological diseases undergo autologous or allogeneic BM transplantations or receive gene therapy, which is the only curative treatment currently available. Therefore, our experimental setting to study human BM HSC fate and their associated transcriptional signature is compatible with the current therapy strategies

i.e., gene therapy workflow involves harvesting patient HSPCs, their pre-stimulation with early-acting cytokines, lentiviral transduction, and re-transplantation into myeloablated patients (Figure 1.2.1 B). Although one obvious difference to this study is that the presented experiments were not performed in humans. An alternative lineage tracing method allowing to study of human haematopoiesis is through using retroviral integration sites in patients who have undergone gene therapy treatments [240, 251, 252]. As with most of the techniques, this method also has methodological constraints such as bias in fragment amplification of variable length, dependency on the viral integration site, and restriction endonucleases [254]. Overall, our approach to study *in vivo* human haematopoiesis displays the best available method to track the HSC cell fate decisions at the single-cell level.

This thesis lays the groundwork for various future studies shedding more light on the underappreciated human HSC heterogeneity. Although it is not covered in this study, the stability of the fates of the HSC clones identified herein could be confirmed using serial transplantations. This could also shed light onto the suggested higher self-renewal activity (inferred from the gene signatures) of biased/restricted subtypes. However, it should be noted that this an additional stress factor that may force HSCs to proliferate again. Besides it is also technically challenging to purify enough barcoded HSCs and their lineage output to infer HSC fates and analyse single HSC transcriptomes from primary recipients as well as to maintain sufficient cell doses for secondary transplantations. Moreover, human HSC heterogeneity could be studied epigenetically using lentiviral barcoding and single-cell assay for transposase-accessible chromatin with sequencing (scATAC-seq) [369]. For this, optimisation of targeted barcode amplification along with the

epigenome is required. Furthermore, it may also be relevant to identify lineage bias and restriction of human HSCs isolated from BM of healthy young and old individuals to reveal age-associated changes in HSC clonal diversity and the associated molecular signatures. However, engraftment and regeneration capacity of aged HSCs in murine xenografts compared to the young may not be efficient [370, 371]. Overall, these are exciting future prospects highlighting the need to further explore human HSC heterogeneity.

Chapter 5: Conclusions

This work aimed to characterise human adult BM HSC heterogeneity, particularly focusing on the identification of human HSCs with distinct lineage outputs *in vivo* and their transcriptional differences. We used functional single-cell analysis, i.e., cellular barcoding, in combination with single-cell transcriptomics to address this question. In Chapter 3, I presented optimisation and establishment of the experimental protocols to purify human adult BM-derived HSCs, their transduction with a lentiviral barcode library by minimal *in vitro* manipulation, and their stable engraftment in an immunodeficient mouse model up to 4 months post-transplantation. In addition, I designed multi-colour FACS panels to identify barcoded human HSCs along with their barcoded cellular *in vivo* output (B-Ly, My, Mk, and Ery lineages) in murine BM. Moreover, I successfully purified barcoded single HSCs and their all-barcoded mature cell output by FACS, despite the rarity of human barcoded HSCs and the human Mk lineage in murine xenografts. Furthermore, I performed scRNA-seq of barcoded HSCs and bulk RNA-seq of their barcoded cell outputs. The latter required optimisation of cDNA libraries generated using the Mk and My samples. By testing various conditions that may affect the quality of cDNA libraries, we found that the partial degradation observed in the Mk cDNA libraries was due to a small diameter nozzle (85 μm) used to purify them by FACS.

In Chapter 4, I performed computational analyses of the scRNA-seq data to identify and correlate the barcodes in single HSCs to their mature lineage outputs as well as to reveal transcriptional differences between functional HSC subtypes. Using hierarchical clustering of HSC clones based on their lineage contribution, we have uncovered functionally distinct six HSC subtypes within the phenotypic HSC compartment, and they were referred to as Mk-restricted, Ery-restricted, MkEry-biased, EryMy-biased, B-Ly-biased, and Multilineage. Next, the transcriptomes of the HSC subtypes were statistically compared to each other to reveal underlying molecular differences and any initiated transcriptional lineage priming programmes using DGE analysis and GSEA. We identified the human Mk-restricted HSC subtype as a functionally and molecularly distinct HSC subset that was marked with HSC quiescence and self-renewal transcriptional signatures and primed towards the generation of Mk lineage. Moreover, the Mk-restricted, Ery-restricted, MkEry-biased, and EryMy-biased HSC subtypes together (referred to as biased/restricted subtypes) showed the activation of transcriptional priming towards Mk and Ery lineage differentiation when compared to the Multilineage subtype. The significant enrichment of Subset 2 $49f^{+}$ gene-sets (from Belluschi *et al*) in the Multilineage subtype seemed to predict their My/Ly output. Taken together, these results suggest that the lineage-priming programmes seem to be established in functional human HSC subtypes, highlighting that the functional phenotype matches with the molecular phenotype of human HSCs.

Further molecular analysis also revealed that there might be metabolic differences between the Multilineage and biased/restricted HSC subtypes. The biased/restricted HSC subtypes were enriched for HSC and self-renewal related as

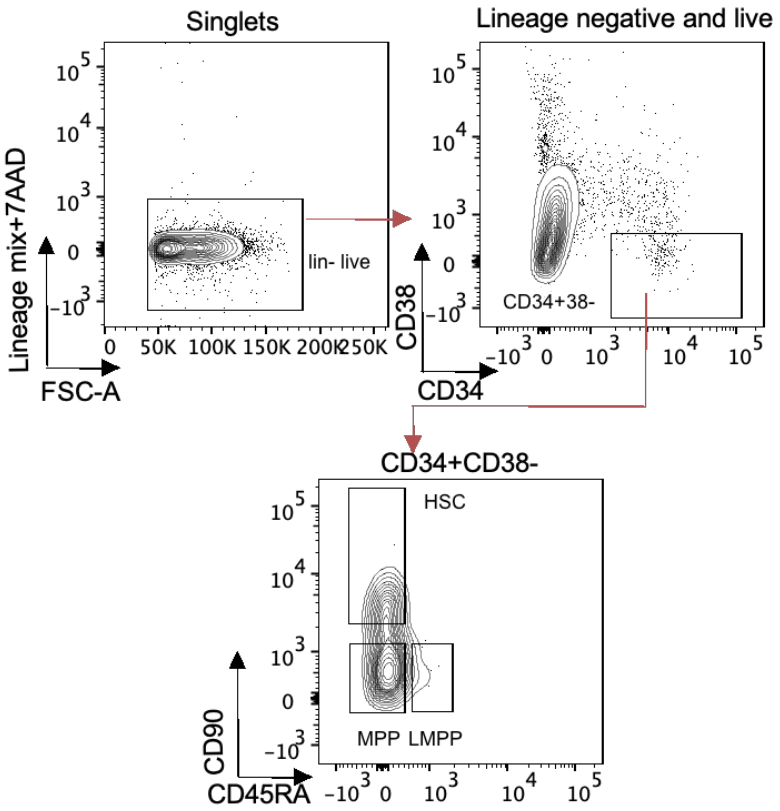
well as glycolysis and asymmetric division related transcriptional signatures. This indicates the maintenance of biased/restricted HSC subtypes in a quiescent state and protection from possible damage caused by cellular stress. In contrast, the Multilineage subtype showed hallmarks of HSC activation, a metabolic switch to oxidative phosphorylation, and increased protein synthesis. This might be correlated with their high repopulating ability. Overall, this gives insight into the underlying cellular and molecular processes in the Multilineage and biased/restricted HSC subtypes.

A major hurdle restricting the success of stem cell-based therapies is a lack of our understanding of the human HSC pool constitution. A better definition of the properties of a population of donor cells, for instance, the number of reconstituting human HSCs and their capacity to reconstitute cellular lineages (short-term: neutrophils and platelets; long-term: lymphocytes) will help to reduce morbidity or mortality associated with HSC transplantations [34, 35, 38, 46]. Characterisation of the human HSC heterogeneity in terms of the molecular signatures of specific HSC subtypes with distinct lineage outputs would allow clinicians to predict how transplanted HSCs may behave in patients. This will be crucial for the development of patient-specific treatments, for example, eliminating the disease-affected HSC subtype while maintaining the healthy subpopulation for the haematopoietic function [372]. The identification of megakaryocyte/platelet restricted, or biased HSC subtypes may help to develop *ex vivo* expansion protocols suited to the requirements of these subsets to enhance platelet output, which is likely to improve patient's recovery and reduce morbidity and mortality after transplantations.

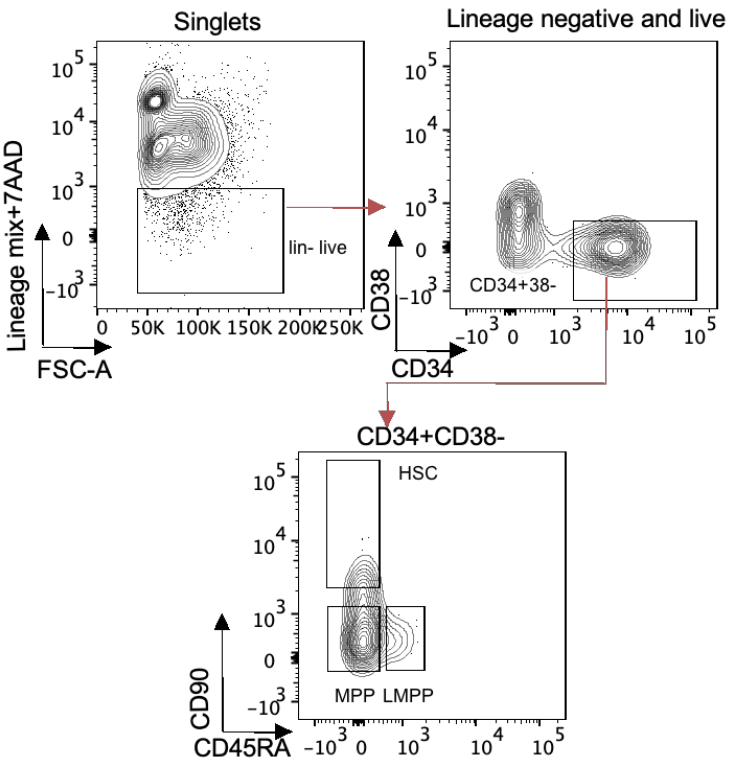
The biological causes of HSC heterogeneity are still a matter of intense research. It might be mainly regulated by intrinsic mechanisms (i.e., transcription, epigenetic alterations, asymmetric cell divisions, DNA mutation, cell cycle, cell size, ageing) [9-13, 197]. However, it is likely that extrinsic signals by the BM niche (i.e., cell-cell contacts, signalling pathways, cytokines) may play a role in HSC heterogeneity, meaning that the progeny of different HSC subsets could show different sensitivity to these signals. Thus, this behaviour may influence their cell fate choices by supporting survival and proliferation [215, 216]. Therefore, integrative approaches that enable functional, as well as molecular readouts, for example, on the transcriptome, chromatin accessibility, DNA methylation, histone modifications, or the proteome, simultaneously, might be needed to fully understand and characterise the complex lineage choices.

Overall, my work using combined barcode tracking and single-HSC transcriptome analysis provides a few new pieces in the puzzle of human HSC heterogeneity by revealing haematopoietic fate decisions and their associated transcriptional signatures. The gene signatures identified herein are associated with specific HSC lineage output and can be further examined to identify potential molecular and/or surface markers that could help to determine human HSC subsets with specific haematopoietic fate restriction. To our knowledge, this is the first study to demonstrate the HSC heterogeneity in the human adult BM using both functional and molecular analysis.

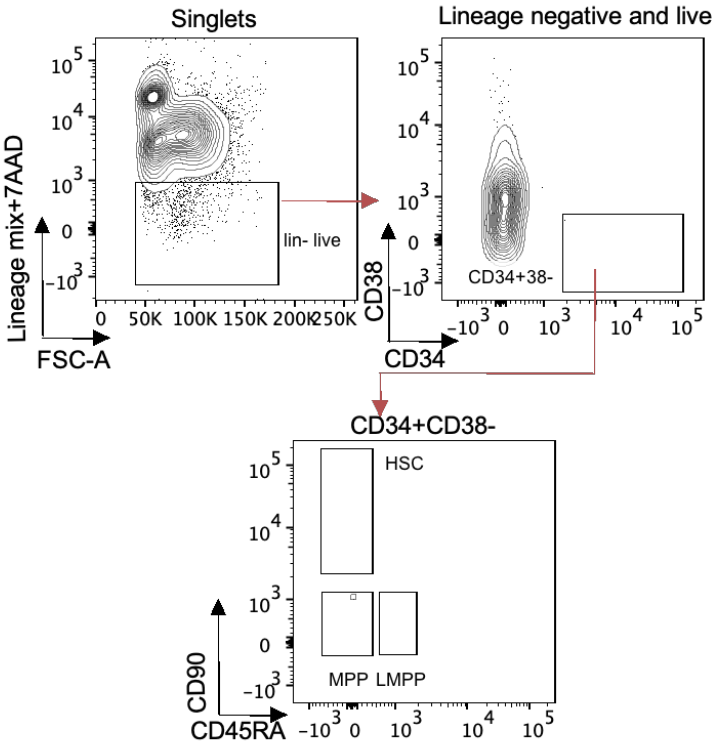
Appendix



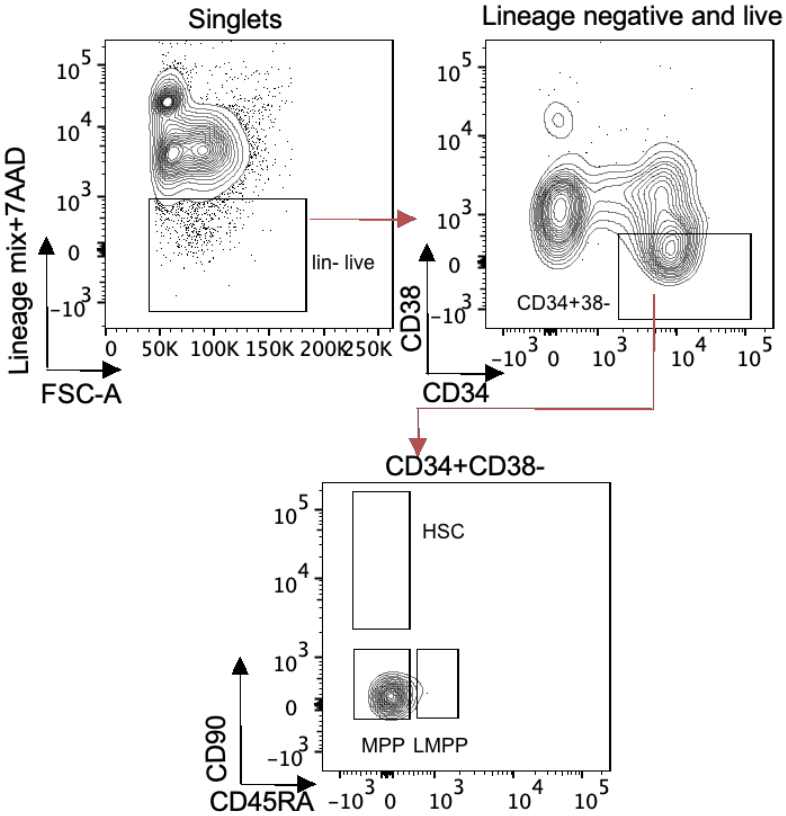
Additional Figure 1. Fluorescence minus one control of Lineage PE-Cy5 that is used to set the Lineage and viability gate for human HSC purification.



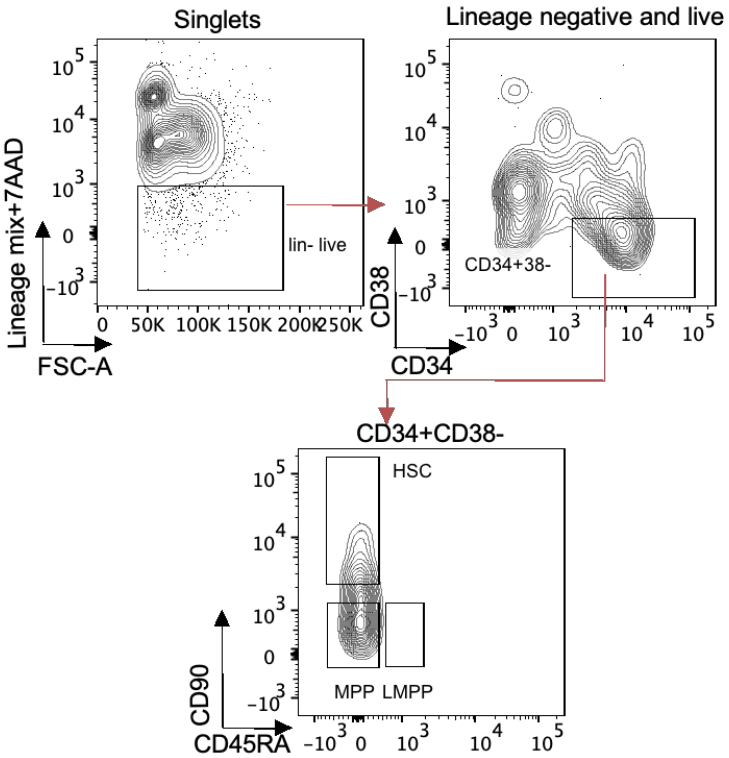
Additional Figure 1. Fluorescence minus one control of CD38 PE-TxRED that is used to set the CD38 gate for human HSC purification.



Additional Figure 3. Fluorescence minus one control of CD34 APC that is used to set the CD34 gate for human HSC purification.



Additional Figure 4. Fluorescence minus one control of CD90 PE that is used to set the CD90 gate for HSC purification.



Additional Figure 5. Fluorescence minus one control of CD45RA FITC that is used to set the CD45RA gate for human HSC purification.

Bibliography

1. Doulatov, S., et al., *Hematopoiesis: a human perspective.* Cell Stem Cell, 2012. **10**(2): p. 120-36.

2. Ogawa, M., *Differentiation and proliferation of hematopoietic stem cells.* Blood, 1993. **81**(11): p. 2844-53.

3. Morrison, S.J., N. Uchida, and I.L. Weissman, *The biology of hematopoietic stem cells.* Annu Rev Cell Dev Biol, 1995. **11**: p. 35-71.

4. Spangrude, G.J. and G.R. Johnson, *Resting and activated subsets of mouse multipotent hematopoietic stem cells.* Proc Natl Acad Sci U S A, 1990. **87**(19): p. 7433-7.

5. Laurenti, E. and B. Gottgens, *From haematopoietic stem cells to complex differentiation landscapes.* Nature, 2018. **553**(7689): p. 418-426.

6. Morrison, S.J. and J. Kimble, *Asymmetric and symmetric stem-cell divisions in development and cancer.* Nature, 2006. **441**(7097): p. 1068-1074.

7. Loeffler, D. and T. Schroeder, *Symmetric and asymmetric activation of hematopoietic stem cells.* Current Opinion in Hematology, 2021. **28**(4).

8. Schofield, R., *The relationship between the spleen colony-forming cell and the haemopoietic stem cell.* Blood Cells, 1978. **4**(1-2): p. 7-25.

9. Zaret, K.S. and S.E. Mango, *Pioneer transcription factors, chromatin dynamics, and cell fate control.* Curr Opin Genet Dev, 2016. **37**: p. 76-81.

10. 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.

11. Ramasamy, S.K., et al., *Regulation of Hematopoiesis and Osteogenesis by Blood Vessel-Derived Signals.* Annu Rev Cell Dev Biol, 2016. **32**(1): p. 649-675.

12. Raynaud, C.M., et al., *Endothelial cells provide a niche for placental hematopoietic stem/progenitor cell expansion through broad transcriptomic modification.* Stem Cell Res, 2013. **11**(3): p. 1074-90.

13. Metcalf, D., *Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation.* Blood, 1998. **92**(2): p. 345-7; discussion 352.

14. Laurenti, E., et al., *CDK6 levels regulate quiescence exit in human hematopoietic stem cells.* Cell Stem Cell, 2015. **16**(3): p. 302-13.

15. Catlin, S.N., et al., *The replication rate of human hematopoietic stem cells in vivo.* Blood, 2011. **117**(17): p. 4460-6.

16. Wilson, A., et al., *Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair.* Cell, 2008. **135**(6): p. 1118-29.

17. Simsek, T., et al., *The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche.* Cell Stem Cell, 2010. **7**(3): p. 380-90.

18. Rossi, D.J., et al., *Hematopoietic stem cell quiescence attenuates DNA damage response and permits DNA damage accumulation during aging.* Cell Cycle, 2007. **6**(19): p. 2371-6.

19. Pinho, S. and P.S. Frenette, *Haematopoietic stem cell activity and interactions with the niche.* Nat Rev Mol Cell Biol, 2019. **20**(5): p. 303-320.

20. Crane, G.M., E. Jeffery, and S.J. Morrison, *Adult haematopoietic stem cell niches.* Nat Rev Immunol, 2017. **17**(9): p. 573-590.

21. Essers, M.A., et al., *IFNalpha activates dormant haematopoietic stem cells in vivo.* Nature, 2009. **458**(7240): p. 904-8.

22. Welte, K., et al., *Purification and biochemical characterization of human pluripotent hematopoietic colony-stimulating factor.* Proc Natl Acad Sci U S A, 1985. **82**(5): p. 1526-30.

23. Wright, D.E., et al., *Physiological migration of hematopoietic stem and progenitor cells.* Science, 2001. **294**(5548): p. 1933-6.

24. Caocci, G., M. Greco, and G. La Nasa, *Bone Marrow Homing and Engraftment Defects of Human Hematopoietic Stem and Progenitor Cells.* Mediterr J Hematol Infect Dis, 2017. **9**(1): p. e2017032.

25. Suárez-Álvarez, B., A. López-Vázquez, and C. López-Larrea, *Mobilization and Homing of Hematopoietic Stem Cells*, in *Stem Cell Transplantation*, C. López-Larrea, A. López-Vázquez, and B. Suárez-Álvarez, Editors. 2012, Springer US: New York, NY. p. 152-170.

26. Ratajczak, M.Z. and M. Suszynska, *Emerging Strategies to Enhance Homing and Engraftment of Hematopoietic Stem Cells.* Stem Cell Rev Rep, 2016. **12**(1): p. 121-8.

27. Jacobson, L.O., et al., *Recovery from radiation injury.* Science, 1951. **113**(2940): p. 510-11.

28. Jacobson, L.O., E.K. Marks, and et al., *The role of the spleen in radiation injury.* Proc Soc Exp Biol Med, 1949. **70**(4): p. 740-2.

29. Lorenz, E., C. Congdon, and D. Uphoff, *Modification of acute irradiation injury in mice and guinea-pigs by bone marrow injections.* Radiology, 1952. **58**(6): p. 863-77.

30. Thomas, E.D., et al., *Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy.* New England Journal of Medicine, 1957. **257**(11): p. 491-496.

31. Goodman, J.W. and G.S. Hodgson, *Evidence for stem cells in the peripheral blood of mice.* Blood, 1962. **19**: p. 702-14.

32. Korbling, M., et al., *With first successful allogeneic transplantations of apheresis-derived hematopoietic progenitor cells reported, can the recruitment of volunteer matched, unrelated stem cell donors be expanded substantially?* Blood, 1995. **86**(3): p. 1235.

33. Gluckman, E., et al., *Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling.* N Engl J Med, 1989. **321**(17): p. 1174-8.

34. Daikeler, T., et al., *Allogeneic hematopoietic SCT for patients with autoimmune diseases.* Bone Marrow Transplant, 2009. **44**(1): p. 27-33.

35. Hirano, M., et al., *Allogeneic stem cell transplantation corrects biochemical derangements in MNGIE.* Neurology, 2006. **67**(8): p. 1458-60.

36. Lucarelli, G., et al., *Hematopoietic stem cell transplantation in thalassemia and sickle cell anemia.* Cold Spring Harb Perspect Med, 2012. **2**(5): p. a011825.

37. Wayne, A.S., K. Baird, and R.M. Egeler, *Hematopoietic stem cell transplantation for leukemia.* Pediatr Clin North Am, 2010. **57**(1): p. 1-25.

38. Hatzimichael, E. and M. Tuthill, *Hematopoietic stem cell transplantation.* Stem Cells Cloning, 2010. **3**: p. 105-17.

39. Henig, I. and T. Zuckerman, *Hematopoietic stem cell transplantation-50 years of evolution and future perspectives.* Rambam Maimonides Med J, 2014. **5**(4): p. e0028.

40. Handgretinger, R., et al., *Isolation and transplantation of autologous peripheral CD34+ progenitor cells highly purified by magnetic-activated cell sorting.* Bone Marrow Transplant, 1998. **21**(10): p. 987-93.

41. Verfaillie, C.M., et al., *Kinetics of engraftment of CD34(-) and CD34(+) cells from mobilized blood differs from that of CD34(-) and CD34(+) cells from bone marrow.* Experimental Hematology, 2000. **28**(9): p. 1071-1079.

42. Sidney, L.E., et al., *Concise review: evidence for CD34 as a common marker for diverse progenitors.* Stem Cells, 2014. **32**(6): p. 1380-9.

43. van Lochem, E.G., et al., *Immunophenotypic differentiation patterns of normal hematopoiesis in human bone marrow: reference patterns for age-related changes and disease-induced shifts.* Cytometry B Clin Cytom, 2004. **60**(1): p. 1-13.

44. Servida, F., et al., *Functional and morphological characterization of immunomagnetically selected CD34+ hematopoietic progenitor cells.* Stem Cells, 1996. **14**(4): p. 430-8.

45. Ogonek, J., et al., *Immune Reconstitution after Allogeneic Hematopoietic Stem Cell Transplantation.* Front Immunol, 2016. **7**: p. 507.

46. Saito, A.M., et al., *Costs of allogeneic hematopoietic cell transplantation with high-dose regimens.* Biol Blood Marrow Transplant, 2008. **14**(2): p. 197-207.

47. Smith, A.R. and J.E. Wagner, *Alternative haematopoietic stem cell sources for transplantation: place of umbilical cord blood.* Br J Haematol, 2009. **147**(2): p. 246-61.

48. Gluckman, E., G. Koegler, and V. Rocha, *Human leukocyte antigen matching in cord blood transplantation.* Semin Hematol, 2005. **42**(2): p. 85-90.

49. Gluckman, E., et al., *Factors associated with outcomes of unrelated cord blood transplant: guidelines for donor choice.* Exp Hematol, 2004. **32**(4): p. 397-407.

50. Futrega, K., W.B. Lott, and M.R. Doran, *Direct bone marrow HSC transplantation enhances local engraftment at the expense of systemic engraftment in NSG mice.* Sci Rep, 2016. **6**: p. 23886.

51. Wagner, J.E., et al., *Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival.* Blood, 2002. **100**(5): p. 1611-8.

52. Rubinstein, P., et al., *Outcomes among 562 recipients of placental-blood transplants from unrelated donors.* N Engl J Med, 1998. **339**(22): p. 1565-77.

53. Aggarwal, R., et al., *Hematopoietic stem cells: transcriptional regulation, ex vivo expansion and clinical application.* Curr Mol Med, 2012. **12**(1): p. 34-49.

54. Chou, S., et al., *Expansion of human cord blood hematopoietic stem cells for transplantation.* Cell Stem Cell, 2010. **7**(4): p. 427-8.

55. Zonari, E., et al., *Efficient Ex Vivo Engineering and Expansion of Highly Purified Human Hematopoietic Stem and Progenitor Cell Populations for Gene Therapy.* Stem Cell Reports, 2017. **8**(4): p. 977-990.

56. Horwitz, M.E., et al., *Umbilical cord blood expansion with nicotinamide provides long-term multilineage engraftment.* J Clin Invest, 2014. **124**(7): p. 3121-8.

57. Kiernan, J., et al., *Clinical Studies of Ex Vivo Expansion to Accelerate Engraftment After Umbilical Cord Blood Transplantation: A Systematic Review.* Transfus Med Rev, 2017. **31**(3): p. 173-182.

58. Aiuti, A., et al., *Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome.* Science, 2013. **341**(6148): p. 1233151.

59. Hacein-Bey Abina, S., et al., *Outcomes following gene therapy in patients with severe Wiskott-Aldrich syndrome.* JAMA, 2015. **313**(15): p. 1550-63.

60. Hacein-Bey-Abina, S., et al., *A modified gamma-retrovirus vector for X-linked severe combined immunodeficiency.* N Engl J Med, 2014. **371**(15): p. 1407-17.

61. Cartier, N., et al., *Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy.* Science, 2009. **326**(5954): p. 818-23.

62. Cicalese, M.P., et al., *Gene Therapy for Adenosine Deaminase Deficiency: A Comprehensive Evaluation of Short- and Medium-Term Safety.* Mol Ther, 2018. **26**(3): p. 917-931.

63. Aiuti, A., et al., *Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement.* Nat Med, 2002. **8**(5): p. 423-5.

64. Cavazzana-Calvo, M., et al., *Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia.* Nature, 2010. **467**(7313): p. 318-22.

65. Thompson, A.A., et al., *Gene Therapy in Patients with Transfusion-Dependent beta-Thalassemia.* N Engl J Med, 2018. **378**(16): p. 1479-1493.

66. Biffi, A., et al., *Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy.* Science, 2013. **341**(6148): p. 1233158.

67. Sessa, M., et al., *Lentiviral haemopoietic stem-cell gene therapy in early-onset metachromatic leukodystrophy: an ad-hoc analysis of a non-randomised, open-label, phase 1/2 trial.* Lancet, 2016. **388**(10043): p. 476-87.

68. Goswami, R., et al., *Gene Therapy Leaves a Vicious Cycle.* Frontiers in Oncology, 2019. **9**: p. 297.

69. Verma, I.M., et al. *Gene Therapy: Promises, Problems and Prospects*. in *Genes and Resistance to Disease*. 2000. Berlin, Heidelberg: Springer Berlin Heidelberg.

70. Morgan, R.A., et al., *Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned.* Cell Stem Cell, 2017. **21**(5): p. 574-590.

71. Naldini, L., *Gene therapy returns to centre stage.* Nature, 2015. **526**(7573): p. 351-60.

72. Herzenberg, L.A., et al., *The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford.* Clin Chem, 2002. **48**(10): p. 1819-27.

73. Mandy, F.F., M. Bergeron, and T. Minkus, *Principles of flow cytometry.* Transfus Sci, 1995. **16**(4): p. 303-14.

74. Bain, B.J., *Multiparameter flow cytometry in the diagnosis of hematologic malignancies A.Porwit and M. C.Béné (eds), Cambridge University Press, Cambridge, 2018, ISBN 978-1-107-50383-0.* British Journal of Haematology, 2018. **182**(2): p. 308-308.

75. Wilkerson, M.J., *Principles and applications of flow cytometry and cell sorting in companion animal medicine.* Vet Clin North Am Small Anim Pract, 2012. **42**(1): p. 53-71.

76. McKinnon, K.M., *Flow Cytometry: An Overview.* Curr Protoc Immunol, 2018. **120**: p. 5 1 1-5 1 11.

77. Maciorowski, Z., P.K. Chattopadhyay, and P. Jain, *Basic Multicolor Flow Cytometry.* Curr Protoc Immunol, 2017. **117**(1): p. 5 4 1-5 4 38.

78. Okada, S., et al., *In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells.* Blood, 1992. **80**(12): p. 3044-50.

79. Spangrude, G.J., S. Heimfeld, and I.L. Weissman, *Purification and characterization of mouse hematopoietic stem cells.* Science, 1988. **241**(4861): p. 58-62.

80. Morrison, S.J. and I.L. Weissman, *The Long-Term Repopulating Subset of Hematopoietic Stem-Cells Is Deterministic and Isolatable by Phenotype.* Immunity, 1994. **1**(8): p. 661-673.

81. Spangrude, G.J., D.M. Brooks, and D.B. Tumas, *Long-term repopulation of irradiated mice with limiting numbers of purified hematopoietic stem cells: in vivo expansion of stem cell phenotype but not function.* Blood, 1995. **85**(4): p. 1006-1016.

82. Osawa, M., et al., *Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell.* Science, 1996. **273**(5272): p. 242-5.

83. 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. 1109-21.

84. Adolfsson, J., et al., *Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity.* Immunity, 2001. **15**(4): p. 659-69.

85. Randall, T.D., et al., *Expression of murine CD38 defines a population of long-term reconstituting hematopoietic stem cells.* Blood, 1996. **87**(10): p. 4057-67.

86. Okuno, Y., et al., *Differential regulation of the human and murine CD34 genes in hematopoietic stem cells.* Proc Natl Acad Sci U S A, 2002. **99**(9): p. 6246-51.

87. Terstappen, L.W., et al., *Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells.* Blood, 1991. **77**(6): p. 1218-27.

88. Baum, C.M., et al., *Isolation of a candidate human hematopoietic stem-cell population.* Proc Natl Acad Sci U S A, 1992. **89**(7): p. 2804-8.

89. Sitnicka, E., et al., *Human CD34+ hematopoietic stem cells capable of multilineage engrafting NOD/SCID mice express flt3: distinct flt3 and c-kit expression and response patterns on mouse and candidate human hematopoietic stem cells.* Blood, 2003. **102**(3): p. 881-6.

90. Ebihara, Y., et al., *Reconstitution of human haematopoiesis in non-obese diabetic/severe combined immunodeficient mice by clonal cells expanded from single CD34+CD38- cells expressing Flk2/Flt3.* Br J Haematol, 2002. **119**(2): p. 525-34.

91. Kawashima, I., et al., *CD34(+) human marrow cells that express low levels of Kit protein are enriched for long-term marrow-engrafting cells.* Blood, 1996. **87**(10): p. 4136-4142.

92. Sakabe, H., et al., *Haematopoietic action of flt3 ligand on cord blood-derived CD34-positive cells expressing different levels of flt3 or c-kit tyrosine kinase receptor: comparison with stem cell factor.* Eur J Haematol, 1998. **60**(5): p. 297-306.

93. 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.* J Immunol, 1984. **133**(1): p. 157-65.

94. Berenson, R.J., et al., *Antigen CD34+ marrow cells engraft lethally irradiated baboons.* J Clin Invest, 1988. **81**(3): p. 951-5.

95. Miltenyi, S., et al., *High gradient magnetic cell separation with MACS.* Cytometry, 1990. **11**(2): p. 231-8.

96. 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. 635-45.

97. Notta, F., et al., *Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment.* Science, 2011. **333**(6039): p. 218-21.

98. Notta, F., et al., *Distinct routes of lineage development reshape the human blood hierarchy across ontogeny.* Science, 2016. **351**(6269): p. aab2116.

99. Lorenz, E., C. Congdon, and D. Uphoff, *Modification of Acute Irradiation Injury in Mice and Guinea-Pigs by Bone Marrow Injections.* Radiology, 1952. **58**(6): p. 863-877.

100. Till, J.E. and C.E. Mc, *A direct measurement of the radiation sensitivity of normal mouse bone marrow cells.* Radiat Res, 1961. **14**: p. 213-22.

101. Notta, F., et al., *Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment.* Science, 2011. **333**(6039): p. 218.

102. Kent, D.G., et al., *Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential.* Blood, 2009. **113**(25): p. 6342-50.

103. 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.

104. Parekh, C. and G.M. Crooks, *Critical differences in hematopoiesis and lymphoid development between humans and mice.* J Clin Immunol, 2013. **33**(4): p. 711-5.

105. Shultz, L.D., F. Ishikawa, and D.L. Greiner, *Humanized mice in translational biomedical research.* Nat Rev Immunol, 2007. **7**(2): p. 118-30.

106. Cheung, A.M., et al., *Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice.* Blood, 2013. **122**(18): p. 3129-37.

107. Rahmig, S., et al., *Improved Human Erythropoiesis and Platelet Formation in Humanized NSGW41 Mice.* Stem Cell Reports, 2016. **7**(4): p. 591-601.

108. Cosgun, K.N., et al., *Kit regulates HSC engraftment across the human-mouse species barrier.* Cell Stem Cell, 2014. **15**(2): p. 227-38.

109. Escobar, G., et al., *Genetic engineering of hematopoiesis for targeted IFN-alpha delivery inhibits breast cancer progression.* Sci Transl Med, 2014. **6**(217): p. 217ra3.

110. Ishikawa, F., et al., *Development of functional human blood and immune systems in NOD/SCID/IL2 receptor {gamma} chain(null) mice.* Blood, 2005. **106**(5): p. 1565-73.

111. Mazurier, F., et al., *Lentivector-mediated clonal tracking reveals intrinsic heterogeneity in the human hematopoietic stem cell compartment and culture-induced stem cell impairment.* Blood, 2003. **103**(2): p. 545.

112. Vormoor, J., et al., *Immature human cord blood progenitors engraft and proliferate to high levels in severe combined immunodeficient mice.* Blood, 1994. **83**(9): p. 2489-97.

113. Kim, D.K., et al., *Comparison of hematopoietic activities of human bone marrow and umbilical cord blood CD34 positive and negative cells.* Stem Cells, 1999. **17**(5): p. 286-94.

114. Lansdorp, P.M., W. Dragowska, and H. Mayani, *Ontogeny-related changes in proliferative potential of human hematopoietic cells.* J Exp Med, 1993. **178**(3): p. 787-91.

115. Lu, L., et al., *Enrichment, characterization, and responsiveness of single primitive CD34 human umbilical cord blood hematopoietic progenitors with high proliferative and replating potential.* Blood, 1993. **81**(1): p. 41-8.

116. Kita, K., et al., *Cord blood-derived hematopoietic stem/progenitor cells: current challenges in engraftment, infection, and ex vivo expansion.* Stem Cells Int, 2011. **2011**: p. 276193.

117. da Silva, C.L., et al., *Differences amid bone marrow and cord blood hematopoietic stem/progenitor cell division kinetics.* J Cell Physiol, 2009. **220**(1): p. 102-11.

118. Wang, J.C., 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. 3919-24.

119. Theunissen, K. and C.M. Verfaillie, *A multifactorial analysis of umbilical cord blood, adult bone marrow and mobilized peripheral blood progenitors using the improved ML-IC assay.* Experimental Hematology, 2005. **33**(2): p. 165-172.

120. Holyoake, T., *Functional differences between transplantable human hematopoietic stem cells from fetal liver, cord blood, and adult marrow.* Experimental Hematology, 1999. **27**(9): p. 1418-1427.

121. Hugle, T. and T. Daikeler, *Stem cell transplantation for autoimmune diseases.* Haematologica, 2010. **95**(2): p. 185-8.

122. Yong, K.S.M., Z. Her, and Q. Chen, *Humanized Mice as Unique Tools for Human-Specific Studies.* Arch Immunol Ther Exp (Warsz), 2018. **66**(4): p. 245-266.

123. Shultz, L.D., et al., *Humanized mice for immune system investigation: progress, promise and challenges.* Nat Rev Immunol, 2012. **12**(11): p. 786-98.

124. 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.

125. Mosier, D.E., et al., *Transfer of a functional human immune system to mice with severe combined immunodeficiency.* Nature, 1988. **335**(6187): p. 256-259.

126. Kamel-Reid, S. and J.E. Dick, *Engraftment of immune-deficient mice with human hematopoietic stem cells.* Science, 1988. **242**(4886): p. 1706-9.

127. Audige, A., et al., *Long-term leukocyte reconstitution in NSG mice transplanted with human cord blood hematopoietic stem and progenitor cells.* BMC Immunol, 2017. **18**(1): p. 28.

128. 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.

129. Shultz, L.D., et al., *Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells.* J Immunol, 2005. **174**(10): p. 6477-89.

130. Ito, M., et al., *NOD/SCID/gamma(c)(null) mouse: an excellent recipient mouse model for engraftment of human cells.* Blood, 2002. **100**(9): p. 3175-82.

131. Notta, F., S. Doulatov, and J.E. Dick, *Engraftment of human hematopoietic stem cells is more efficient in female NOD/SCID/IL-2Rgc-null recipients.* Blood, 2010. **115**(18): p. 3704-7.

132. Wunderlich, M., et al., *Improved multilineage human hematopoietic reconstitution and function in NSGS mice.* PLoS One, 2018. **13**(12): p. e0209034.

133. Billerbeck, E., et al., *Development of human CD4+FoxP3+ regulatory T cells in human stem cell factor-, granulocyte-macrophage colony-stimulating factor-, and interleukin-3-expressing NOD-SCID IL2Rgamma(null) humanized mice.* Blood, 2011. **117**(11): p. 3076-86.

134. Wunderlich, M., et al., *AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3.* Leukemia, 2010. **24**(10): p. 1785-8.

135. Takagi, S., et al., *Membrane-bound human SCF/KL promotes in vivo human hematopoietic engraftment and myeloid differentiation.* Blood, 2012. **119**(12): p. 2768-77.

136. 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. 341-7.

137. Traggiai, E., et al., *Development of a human adaptive immune system in cord blood cell-transplanted mice.* Science, 2004. **304**(5667): p. 104-7.

138. Rongvaux, A., et al., *Development and function of human innate immune cells in a humanized mouse model.* Nat Biotechnol, 2014. **32**(4): p. 364-72.

139. Reinisch, A., et al., *Epigenetic and in vivo comparison of diverse MSC sources reveals an endochondral signature for human hematopoietic niche formation.* Blood, 2015. **125**(2): p. 249-60.

140. Reinisch, A., et al., *A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells.* Nat Med, 2016. **22**(7): p. 812-21.

141. Antonelli, A., et al., *Establishing human leukemia xenograft mouse models by implanting human bone marrow-like scaffold-based niches.* Blood, 2016. **128**(25): p. 2949-2959.

142. Chen, Y., et al., *Human extramedullary bone marrow in mice: a novel in vivo model of genetically controlled hematopoietic microenvironment.* Blood, 2012. **119**(21): p. 4971-80.

143. Reinisch, A., et al., *Generation and use of a humanized bone-marrow-ossicle niche for hematopoietic xenotransplantation into mice.* Nat Protoc, 2017. **12**(10): p. 2169-2188.

144. Abarrategi, A., et al., *Modeling the human bone marrow niche in mice: From host bone marrow engraftment to bioengineering approaches.* J Exp Med, 2018. **215**(3): p. 729-743.

145. Pang, W.W., et al., *Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20012-7.

146. Dorrell, C., et al., *Expansion of human cord blood CD34(+)CD38(-) cells in ex vivo culture during retroviral transduction without a corresponding increase in SCID repopulating cell (SRC) frequency: dissociation of SRC phenotype and function.* Blood, 2000. **95**(1): p. 102-110.

147. Guenechea, G., et al., *Transduction of human CD34+ CD38- bone marrow and cord blood-derived SCID-repopulating cells with third-generation lentiviral vectors.* Mol Ther, 2000. **1**(6): p. 566-73.

148. McIntosh, B.E., et al., *Nonirradiated NOD,B6.SCID Il2rgamma-/- Kit(W41/W41) (NBSGW) mice support multilineage engraftment of human hematopoietic cells.* Stem Cell Reports, 2015. **4**(2): p. 171-80.

149. Waskow, C., et al., *Hematopoietic stem cell transplantation without irradiation.* Nat Methods, 2009. **6**(4): p. 267-9.

150. Wang, X. and I. Riviere, *Genetic Engineering and Manufacturing of Hematopoietic Stem Cells.* Mol Ther Methods Clin Dev, 2017. **5**: p. 96-105.

151. Milone, M.C. and U. O'Doherty, *Clinical use of lentiviral vectors.* Leukemia, 2018. **32**(7): p. 1529-1541.

152. Naldini, L., et al., *In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector.* Science, 1996. **272**(5259): p. 263-7.

153. Evans, J.T., et al., *Human cord blood CD34+CD38- cell transduction via lentivirus-based gene transfer vectors.* Hum Gene Ther, 1999. **10**(9): p. 1479-89.

154. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration.* Nat Biotechnol, 2006. **24**(6): p. 687-96.

155. Mitchell, R.S., et al., *Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences.* PLoS Biol, 2004. **2**(8): p. E234.

156. De Palma, M., et al., *Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells.* Blood, 2005. **105**(6): p. 2307-15.

157. Cattoglio, C., et al., *Hot spots of retroviral integration in human CD34+ hematopoietic cells.* Blood, 2007. **110**(6): p. 1770-1778.

158. Hacein-Bey-Abina, S., et al., *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.* Science, 2003. **302**(5644): p. 415-9.

159. Gandara, C., V. Affleck, and E.A. Stoll, *Manufacture of Third-Generation Lentivirus for Preclinical Use, with Process Development Considerations for Translation to Good Manufacturing Practice.* Hum Gene Ther Methods, 2018. **29**(1): p. 1-15.

160. Milone, M.C. and U. O'Doherty, *Clinical use of lentiviral vectors.* Leukemia, 2018. **32**(7): p. 1529-1541.

161. Millington, M., et al., *Towards a clinically relevant lentiviral transduction protocol for primary human CD34 hematopoietic stem/progenitor cells.* PLoS One, 2009. **4**(7): p. e6461.

162. Uchida, N., et al., *Optimal conditions for lentiviral transduction of engrafting human CD34+ cells.* Gene Ther, 2011. **18**(11): p. 1078-86.

163. Park, S.W., C.-W. Pyo, and S.-Y. Choi, *High-efficiency lentiviral transduction of primary human CD34+ hematopoietic cells with low-dose viral inocula.* Biotechnology Letters, 2015. **37**(2): p. 281-288.

164. Hoffman, R., et al., *The in vitro and in vivo effects of stem cell factor on human hematopoiesis.* Stem Cells, 1993. **11 Suppl 2**(S2): p. 76-82.

165. Namikawa, R., M.O. Muench, and M.G. Roncarolo, *Regulatory roles of the ligand for Flk2/Flt3 tyrosine kinase receptor on human hematopoiesis.* Stem Cells, 1996. **14**(4): p. 388-95.

166. Yagi, M., et al., *Sustained ex vivo expansion of hematopoietic stem cells mediated by thrombopoietin.* Proc Natl Acad Sci U S A, 1999. **96**(14): p. 8126-31.

167. Zielske, S.P. and S.L. Gerson, *Cytokines, including stem cell factor alone, enhance lentiviral transduction in nondividing human LTCIC and NOD/SCID repopulating cells.* Mol Ther, 2003. **7**(3): p. 325-33.

168. Sutton, R.E., et al., *Transduction of human progenitor hematopoietic stem cells by human immunodeficiency virus type 1-based vectors is cell cycle dependent.* J Virol, 1999. **73**(5): p. 3649-60.

169. Chao, M.P., J. Seita, and I.L. Weissman, *Establishment of a normal hematopoietic and leukemia stem cell hierarchy.* Cold Spring Harb Symp Quant Biol, 2008. **73**: p. 439-49.

170. Bryder, D., D.J. Rossi, and I.L. Weissman, *Hematopoietic stem cells: the paradigmatic tissue-specific stem cell.* Am J Pathol, 2006. **169**(2): p. 338-46.

171. Stelzer, G.T., K.E. Shults, and M.R. Loken, *CD45 gating for routine flow cytometric analysis of human bone marrow specimens.* Ann N Y Acad Sci, 1993. **677**(1): p. 265-80.

172. Shah, V.O., C.I. Civin, and M.R. Loken, *Flow cytometric analysis of human bone marrow. IV. Differential quantitative expression of T-200 common leukocyte antigen during normal hemopoiesis.* J Immunol, 1988. **140**(6): p. 1861-7.

173. Kawamoto, H. and N. Minato, *Myeloid cells.* Int J Biochem Cell Biol, 2004. **36**(8): p. 1374-9.

174. Favaloro, E.J., et al., *Further characterization of human myeloid antigens (gp160,95; gp150; gp67): investigation of epitopic heterogeneity and non-haemopoietic distribution using panels of monoclonal antibodies belonging to CD-11b, CD-13 and CD-33.* Br J Haematol, 1988. **69**(2): p. 163-71.

175. Linden, M., J.M. Ward, and S. Cherian, *Hematopoietic and Lymphoid Tissues*, in *Comparative Anatomy and Histology*, P.M. Treuting and S.M. Dintzis, Editors. 2012, Academic Press: San Diego. p. 309-338.

176. Ekpenyong, A.E., et al., *Mechanical deformation induces depolarization of neutrophils.* Sci Adv, 2017. **3**(6): p. e1602536.

177. Loken, M.R., et al., *Flow cytometric analysis of human bone marrow. II. Normal B lymphocyte development.* Blood, 1987. **70**(5): p. 1316-24.

178. Lucio, P., et al., *Flow cytometric analysis of normal B cell differentiation: a frame of reference for the detection of minimal residual disease in precursor-B-ALL.* Leukemia, 1999. **13**(3): p. 419-27.

179. Samstag, Y., F. Emmrich, and T. Staehelin, *Activation of human T lymphocytes: differential effects of CD3- and CD8-mediated signals.* Proc Natl Acad Sci U S A, 1988. **85**(24): p. 9689-93.

180. Nagler, A., et al., *Comparative studies of human FcRIII-positive and negative natural killer cells.* J Immunol, 1989. **143**(10): p. 3183-91.

181. Strokotov, D.I., et al., *Is there a difference between T- and B-lymphocyte morphology?* J Biomed Opt, 2009. **14**(6): p. 064036.

182. Noetzli, L.J., S.L. French, and K.R. Machlus, *New Insights Into the Differentiation of Megakaryocytes From Hematopoietic Progenitors.* Arterioscler Thromb Vasc Biol, 2019. **39**(7): p. 1288-1300.

183. Mazzi, S., et al., *Megakaryocyte and polyploidization.* Exp Hematol, 2018. **57**: p. 1-13.

184. Lu, S.-J., et al., *Blood Components from Pluripotent Stem Cells*, in *Principles of Tissue Engineering*, R. Lanza, R. Langer, and J. Vacanti, Editors. 2014, Academic Press: Boston. p. 1041-1057.

185. Miyawaki, K., et al., *Identification of unipotent megakaryocyte progenitors in human hematopoiesis.* Blood, 2017. **129**(25): p. 3332-3343.

186. Loken, M.R., et al., *Flow cytometric analysis of human bone marrow: I. Normal erythroid development.* Blood, 1987. **69**(1): p. 255-63.

187. Yeo, J.H., Y.W. Lam, and S.T. Fraser, *Cellular dynamics of mammalian red blood cell production in the erythroblastic island niche.* Biophys Rev, 2019. **11**(6): p. 873-894.

188. de Vries, E., et al., *Correction for erythroid cell contamination in microassay for immunophenotyping of neonatal lymphocytes.* Arch Dis Child Fetal Neonatal Ed, 1999. **80**(3): p. F226-9.

189. 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.

190. Akashi, K., et al., *A clonogenic common myeloid progenitor that gives rise to all myeloid lineages.* Nature, 2000. **404**(6774): p. 193-7.

191. Manz, M.G., et al., *Prospective isolation of human clonogenic common myeloid progenitors.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11872-7.

192. Adolfsson, J., et al., *Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment.* Cell, 2005. **121**(2): p. 295-306.

193. Mansson, R., et al., *Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors.* Immunity, 2007. **26**(4): p. 407-19.

194. Pronk, C.J., et al., *Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy.* Cell Stem Cell, 2007. **1**(4): p. 428-42.

195. Karamitros, D., et al., *Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells.* Nat Immunol, 2018. **19**(1): p. 85-97.

196. Goodell, M.A., et al., *Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo.* J Exp Med, 1996. **183**(4): p. 1797-806.

197. Muller-Sieburg, C.E., et al., *Stem cell heterogeneity: implications for aging and regenerative medicine.* Blood, 2012. **119**(17): p. 3900-7.

198. Ema, H., et al., *Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice.* Dev Cell, 2005. **8**(6): p. 907-14.

199. Morita, Y., H. Ema, and H. Nakauchi, *Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment.* J Exp Med, 2010. **207**(6): p. 1173-82.

200. Dykstra, B., et al., *Long-term propagation of distinct hematopoietic differentiation programs in vivo.* Cell Stem Cell, 2007. **1**(2): p. 218-29.

201. Muller-Sieburg, C.E., et al., *Deterministic regulation of hematopoietic stem cell self-renewal and differentiation.* Blood, 2002. **100**(4): p. 1302-9.

202. Carrelha, J., et al., *Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells.* Nature, 2018. **554**(7690): p. 106-111.

203. Sanjuan-Pla, A., et al., *Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy.* Nature, 2013. **502**(7470): p. 232-6.

204. Yamamoto, R., et al., *Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells.* Cell, 2013. **154**(5): p. 1112-1126.

205. Belluschi, S., et al., *Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors.* Nat Commun, 2018. **9**(1): p. 4100.

206. Sieburg, H.B., et al., *The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets.* Blood, 2006. **107**(6): p. 2311-6.

207. 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.

208. Cho, R.H., H.B. Sieburg, and C.E. Muller-Sieburg, *A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells.* Blood, 2008. **111**(12): p. 5553-61.

209. Beerman, I., et al., *Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion.* Proc Natl Acad Sci U S A, 2010. **107**(12): p. 5465-70.

210. Muller-Sieburg, C.E. and H.B. Sieburg, *The GOD of hematopoietic stem cells: a clonal diversity model of the stem cell compartment.* Cell Cycle, 2006. **5**(4): p. 394-8.

211. Denis, C., et al., *A mouse model of severe von Willebrand disease: defects in hemostasis and thrombosis.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9524-9.

212. Grover, A., et al., *Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells.* Nat Commun, 2016. **7**: p. 11075.

213. Haas, S., et al., *Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors.* Cell Stem Cell, 2015. **17**(4): p. 422-34.

214. Knapp, D., et al., *Single-cell analysis identifies a CD33(+) subset of human cord blood cells with high regenerative potential.* Nat Cell Biol, 2018. **20**(6): p. 710-720.

215. Schroeder, T., *Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior.* Cell Stem Cell, 2010. **6**(3): p. 203-7.

216. McKenzie, J.L., et al., *Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment.* Nat Immunol, 2006. **7**(11): p. 1225-33.

217. Carrelha, J., et al., *Single-cell lineage tracing approaches in hematology research: technical considerations.* Exp Hematol, 2020. **89**: p. 26-36.

218. Hsu, Y.C., *Theory and Practice of Lineage Tracing.* Stem Cells, 2015. **33**(11): p. 3197-204.

219. Fink, J., A. Andersson-Rolf, and B.K. Koo, *Adult stem cell lineage tracing and deep tissue imaging.* BMB Rep, 2015. **48**(12): p. 655-67.

220. Serbedzija, G.N., M. Bronner-Fraser, and S.E. Fraser, *A vital dye analysis of the timing and pathways of avian trunk neural crest cell migration.* Development, 1989. **106**(4): p. 809-16.

221. Woo, K. and S.E. Fraser, *Order and coherence in the fate map of the zebrafish nervous system.* Development, 1995. **121**(8): p. 2595-609.

222. Kretzschmar, K. and F.M. Watt, *Lineage tracing.* Cell, 2012. **148**(1-2): p. 33-45.

223. Vogt, W., *Gestaltungsanalyse am Amphibienkeim mit Örtlicher Vitalfärbung.* Wilhelm Roux Arch Entwickl Mech Org, 1929. **120**(1): p. 384-706.

224. Mitchell, J.B., et al., *Radiosensitization of hematopoietic precursor cells (CFUc) in glioblastoma patients receiving intermittent intravenous infusions of bromodeoxyuridine (BUdR).* Int J Radiat Oncol Biol Phys, 1983. **9**(4): p. 457-63.

225. Morstyn, G., et al., *In vivo incorporation of bromodeoxyuridine into proliferating cells in the marrow and its effects on granulocyte-macrophage progenitor cells.* Exp Hematol, 1985. **13**(4): p. 289-94.

226. Caldwell, M.A., X. He, and C.N. Svendsen, *5-Bromo-2'-deoxyuridine is selectively toxic to neuronal precursors in vitro.* Eur J Neurosci, 2005. **22**(11): p. 2965-70.

227. 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.

228. Ju, Y.S., et al., *Somatic mutations reveal asymmetric cellular dynamics in the early human embryo.* Nature, 2017. **543**(7647): p. 714-718.

229. Lodato, M.A., et al., *Somatic mutation in single human neurons tracks developmental and transcriptional history.* Science, 2015. **350**(6256): p. 94-98.

230. Lee-Six, H., et al., *Population dynamics of normal human blood inferred from somatic mutations.* Nature, 2018. **561**(7724): p. 473-478.

231. Behjati, S., et al., *Genome sequencing of normal cells reveals developmental lineages and mutational processes.* Nature, 2014. **513**(7518): p. 422-425.

232. Shlush, L.I., et al., *Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability.* Blood, 2012. **120**(3): p. 603-12.

233. Bolli, N., et al., *Heterogeneity of genomic evolution and mutational profiles in multiple myeloma.* Nat Commun, 2014. **5**(1): p. 2997.

234. Chu, W.K., et al., *Ultraaccurate genome sequencing and haplotyping of single human cells.* Proc Natl Acad Sci U S A, 2017. **114**(47): p. 12512-12517.

235. Biezuner, T., et al., *A generic, cost-effective, and scalable cell lineage analysis platform.* Genome Res, 2016. **26**(11): p. 1588-1599.

236. Ludwig, L.S., et al., *Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics.* Cell, 2019. **176**(6): p. 1325-1339 e22.

237. Cornils, K., et al., *Multiplexing clonality: combining RGB marking and genetic barcoding.* Nucleic Acids Res, 2014. **42**(7): p. e56.

238. Livet, J., et al., *Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system.* Nature, 2007. **450**(7166): p. 56-62.

239. de Roo, J.J., et al., *Development of an in vivo model to study clonal lineage relationships in hematopoietic cells using Brainbow2.1/Confetti mice.* Future Sci OA, 2019. **5**(10): p. FSO427.

240. Kustikova, O., et al., *Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking.* Science, 2005. **308**(5725): p. 1171-4.

241. Maetzig, T., et al., *Polyclonal fluctuation of lentiviral vector-transduced and expanded murine hematopoietic stem cells.* Blood, 2011. **117**(11): p. 3053-64.

242. Pei, W., et al., *Polylox barcoding reveals haematopoietic stem cell fates realized in vivo.* Nature, 2017. **548**(7668): p. 456-460.

243. Pei, W., et al., *Resolving fate and transcriptome of hematopoietic stem cell clones.* bioRxiv, 2020: p. 2020.03.25.008433.

244. 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**(6): p. 1410-1422 e27.

245. McKenna, A., et al., *Whole-organism lineage tracing by combinatorial and cumulative genome editing.* Science, 2016. **353**(6298): p. aaf7907.

246. Sun, J., et al., *Clonal dynamics of native haematopoiesis.* Nature, 2014. **514**(7522): p. 322-7.

247. Rodriguez-Fraticelli, A.E., et al., *Clonal analysis of lineage fate in native haematopoiesis.* Nature, 2018. **553**(7687): p. 212-216.

248. Belderbos, M.E., et al., *Donor-to-Donor Heterogeneity in the Clonal Dynamics of Transplanted HumanCord Blood Stem Cellsin Murine Xenografts.* Biol Blood Marrow Transplant, 2020. **26**(1): p. 16-25.

249. Verovskaya, E., et al., *Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding.* Blood, 2013. **122**(4): p. 523-32.

250. Gerrits, A., et al., *Cellular barcoding tool for clonal analysis in the hematopoietic system.* Blood, 2010. **115**(13): p. 2610-8.

251. Biasco, L., et al., *In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases.* Cell Stem Cell, 2016. **19**(1): p. 107-19.

252. Wang, G.P., et al., *Dynamics of gene-modified progenitor cells analyzed by tracking retroviral integration sites in a human SCID-X1 gene therapy trial.* Blood, 2010. **115**(22): p. 4356-66.

253. Niederer, H.A. and C.R. Bangham, *Integration site and clonal expansion in human chronic retroviral infection and gene therapy.* Viruses, 2014. **6**(11): p. 4140-64.

254. Bystrykh, L.V., et al., *Counting stem cells: methodological constraints.* Nat Methods, 2012. **9**(6): p. 567-74.

255. Pei, W., et al., *Using Cre-recombinase-driven Polylox barcoding for in vivo fate mapping in mice.* Nat Protoc, 2019. **14**(6): p. 1820-1840.

256. 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.

257. Bystrykh, L.V., G. de Haan, and E. Verovskaya, *Barcoded Vector Libraries and Retroviral or Lentiviral Barcoding of Hematopoietic Stem Cells*, in *Hematopoietic Stem Cell Protocols*, K.D. Bunting and C.-K. Qu, Editors. 2014, Springer New York: New York, NY. p. 345-360.

258. Perie, L., et al., *Determining lineage pathways from cellular barcoding experiments.* Cell Rep, 2014. **6**(4): p. 617-24.

259. Schepers, K., et al., *Dissecting T cell lineage relationships by cellular barcoding.* J Exp Med, 2008. **205**(10): p. 2309-18.

260. Porter, S.N., et al., *Lentiviral and targeted cellular barcoding reveals ongoing clonal dynamics of cell lines in vitro and in vivo.* Genome Biol, 2014. **15**(5): p. R75.

261. Bhang, H.E., et al., *Studying clonal dynamics in response to cancer therapy using high-complexity barcoding.* Nat Med, 2015. **21**(5): p. 440-8.

262. Nguyen, L.V., et al., *DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts.* Nat Commun, 2014. **5**: p. 5871.

263. Lu, R., et al., *Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding.* Nat Biotechnol, 2011. **29**(10): p. 928-33.

264. Verovskaya, E., et al., *Asymmetry in skeletal distribution of mouse hematopoietic stem cell clones and their equilibration by mobilizing cytokines.* J Exp Med, 2014. **211**(3): p. 487-97.

265. Kristiansen, T.A., et al., *Cellular Barcoding Links B-1a B Cell Potential to a Fetal Hematopoietic Stem Cell State at the Single-Cell Level.* Immunity, 2016. **45**(2): p. 346-57.

266. Lu, R., et al., *Clonal-level lineage commitment pathways of hematopoietic stem cells in vivo.* Proc Natl Acad Sci U S A, 2019. **116**(4): p. 1447-1456.

267. Naik, S.H., et al., *Diverse and heritable lineage imprinting of early haematopoietic progenitors.* Nature, 2013. **496**(7444): p. 229-32.

268. Perie, L., et al., *The Branching Point in Erythro-Myeloid Differentiation.* Cell, 2015. **163**(7): p. 1655-62.

269. Wu, C., et al., *Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells.* Cell Stem Cell, 2014. **14**(4): p. 486-499.

270. Kim, S., et al., *Dynamics of HSPC repopulation in nonhuman primates revealed by a decade-long clonal-tracking study.* Cell Stem Cell, 2014. **14**(4): p. 473-85.

271. Yu, K.R., et al., *The Impact of Aging of Hematopoietic Stem and Progenitor Cells (HSPCs) in Non-Human Primates As Interrogated By Genetic Barcode Clonal Tracking.* Blood, 2015. **126**(23): p. 1151.

272. Yu, K.R., et al., *The impact of aging on primate hematopoiesis as interrogated by clonal tracking.* Blood, 2018. **131**(11): p. 1195-1205.

273. Espinoza, D.A., et al., *Aberrant Clonal Hematopoiesis following Lentiviral Vector Transduction of HSPCs in a Rhesus Macaque.* Mol Ther, 2019. **27**(6): p. 1074-1086.

274. Belderbos, M.E., et al., *Clonal selection and asymmetric distribution of human leukemia in murine xenografts revealed by cellular barcoding.* Blood, 2017. **129**(24): p. 3210-3220.

275. Maetzig, T., et al., *Lentiviral Fluorescent Genetic Barcoding for Multiplex Fate Tracking of Leukemic Cells.* Mol Ther Methods Clin Dev, 2017. **6**: p. 54-65.

276. Spanjaard, B., et al., *Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars.* Nat Biotechnol, 2018. **36**(5): p. 469-473.

277. Alemany, A., et al., *Whole-organism clone tracing using single-cell sequencing.* Nature, 2018. **556**(7699): p. 108-112.

278. Weinreb, C., et al., *Lineage tracing on transcriptional landscapes links state to fate during differentiation.* Science, 2020. **367**(6479): p. eaaw3381.

279. Rodriguez-Fraticelli, A.E., et al., *Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis.* Nature, 2020. **583**(7817): p. 585-589.

280. Lyne, A.M., et al., *A track of the clones: new developments in cellular barcoding.* Exp Hematol, 2018. **68**: p. 15-20.

281. Kalia, A. and R.P. Gupta, *Proteomics: a paradigm shift.* Crit Rev Biotechnol, 2005. **25**(4): p. 173-98.

282. Frith, M.C., M. Pheasant, and J.S. Mattick, *The amazing complexity of the human transcriptome.* Eur J Hum Genet, 2005. **13**(8): p. 894-7.

283. Chen, M. and J.L. Manley, *Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.* Nat Rev Mol Cell Biol, 2009. **10**(11): p. 741-54.

284. Tian, B. and J.L. Manley, *Alternative cleavage and polyadenylation: the long and short of it.* Trends Biochem Sci, 2013. **38**(6): p. 312-20.

285. Velten, L., et al., *Human haematopoietic stem cell lineage commitment is a continuous process.* Nat Cell Biol, 2017. **19**(4): p. 271-281.

286. Pellin, D., et al., *A comprehensive single cell transcriptional landscape of human hematopoietic progenitors.* Nat Commun, 2019. **10**(1): p. 2395.

287. Hay, S.B., et al., *The Human Cell Atlas bone marrow single-cell interactive web portal.* Exp Hematol, 2018. **68**: p. 51-61.

288. Chen, L., et al., *Transcriptional diversity during lineage commitment of human blood progenitors.* Science, 2014. **345**(6204): p. 1251033.

289. Novershtern, N., et al., *Densely interconnected transcriptional circuits control cell states in human hematopoiesis.* Cell, 2011. **144**(2): p. 296-309.

290. Tikhonova, A.N., et al., *The bone marrow microenvironment at single-cell resolution.* Nature, 2019. **569**(7755): p. 222-228.

291. Ziegenhain, C., et al., *Comparative Analysis of Single-Cell RNA Sequencing Methods.* Mol Cell, 2017. **65**(4): p. 631-643 e4.

292. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**(1): p. 14049.

293. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.

294. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.

295. Watcham, S., I. Kucinski, and B. Gottgens, *New insights into hematopoietic differentiation landscapes from single-cell RNA sequencing.* Blood, 2019. **133**(13): p. 1415-1426.

296. Zhang, X., et al., *Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems.* Mol Cell, 2019. **73**(1): p. 130-142 e5.

297. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2.* Nat Protoc, 2014. **9**(1): p. 171-81.

298. Hashimshony, T., et al., *CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq.* Genome Biol, 2016. **17**(1): p. 77.

299. Bagnoli, J.W., et al., *Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq.* Nat Commun, 2018. **9**(1): p. 2937.

300. Picelli, S., *Single-cell RNA-sequencing: The future of genome biology is now.* RNA Biol, 2017. **14**(5): p. 637-650.

301. Svensson, V., et al., *Power analysis of single-cell RNA-sequencing experiments.* Nat Methods, 2017. **14**(4): p. 381-387.

302. Chen, G., B. Ning, and T. Shi, *Single-Cell RNA-Seq Technologies and Related Computational Data Analysis.* Front Genet, 2019. **10**: p. 317.

303. Picelli, S., *Full-Length Single-Cell RNA Sequencing with Smart-seq2*, in *Single Cell Methods: Sequencing and Proteomics*, V. Proserpio, Editor. 2019, Springer New York: New York, NY. p. 25-44.

304. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nat Methods, 2013. **10**(11): p. 1096-8.

305. Giustacchini, A., et al., *Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia.* Nat Med, 2017. **23**(6): p. 692-702.

306. Rodriguez-Meira, A., et al., *Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing.* Mol Cell, 2019. **73**(6): p. 1292-1305 e8.

307. Wang, Z., et al., *Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies.* Signal Transduct Target Ther, 2021. **6**(1): p. 24.

308. Paul, F., et al., *Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.* Cell, 2015. **163**(7): p. 1663-77.

309. Giladi, A., et al., *Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis.* Nat Cell Biol, 2018. **20**(7): p. 836-846.

310. Drissen, R., et al., *Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing.* Nature Immunology, 2016. **17**: p. 666.

311. Fujiwara, Y., et al., *Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12355-8.

312. Shivdasani, R.A., et al., *A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development.* EMBO J, 1997. **16**(13): p. 3965-73.

313. Migliaccio, A.R., et al., *GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant.* J Exp Med, 2003. **197**(3): p. 281-96.

314. Nei, Y., et al., *GATA-1 regulates the generation and function of basophils.* Proc Natl Acad Sci U S A, 2013. **110**(46): p. 18620-5.

315. Yu, C., et al., *Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo.* J Exp Med, 2002. **195**(11): p. 1387-95.

316. Jacobsen, S.E.W. and C. Nerlov, *Haematopoiesis in the era of advanced single-cell technologies.* Nat Cell Biol, 2019. **21**(1): p. 2-8.

317. Drissen, R., et al., *Identification of two distinct pathways of human myelopoiesis.* Sci Immunol, 2019. **4**(35): p. eaau7148.

318. Guilliams, M., A. Mildner, and S. Yona, *Developmental and Functional Heterogeneity of Monocytes.* Immunity, 2018. **49**(4): p. 595-613.

319. Xie, X., et al., *Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection.* Nat Immunol, 2020. **21**(9): p. 1119-1133.

320. Zheng, S., et al., *Molecular transitions in early progenitors during human cord blood hematopoiesis.* Mol Syst Biol, 2018. **14**(3): p. e8041.

321. Mazurier, F., et al., *Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells.* Nat Med, 2003. **9**(7): p. 959-63.

322. Krueger, F. *TrimGalore!* 2012; Available from: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore.

323. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

324. Liao, Y., G.K. Smyth, and W. Shi, *The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote.* Nucleic Acids Res, 2013. **41**(10): p. e108.

325. Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a single report.* Bioinformatics, 2016. **32**(19): p. 3047-8.

326. Satija, R., et al., *Spatial reconstruction of single-cell gene expression data.* Nat Biotechnol, 2015. **33**(5): p. 495-502.

327. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society Series B-Statistical Methodology, 1995. **57**(1): p. 289-300.

328. Wilcoxon, F., *Individual Comparisons by Ranking Methods.* Biometrics Bulletin, 1945. **1**(6): p. 80-83.

329. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.

330. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.

331. Xie, S.Z., et al., *Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal.* Cell Stem Cell, 2019. **25**(5): p. 639-653 e7.

332. Kowalczyk, M.S., et al., *Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells.* Genome Res, 2015. **25**(12): p. 1860-72.

333. Durinck, S., et al., *BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis.* Bioinformatics, 2005. **21**(16): p. 3439-3440.

334. Durinck, S., et al., *Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt.* Nature Protocols, 2009. **4**(8): p. 1184-1191.

335. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.

336. Mazurier, F., et al., *Lentivector-mediated clonal tracking reveals intrinsic heterogeneity in the human hematopoietic stem cell compartment and culture-induced stem cell impairment.* Blood, 2004. **103**(2): p. 545-52.

337. Dickson, K.A., M.C. Haigis, and R.T. Raines, *Ribonuclease inhibitor: structure and function.* Prog Nucleic Acid Res Mol Biol, 2005. **80**: p. 349-74.

338. Berardi, A.C., et al., *Functional isolation and characterization of human hematopoietic stem cells.* Science, 1995. **267**(5194): p. 104-8.

339. Shariatmadar, S., et al., *Electronic volume of CD34 positive cells from peripheral blood apheresis samples.* Cytometry B Clin Cytom, 2008. **74**(3): p. 182-8.

340. Rosales, C., *Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types?* Front Physiol, 2018. **9**(113): p. 113.

341. Ng, Y.Y., et al., *Isolation of human and mouse hematopoietic stem cells.* Methods Mol Biol, 2009. **506**: p. 13-21.

342. Uchida, N., et al., *High-Efficiency Lentiviral Transduction of Human CD34(+) Cells in High-Density Culture with Poloxamer and Prostaglandin E2.* Mol Ther Methods Clin Dev, 2019. **13**: p. 187-196.

343. Yahata, T., et al., *A highly sensitive strategy for SCID-repopulating cell assay by direct injection of primitive human hematopoietic cells into NOD/SCID mice bone marrow.* Blood, 2003. **101**(8): p. 2905-13.

344. McKenzie, J.L., et al., *Human short-term repopulating stem cells are efficiently detected following intrafemoral transplantation into NOD/SCID recipients depleted of CD122⁺ cells.* Blood, 2005. **106**(4): p. 1259.

345. Rodriguez-Meira, A., et al., *Unraveling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA-sequencing.* bioRxiv, 2018.

346. Rodriguez-Meira, A., et al., *TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing.* STAR Protoc, 2020. **1**(3): p. 100125.

347. Laurenti, E. and J.E. Dick, *Molecular and functional characterization of early human hematopoiesis.* Ann N Y Acad Sci, 2012. **1266**(1): p. 68-71.

348. Stegelmeier, A.A., et al., *Myeloid Cells during Viral Infections and Inflammation.* Viruses, 2019. **11**(2): p. 168.

349. Rodriguez-Fraticelli, A.E. and F. Camargo, *Systems analysis of hematopoiesis using single-cell lineage tracing.* Current Opinion in Hematology, 2021. **28**(1).

350. Vitali, C., et al., *SOCS2 Controls Proliferation and Stemness of Hematopoietic Cells under Stress Conditions and Its Deregulation Marks Unfavorable Acute Leukemias.* Cancer Res, 2015. **75**(11): p. 2387-99.

351. Kataoka, K., et al., *Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity.* J Exp Med, 2011. **208**(12): p. 2403-16.

352. Calvanese, V., et al., *MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment.* Nature, 2019. **576**(7786): p. 281-286.

353. Komorowska, K., et al., *Hepatic Leukemia Factor Maintains Quiescence of Hematopoietic Stem Cells and Protects the Stem Cell Pool during Regeneration.* Cell Rep, 2017. **21**(12): p. 3514-3523.

354. Gazit, R., et al., *Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells.* Stem Cell Reports, 2013. **1**(3): p. 266-80.

355. Becker-Herman, S., et al., *CD74 is a regulator of hematopoietic stem cell maintenance.* PLoS Biol, 2021. **19**(3): p. e3001121.

356. Pina, C., et al., *MLLT3 regulates early human erythroid and megakaryocytic cell fate.* Cell Stem Cell, 2008. **2**(3): p. 264-73.

357. Mayer, B., et al., *Vasopressin stimulates the proliferation and differentiation of red blood cell precursors and improves recovery from anemia.* Sci Transl Med, 2017. **9**(418): p. eaao1632.

358. Cabezas-Wallscheid, N., et al., *Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis.* Cell Stem Cell, 2014. **15**(4): p. 507-522.

359. Grandi, A., et al., *ERMP1, a novel potential oncogene involved in UPR and oxidative stress defense, is highly expressed in human cancer.* Oncotarget, 2016. **7**(39): p. 63596-63610.

360. Lin, K.K., et al., *CD81 is essential for the re-entry of hematopoietic stem cells to quiescence following stress-induced proliferation via deactivation of the Akt pathway.* PLoS Biol, 2011. **9**(9): p. e1001148.

361. Cabezas-Wallscheid, N., et al., *Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy.* Cell, 2017. **169**(5): p. 807-823 e19.

362. Sun, D., et al., *Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal.* Cell Stem Cell, 2014. **14**(5): p. 673-88.

363. Ito, K., et al., *A PML-PPAR-delta pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance.* Nat Med, 2012. **18**(9): p. 1350-8.

364. Papa, L., M. Djedaini, and R. Hoffman, *Mitochondrial Role in Stemness and Differentiation of Hematopoietic Stem Cells.* Stem Cells Int, 2019. **2019**: p. 4067162.

365. van Galen, P., et al., *The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress.* Nature, 2014. **510**(7504): p. 268-72.

366. Sigurdsson, V. and K. Miharada, *Regulation of unfolded protein response in hematopoietic stem cells.* Int J Hematol, 2018. **107**(6): p. 627-633.

367. Suda, T., K. Takubo, and G.L. Semenza, *Metabolic regulation of hematopoietic stem cells in the hypoxic niche.* Cell Stem Cell, 2011. **9**(4): p. 298-310.

368. Lee-Six, H. and D.G. Kent, *Tracking hematopoietic stem cells and their progeny using whole-genome sequencing.* Exp Hematol, 2020. **83**: p. 12-24.

369. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**(1): p. 21 29 1-21 29 9.

370. de Haan, G. and S.S. Lazare, *Aging of hematopoietic stem cells.* Blood, 2018. **131**(5): p. 479-487.

371. Kuribayashi, W., et al., *Limited rejuvenation of aged hematopoietic stem cells in young bone marrow niche.* Journal of Experimental Medicine, 2020. **218**(3).

372. Crisan, M. and E. Dzierzak, *The many faces of hematopoietic stem cell heterogeneity.* Development, 2016. **143**(24): p. 4571-4581.