Human Haematopoietic Stem Cell Heterogeneity in Postnatal Haematopoiesis and Ontogeny

Semiramis Popova

University College London

and

The Francis Crick Institute

PhD Supervisor: Prof. Dominique Bonnet

Secondary Supervisor: Dr. Owen Williams

A thesis submitted for the degree of

Doctor of Philosophy

University College London

March 2021

Declaration

I Semiramis Popova confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Haematopoietic stem cell (HSC) transplants are upheld as one of the most successful therapies in regenerative medicine. While improved purification and functionality assays have advanced understanding of steady-state haematopoiesis and the human *bona fide* HSC, evidence suggests significant heterogeneity exists within the HSC compartment in post-natal and pre-natal haematopoiesis.

In post-natal haematopoiesis, both $CD34^+$ and $CD34^-$ cells possess robust *in vivo* repopulating potential. $CD34^-$ repopulating cells, however, exhibit distinct repopulation kinetics, capacity to produce functional $CD34^+$ repopulating cells, and accordingly have been speculated to reside at the apex of the human haematopoietic hierarchy. But a low repopulating cell frequency has hindered efforts to study these HSCs. We thus aim to improve purification of $CD34^-$ HSCs and further expand the knowledge of this immature stem pool. We successfully identified an additional positive selection marker, CD117 (c-Kit). Through limiting dilution analysis and serial transplantations of enriched $CD34^-$ repopulating cells in enhanced NSG mouse models, we observed repopulation and lineage commitment kinetics. To investigate the molecular mechanisms, we conducted single-cell RNAseq. With these new data we have asserted the importance of human $CD34^-$ HSCs their enormous therapeutic potential.

In human foetal haematopoiesis, it is unknown whether $CD34^-$ repopulating cells emerge during ontogeny and play a role in foetal haematopoiesis. While reports for HSC-purifying markers have been produced, much of these studies have been restricted to foetal liver, a single gestational stage, and the $CD34^+$ population. In humans, little is known about how the HSC cell surface marker phenotype adapts to the dynamic niches in the liver during expansion, homing to the bone marrow, and bone marrow colonisation. To address this, we optimised a multi-parameter flow cytometry panel and used it to investigate the expression of a number of reported HSC cell surface markers across first and second trimester liver and bone marrow. Through a combination of high-dimensional data analysis and mathematical modelling we have produced an antigen-based map of foetal haematopoietic stem cell and progenitor dynamics.

Impact Statement

The identification of HSCs was ultimately advanced by the application of flow cytometry and the use of purifying cell surfaces markers, which enabled the isolation of prospective HSC populations. Parallel developments in *in vitro* HSC functionality assays and immunodeficient mouse models led to more reliable readouts of HSC multilineage reconstitution and self-renewal capacity. In humans, the CD34 antigen was established early on as purification marker for HSCs. However, these findings were followed by reports that cells negative for CD34 expression are also capable of repopulating immunodeficient mice. Furthermore, $CD34^-$ HSCs were shown to be highly quiescent and able to produce functional $CD34^+$, placing $CD34^-$ HSC at the top of the human haematopoietic hierarchy. These findings presented $CD34^-$ HSCs as a population with enormous clinical potential; if they are capable of giving rise to $CD34^+$ HSC they can arguably be a better source for HSC expansion and haematopoietic stem cell transplants (HSCT). However, $CD34^-$ HSCs have proven to be challenging to study. Due to their poor *in vitro* differentiation and self-renewal capacity, their functionality can only be properly assayed using *in vivo* repopulation assays. But even *in vivo* repopulation studies pose a challenge as $CD34^-$ HSCs take approximately six weeks longer than $CD34^+$ HSCs to engraft in conditioned mouse recipients. On the other hand, any attempts to study the $CD34^-$ HSC population at the molecular level have also been hampered by the fact that it cannot be isolated at high purity.

In this thesis we present work that has successfully addressed some of the major challenges in studying $CD34^-$ HSCs. We have identified a new positive selection marker, CD117, which has enabled us to isolate $CD34^-$ HSCs at significantly higher purity. We have also identified new immunodeficient mouse strains that can better reveal the functionality of $CD34^-$ HSCs. We have also made advancements in elucidating the molecular profile of this population. Through the application of single cell transcriptomics we have revealed that the $CD34^-$ HSC population may represent a larger proportion of cells within the $CD34^-$ compartment than previously estimated by *in vivo* functionality readouts. We have also demonstrated that similarly to reports focused on $CD34^+$ HSCs, a level of heterogeneity may also exist among $CD34^-$ HSCs.

In this thesis we have also presented a strategy for the design and optimisation of high-parameter flow cytometry panel, which we successfully implemented to collect data on the cell surface marker expression dynamics in haematopoietic populations during ontogeny. Through clustering and dimensionality reduction analysis, we demonstrated that this can be a viable approach to study haematopoietic population dynamics during development. We also demonstrated that despite the limited number of variables measured compared to other single cell approaches, high-parameter flow cytometry can still be an effective tool in identifying candidate HSC populations.

Acknowledgement

First and foremost, I need to thank my supervisor, Prof. Dominique Bonnet. This project would have been impossible without her patience, encouragement and support. No matter how discouraging things got at times, I always left Dominque's office (after a meeting that usually overran by an hour) with recharged enthusiasm, a battle plan, and a head buzzing with ideas and possibilities. Regardless of coming grant deadlines, vivas, panels, conferences, she always made the time for a meeting, to read last-minute reports, or join for a cake to celebrate someone's birthday.

The past 4.5 years would have also been very different without the amazing Bonnet lab members, former and current. In addition to the hilarious banter which made coming into the office and working in tissue culture a pleasure, I was also exposed to amazing role models (both as human beings and scientists). The lessons I've learned from all of these guys while watching them tackle the highs and lows of a career in research or life in general could be the subject of a few PhDs in awesomeness. I'll be eternally grateful that when things got a bit bleak, I could go into work and count on a few laughs and the reminder that there are still things to celebrate. For that I also need to extend my gratitude to my friends, who were there in moments of need.

Strange, but very important is for me to also thank the NHS, which took amazing care of me when I needed it. For that, I'll be forever their champion.

Without the support family, I wouldn't have made it through school, let alone a PhD. Through the last five years, they've rolled with a few heavy punches, teaching me this and that about resiliency and taking it on the chin.

Table of Contents

Table of figures

List of tables

Abbreviations

AGM – Aorta-gonad-mesonephros

AML – Acute myeloid leukaemia

BM – Bone marrow

CFU – Colony-forming unit

CFU-S – Splenic colony-forming unit

CFU-C – Colony-forming unit culture

CLP – Common lymphoid progenitor

CMP – Common myeloid progenitor

EPO – Erythropoietin

ETP – Early thymic precursor

FACS – Fluorescence activated cell sorting

FBM – Foetal bone marrow

FL – Foetal liver

GM-CSF – Granulocyte/macrophage-colony stimulating factor

G-CSF – Granulocyte-colony stimulating factor

GMP – Granulocyte-monocyte progenitor

HSC – Haematopoietic stem cell

HSPC – Haematopoietic stem and progenitor cell

HSCT – Haematopoietic stem cell transplant

IAHC – Intra-aortic haematopoietic cluster

IL-3 – Interleukin-3

IV – Intravenously

LMPP – Lympho-myeloid primed progenitor

LSK – $Lin^{-}/Sca\text{-}1^{+}/Kit^{+}$

LTC-IC – Long-term cell – initiation cell

LT-HSC – Long-term HSC

MEP – Megakaryocyte-erythrocyte progenitor

Mk - Megakaryocyte

MLP – Multilymphoid progenitor

MDS – Multidimensional scaling

MNC – Mononuclear cell

MPP – Multipotent progenitor

NK (cell) – Natural killer cell

NOD – Non-obese diabetic

NSG - NOD/SCID/IL2Rγnull

NSG-S – NSG-SGM3

c-Kit^{Mutant} - NSG-Kit$^{W41/W41}$

PBS – Phosphate-buffered saline

PCA – Principal component analysis

SCID - Severe combined immune-deficient

SCF – Stem cell factor

SOM – Self-organising map

ST-HSC – Short-term HSC

SRC – SCID-repopulating cell

TPO - Thrombopoietin

UCB/CB – Umbilical cord blood

UMAP – Uniform manifold approximation and projection

Chapter 1. Introduction

Following the establishment of the cell as the basic biological unit, it was Walther Flemming's discovery of chromosomes and cell division that laid the basis for the theory of tissue regeneration in the late 19th century (Mozazarello et al., 2001). This exciting new field of investigation was bolstered by Giulio Bizzozero, who observed an amplified blood-cell-forming phenomenon in the bone marrow of pigeons and chickens following controlled bleeding (Mozazarello et al., 2001). Adopting Flemming's methods, Bizzozero went on to investigate the regenerative capabilities of multiple tissues and categorized his findings according to the rate of tissue turnover. To the group with the greatest proliferative capacity, termed "labile", he ascribed a number of glandular organs such as the gut, testicle, and ovary, but also the spleen and the bone marrow (Mozazarello et al., 2001).

In parallel, other 19th century scientists were putting forward the idea of a "stem cell". Earnst Heackel first proposed that all multicellular organisms originate from a single primordial cell. The theory was quickly expanded by Weissman and Boveri, who introduced the concept of segregation of the germ cells in the early embryo and the definition of non-committed cells in the zygote as stem cells. These early theories, however, were solely centred around embryonic development. It wasn't until decades later that the stem cell and Bizzozero's work on adult tissue regeneration would converge to establish the significance of the adult stem cell.

Haematology has been a trailblazing field that has benefited greatly from a productive interchange between clinical and basic biological research. It was before the existence of the haematopoietic stem cell (HSC) was experimentally confirmed that the first bone marrow transplants were being tested in the clinic to treat patients with haematological cancers (Thomas et al., 1957, 1959). While falling short of curative, these early clinical trials demonstrated the potency of the regenerative capacity of the bone marrow and its undisputed therapeutic potential. It was a few years later, that in seminal series of papers, Till and McCulloch provided categorical evidence of the existence of a bone marrow-

derived cells that could give rise to splenic colonies composed of both committed and immature haematopoietic cells (Becker et al., 1963; Till & McCulloch, 1961). Furthermore, their experiments involving the injection of donor mouse bone marrow cells into lethally irradiated recipient mice introduced the basic principles for assays, which would later become fundamental in stem cell and cancer research. In the following two decades splenic colony-forming units (CFU-S) became a standard *in vivo* test that also enabled the first attempts at estimating the stem cell frequency within a population (Becker et al., 1963; Till et al., 1964). However, the dozen CFU-S that usually formed per recipient were derivative of single cells, implying only a small fraction of the tested population were being captured in the assay. In a serial transplantation experiment of single splenic colonies, the majority of secondary recipients failed to form CFU-S, a few mice were observed to have one colony per spleen, while a limited number of mice developed over twenty colonies per spleen. The authors recognized that there appeared to be a hierarchy of blood-forming cells as well as the fact that the CFU-S was limited in its ability to study stem cell behaviour on a population level, as it was only able to capture progenies of single cells. Before single HSCs could be studied in isolation, the features and functions of the different haematopoietic compartments first had to be described. In addition, the short time frame of these experiments, only revealed multipotent progenitors, not true stem cells, which we now know need longer than two weeks to begin repopulating mouse recipients. Nonetheless, the CFU-S assay gave rise to its *in vitro* derivative, the colony-forming unit – culture (CFU-C), and set the basic tools that would help advance haemotopoietic research (Bradley & Metcalf, 1966; Pluznik & Sachs, 1965).

1.1 Mouse haematopoiesis

The search of murine HSCs started by separation of cells by size and density, and later with reagents against cell surface antigens such as MHC Class I molecules (Iscove, 1990; Visserc et al., 1984; Worton et al., 1969). But validation of potential HSC populations required assays that could test the different lineage outputs. Development of thymic assays allowed validation of T cell differentiation,

while the *in vitro* assays using bone marrow stroma could maintain B cells in long-term cultures (Ezine et al., 1984; C. A. Whitlock & Witte, 1982). While co-culture with stroma was allowing the long-term maintenance of haematopoietic cells *in vitro*, the ultimate function of HSCs was linked with *in vivo* reconstitution of mouse recipients and rescue from radiation damage (Cheryl A. Whitlock et al., 1987). Consequently two nearly identical strains with congenic alleles for CD45 were developed (F. W. Shen et al., 1986). The nearly identical genetic background removed the risk of graft rejection, while the monoclonal antibodies against the two different CD45 isoforms could distinguish native cells from donor transplant.These advances in assays quickly promoted the search for the murine HSC.

With the identification of monoclonal antibodies against mature lineages such as B220 (B cells), fluorescence-activated cell sorting allowed the subsetting of lineage-negative (Lin^-) cells (Coffman & Weissman, 1981). Thy-1, a marker, initially reported to be expressed by thymic lymphocytes was being shown to also identify cells with CFU-S capacity (Basch & Berman, 1982; Reif & Allen, 1966). It was later shown that Lin^- cells that also expressed low levels of Thy-1 (CD90) were enriched for bone marrow reconstituting cells (Muller-Sieburg et al., 1986). Monoclonal antibodies against Thy-1^{Low} cells identified Sca-1 (CD59) as another purification marker of haematopoietic cells with clonogenic and *in vivo* reconstitution capacity (Aihara et al., 1986; Gerald J. Spangrude et al., 1988). In parallel, Kit (CD117) expression identified a population of self-renewing cells with long-term *in vivo* reconstitution function, defining the true long-term cells as $Lin^-/Sca\text{-}1^+/Kit^+$ (LSK) (Ikuta & Weissman, 1992; Morrison & Weissman, 1994). But it wasn't until a few years later that murine HSC were finally purified. Limiting dilution analysis of $Lin^-/Thy^{Low}/Sca\text{-}1^+$ cells suggested that the population contained a mix of cells with both long-term and transient self-renewal and multilineage reconstitution potential, confirming the existence of both short-term HSCs and multipotent progenitors (Morrison et al., 1997).

Despite rapid improvements in flow cytometry technology and antibody availability, it was still only a fraction of the LSK population that were generating engraftment (G J Spangrude et al., 1995). Hence, the search for better HSC

markers continued. CD48 and CD150 of the SLAM receptor family were found to successfully differentiate between multipotent progenitors (MPPs) and long-term HSCs, with single LSK/CD48$^-$/CD150$^+$ cells generating long-term multilineage reconstitution (Sintes et al., 2008). Expression of SLAM receptors and Flt3 (Flk2, CD135) was later shown to delineate subsets of differentially primed MPPs with distinct reconstitution kinetics (Pietras et al., 2015). These subsets were also shown to be highly dynamic in their self-renewal and output during blood regeneration (Pietras et al., 2015). The pertinent question of whether differentiation occurs at the level of single HSCs or by successive steps of lineage restriction of increasingly mature progenitors was addressed by attempts to further subdivide the LSK compartment and dissection of populations outside it.

Building on reports of mice with impaired lymphopoiesis, CD127 (Interleukin-7 receptor subunit alpha) was used as a positive selection marker to isolate cells with T and B cell reconstitution capacity (Kondo et al., 1997). Interestingly, these common lymphoid progenitors (CLPs) were also defined by low C-Kit and Sca-1 expression, suggesting these HSC markers may be downregulated with cell fate commitment (Kondo et al., 1997). A few years later, the common myeloid progenitor (CMP) was reported to completely lack C-Kit and Sca-1 expression and be positive for CD34 and CD16/32 (Akashi et al., 2000). The lineage readout of CMPs *in vitro* was determined to be myeloid, but also megakaryocute (Mk) and erythroid (Akashi et al., 2000; Manz et al., 2002). Although the multilineage potential of CMPs was still questioned as colony-forming capacity (CFC) assays could not make the distinction between a homogenous multipotent population and a mix of unipotent progenitors, these findings formed the basis of the hypothesis that the first lineage restriction event in HSCs is the bifurcation of the myeloid and lymphoid lineages.

In parallel, search for markers that could further dissect the LSK compartment was also progressing. As well as the SLAM family of proteins, Flt3 became of interest as LSK/Flt3$^+$ cells were shown to have decreased self-renewal and increased proliferation (Adolfsson et al., 2001; Christensen & Weissman, 2001; Passegué et al., 2005). Furthermore, LSK/Flt3$^+$ cells could generate lymphoid and myeloid cells, confirming the existence of lymphoid-primed multipotent progenitor (LMPP) (Adolfsson et al., 2005). These LMPPs were initially reported to be devoid of erythroid and megakaryocyte (Mk) potential both *in vitro* and *in vivo* (Adolfsson et al.,

2005; Yang et al., 2005). This indicated that the first lineage specification event in haematopoiesis not only came earlier than previously suspected, but also could potentially be the priming of Mk/erythroid progenitors (MEPs) rather than specification of the myeloid and lymphoid lineages. But with improvement of *in vitro* assays and sampling schedules for *in vivo* experiments, it appeared that LMPPs defined by the LSK/Flt3^{+} phenotype could generate Mks and erythrocytes at comparable levels to short-term and long-term HSCs (SH-HSC, LT-HSCs) (Forsberg et al., 2006). Further attempts to subdivide the MPP compartment based on VCAM-1 expression, suggested that MEP activity might be restricted to FLT3^{Low}/VCAM-1^{+} cells, while loss of VCAM-1 and gain of Flt3 expression were associated with lymphoid priming (Lai & Kondo, 2006). However, an intermediate population positive for both markers was shown to promiscuously express low levels of both erythroid and lymphoid gene transcripts (Lai & Kondo, 2006). In the end, Flt3 expression didn't necessarily correlate with loss of multipotency, but these findings and the potent MEP activity in the LSK/Flt3^{-} compartment could be reconciled with the fact that the physiological demand for erythrocytes and platelets is significantly higher than that for lymphocytes and myeloid cells during steady-state haematopoiesis. Therefore, it is possible that several mechanisms exist to stimulate the production of MEPs to ensure that an organism's physiological requirements for red blood cells and platelets are met.

1.2 Construction of the human haematopoietic stem cell hierarchy

1.2.1 Adaptation of *in vitro* and *in vivo* assays

While studies of the murine haematopoietic system were rapidly advancing, research on human haematopoiesis was lagging behind. Bone marrow reconstitution studies and optimization of conditions for long-term cultures were driving forward the identification and characterization of murine HSCs, but these assays had to be adapted for research on human haematopoiesis. By evaluating

the efficiency of retroviral gene transfer which relies on active cell division for its integration, important cytokines, growth factors and bone marrow extracellular matrix molecules for the *in vitro* maintenance of human HSCs were identified (Dick et al., 1991; Hughes et al., 1989; Moritz et al., 1994). These findings led to improved colony-forming capacity (CFC) and long-term culture initiating cell (LTC-IC) assays that could be applied to testing the *in vitro* differentiation and self-renewal of potential human HSC populations. The LTC-IC assay, involving culturing of haematopoietic cells on bone marrow feeder cells, was used to identify haematopoietic cells capable of self-renewing for periods of 6 weeks or longer (Sutherland et al., 1989). LTC-IC assays were powerful tools in the identification not only of CD34 as positive human HSC marker, but also $CD90^{+}$ (Thy-1) and the $Lin^{-}/CD34^{+}/CD38^{-/Low}$ populations as sources of cells with potent self-renewing activity (Baum et al., 1992a; Kondo et al., 2003). However, the true test for an HSC, long-term bone marrow multilineage reconstitution, was proving to be more challenging and demanded the engineering of new mouse models.

The lack of mouse models that were immune-permissive to human grafts was addressed with advancements in genetic engineering and generation of mouse strains like the severe combined immune-deficient (SCID) and the beige/nude/xid mice (BNX). Independent groups began reporting the detection of circulating human lymphoid cells in the SCID strain and both lymphoid and myeloid cells in the BNX strain 9 weeks post transplantation of human bone marrow cells (Kamel-Reid & Dick, 1988; McCune et al., 1988). In these early experiments, it became evident that the more attenuated the host immune system, the superior the engraftment of the donor human cells was. In addition to lacking T, B, and NK cells, BNX mice were deficient for lymphokine activated killer cells (LAK), which allowed not only reconstitution of the myeloid lineage, but better maintenance of the donor human HSCs (Andriole et al., 1985; Kamel-Reid & Dick, 1988). Crossing of the SCID strain with non-obese diabetic (NOD) mice produced an enhanced immune-deficient strain, which had a marked reduction of NK cells compared SCID mice that allowed the long-term engraftment of primitive human cells that became known as SCID-repopulating cells (SRCs) (André Larochelle et al., 1996; Shultz et al., 1995). To further improve the efficiency

of human cell engraftment, the NOD/SCID was reengineered by deleting the IL2 receptor gamma chain gene, which resulted into the NSG mouse strain which combined the properties of NOD/SCID and complete lack of NK cells (Kollet et al., 2000; McKenzie et al., 2004, 2006). Further enhancements were to follow on immune-deficient mouse strains, but the NSG became the standard model for xenograft studies.

1.2.2 The Human HSC

By the end of the 80s, several antibodies raised against the immature leukaemia cell line, KG1, had been reported, and shown to positively select for human bone marrow cells with potent CFC (Andrews et al., 1986; Civin et al., 1984; Tindle et al., 1985). Once it was established that all these antibodies recognized CD34 and its expression by colony-forming cells was not only conserved between mice and humans, but CD34 was also similarly glycosylated and phosphorylated in both species, the study of HSCs was rapidly accelerated (Brown et al., 1991; Krause et al., 1996). Results from studies in humans, baboons, rhesus monkeys and mice were consistently demonstrating the potency of the self-renewal of $CD34^+$ bone marrow-derived cells (Berenson et al., 1988, 1991; Krause et al., 1994). These findings were soon followed up by studies showing CD90 expression purified human foetal bone marrow for colony-forming cells as well as SRCs (Baum et al., 1992a). Following up on another marker first described in mouse, $CD34^+/Kit^{High}$ cells were shown to possess colony-forming capacity, whereas HSCs capable of long-term self-renewal were reported to be $CD34^+/Kit^{Low}$ (Gunji et al., 1993; Okada et al., 1992). Sca-1, a member of the Ly6 family of proteins, is not conserved in humans, but a related protein, CD59, was shown to be expressed on virtually all Lin-/CD34+/CD90+ cells, which are enriched for HSCs (Hill et al., 1996). Another early marker for mouse foetal HSCs, CD93 (clone AA4), was also reported to be expressed by human repopulating cells (Anjos-Afonso et al., 2013; Danet et al., 2002; McKearn et al., 1985). CD90, Kit, and CD93 were all HSC markers that had been first described in mouse, but

other highly purifying murine HSC markers proved to be less useful in bringing human HSCs to single cell purity. In mouse, single cell injections of LSK/CD48$^-$/CD150$^+$ resulted in long-term multilineage reconstitution, but in humans repopulating cells were found both in SLAM and non-SLAM populations (Andre Larochelle et al., 2011; Sintes et al., 2008; Zhou et al., 2015). In fact, it appeared that human and murine HSCs were significantly phenotypically different. While mouse HSCs were reported to express CD38, the human HSC was widely accepted to be negative for this marker (Higuchi et al., 2003; Randall et al., 1996).

After CD90 was confirmed to be expressed not only by foetal bone marrow-derived, but also umbilical cord blood (UCB) CD34$^+$ cells with colony-forming capacity , no significant strides were made in identifying other positive selection markers for some years (Baum et al., 1992a; Craig et al., 1993). In a continuation of some early studies showing that CD45^{Low} cells possess potent colony-forming capacity, studies on expression of CD45 isoforms demonstrated that the CD45RA$^-$/CD34$^+$ fraction was enriched for multipotent progenitors that acquired CD45RA expression over a period of time in culture with cytokines (Beverley et al., 1980; Lansdorp et al., 1990; Mayani et al., 1993). Next, it was asserted that the population enriched with CFCs and SCID-repopulating cells was also CD38$^-$ (Bhatia et al., 1998; Hao et al., 1995).

Injection of Lin$^-$/CD34$^+$/CD38$^-$ cells transduced with unique retroviral tags in NOD/SCID mice showed that a number of clones detected up to 6 weeks post-injection dispersed to be replaced by distinct long-term clones at week 12 (Guenechea et al., 2001; Mazurier et al., 2004). Not only that, but in a similar set of experiments, the authors could detect clones in secondary mouse recipients that were undetectable in the primary recipient (McKenzie et al., 2006). It appeared that while the *in vivo* repopulating capacity of short-term HSCs could be captured, highly quiescent HSCs could be missed in primary recipients.

Using the improved NSG mouse model, the CD34$^+$/CD38$^-$/CD45RA$^-$/CD90$^+$ population was shown to have *in vivo* multilineage engraftment that persisted into secondary mouse recipients (Ravindra Majeti et al., 2007). Furthermore, it appeared that the CD34$^+$/CD38$^-$/CD45RA$^-$/CD90$^-$ population also contained cells with robust repopulating capacity and multilineage differentiation, although the generated engraftment was lower than their CD90$^+$ counterpart (Ravindra Majeti et al., 2007).

Another noteworthy observation, however, was that few $CD34^+/CD38^-/CD45RA^-/CD90^-$ could generate engraftment in secondary mouse recipients, revealing that the population may still contain some long-term HSCs that could not be delineated by CD90 expression alone (Ravindra Majeti et al., 2007).

This disparity was not addressed until a few years later, when scientists investigated whether integrins, key mediators in interactions with the niche, could be used to further purify human HSCs (Benveniste et al., 2010; Notta et al., 2016). The $CD34^+/CD38^-/CD45RA^-/CD90^{+/-}$ populations were further demarcated with the addition of CD49f as a marker. The four subsets were shown to possess differential capacities to engraft into NSG mice (Notta et al., 2011). The $CD90^+/CD49f^+$, $CD90^-/CD49f^+$, $CD90^+/CD49f^-$, $CD90^-/CD49f^-$ populations were estimated to contain, 1 in 10.5, 1 in 22.1, 1 in 111.3, and 1 in 735.2 repopulating cells, respectively (Notta et al., 2011). Notably, the engraftment in mice injected with $CD90^-/CD49f^-$ cells peaked between 2 and 4 weeks and then steadily declined, but still produced multiple lineages, indicating the population mostly contains multipotent progenitors (Notta et al., 2011). CD49f brought forward the purification strategy of human *bona fide* HSCs one step further, with about 10% of $CD90^+/CD49f^+$ cells estimated to have repopulating capacity. Nevertheless, a number of long-term repopulating cells could be detected within the $CD90^-/CD49f^+$ fraction, indicating that CD49f and CD90 could not be used to exhaustively delineate human HSCs.

The advancement of mass cytometry and single-cell RNAseq technologies was quickly followed by an unprecedented number of reports summarising vast cell surface marker and transcriptomic data. While much of this new area of research focused on getting snapshots of the entire haematopoietic system, those keen on finally capturing the human HSC and MPP compartments were zooming in on the $CD34^+/CD38^-$. ECMN (Endomucin) was used to divide the $CD34^+/CD38^-/CD45RA^-/CD90^+$ into two, with the $ECMN^+$ fraction successfully repopulating conditioned mouse recipients, while as many as 50, 000 $EMCN^-$ cells failed to generate human engraftment 20 weeks post-injection (Reckzeh et al., 2018). Similarly, single cell gene expression analysis of $CD34^+$ bone marrow-derived cells showed that CD164 expression also demarcates a population with robust *in vivo* multilineage engraftment potential (Pellin et al., 2019). In single cell colony assays, $CD49f^+/CD34^{Low}/CLEC9A^+$ cells appeared to be more quiescent than $CD49f^+/CD34^{High}/CLEC9A^-$ cells (Belluschi et al., 2018). The former also generated long-term multilineage engraftment, while

the latter could only maintain myeloid and lymphoid engraftment (Belluschi et al., 2018). These findings were key to show that even within what was considered a highly enriched HSC compartment, commitment to different lineages is already occurring.

In human, the characterisation of the stages of differentiation began initially with a lymphoid progenitor that could give rise to B, T, natural killer (NK), and dendritic cells (DCs) (Galy et al., 1995; Hokland et al., 1987). Defined by the $CD34^+/CD38^+/CD45RA^-/CD10^+$ phenotype, this population was later shown to have active IgH rearrangements and expression of genes which aligned with B cell commitment (Davi et al., 1997; M. Rossi et al., 2003). In cord blood a lymphoid-primed population was shown to be $CD34^+/CD38^-/CD7^+$ with capacity to generate DCs, NK, and B cells (Hao et al., 2001). In addition, while the $CD10^+$ lymphoid-primed progenitors were also shown to have myelo-erythroid potential, in cord blood $CD34^+/CD38^-/CD45RA^-/CD10^+/CD7^+$ lacked colony-forming capacity, which indicated that human haematopoiesis may follow the model of the lineage restriction hierarchy in mouse (Galy et al., 1995; Hao et al., 2001). In pediatric bone marrow, CD24 was used to positively mark a B cell restricted population, while $Lin^-/CD34^+/CD10^+/CD24^-$ cells possessed a broader lymphoid differentiation spectrum (Six et al., 2007). While all these studies indisputably verified the existence of a multi-lymphoid progenitor (MLP) associated with CD10 expression some of these reports utilised markers such as CD38 and CD45RA, while others didn't. These variable gating strategies posed a challenge in unifying the findings into a universal phenotype for the human CLP. In a study attempting to reconcile all these results, the phenotype of the MLP was defined as $CD34^+/CD38^-/CD45RA^+/CD10^+/CD7^{+/-}$ and shown to be able to reconstitute NSG mice with lymphoid, but also myeloid cells indicating the obligatory lymphoid-restricted progenitor observed in mouse may not exist in human haematopoiesis (Doulatov et al., 2010). With their unique ability to generate not only DCs, but also macrophages and monocytes, cells previously ascribed to granulocyte macrophage progenitor (GMP) population, MLPs were presenting an ambiguity in the pyramid of haematopoiesis. Important functions for DCs include antigen presentation and activating T cells, and one mechanism to coordinate their balanced generation with lymphoid cells could be to restrict them to the same multipotent progenitor. Monocytes and macrophages, on the other hand, are often the first responders in the innate immune response, and in order to meet

urgent physiological demands could have evolved mechanisms to emerge from multiple points in the haematopoietic hierarchy. Interestingly, the MLP populations appears to be highly enriched in pre-natal and early post-natal haematopoiesis and quite rare in adults (Doulatov et al., 2010; M. Rossi et al., 2003).

The first report on the human CMP came just a few years after the mouse. Within the $CD34^{+}/CD38^{+}$ compartment, the CMPs, GMPs, and MEPs were shown to be $CD45RA^{-}/CD123^{Low}$, $CD45RA^{+}/CD123^{Low}$, and $CD45RA^{-}/CD123^{-}$, respectively (Manz et al., 2002). The CMPs and GMPs were able to generate myeloid colonies, while the MEP population was restricted to Mk and erythroid colonies, suggesting the isolation strategy was robustly picking apart the lineage-committed progenitors (Manz et al., 2002). In a similar attempt to delineate these three progenitor populations Flt3 was shown to be a reliable substitute for CD123 (Doulatov et al., 2010).

These population-based studies outlined the paradigm for the step-wise maturation of HSCs into oligopotent progenitors in tandem with the bifurcation of the myeloid and lymphoid lineages. However, questions were still being asked about when are Mks and erythroid progenitors restricted and if HSCs are primed for specific fates early on haematopoiesis. The answers could only be found by assessing the differentiation and self-renewal potential at the level of individual cells, which required technically-challenging and significant scaling-up of the number of assays performed at a time. Nonetheless, these experiments were soon being performed in large numbers by groups studying both murine and human haematopoiesis, and finally shedding new light on processes of lineage commitment.

Much of the human research had exclusively focused on the CD34+/CD38- population as the HSC compartment, while in mouse it had become established years before that the long-term HSC is CD34- (Krause et al., 1996; Novelli et al., 1998; Osawa et al., 1996; Sato et al., 1999). The view of the human HSC being $CD34^{+}$ was becoming deeply-rooted, and attempts to reconcile this interspecies difference involved attempts to show that the mouse and human CD34 genes were differentially regulated (Okuno et al., 2002).

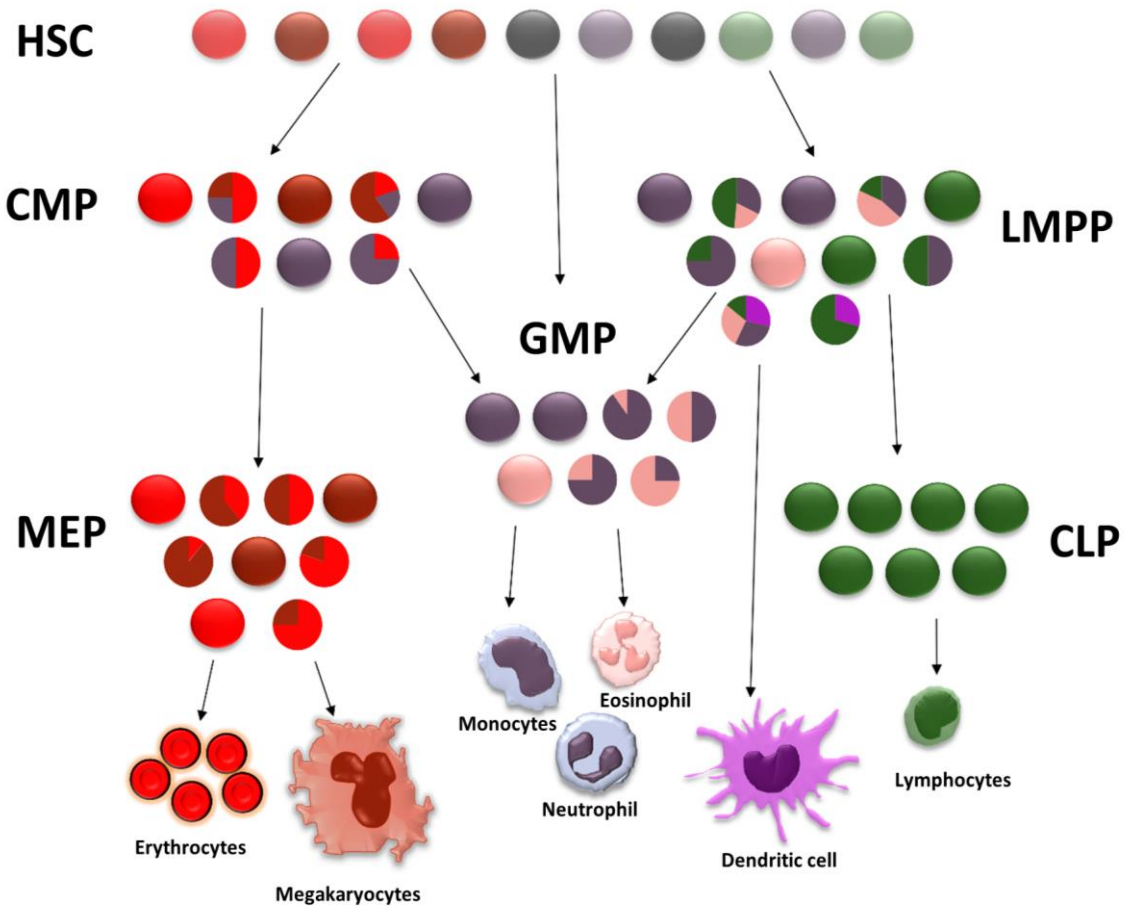

Figure 1.1. Revised haematopoietic stem cell hierarchy

Schematic of revised haematopoietic stem cell hierarchy where faded colours represent priming events (HSC level) and different colours in downstream progenitor populations represent potential cell fate outcomes (red – erythroid; rust – megakaryocytes; purple and pink – myeloid; fuchsia – dendritic cells; green - lymphocytes)

1.2.3 The CD34^{-} haematopoietic stem cell

After Berenson and Krause showed that baboons, humans and mice could be successfully transplanted with CD34^{+} bone marrow-derived cells, CD34 became a universal hallmark for HSCs (Berenson et al., 1988, 1991; Krause et al., 1994). It wasn't long, however, that this notion was put into question, when multiple groups began reporting observations in mouse that both the LSK and Thy^{Low}/Lin^{-}/Sca-1^{+} populations contain CD34^{-} cells with *in vivo* repopulating

capacity (Morel et al., 1998; Osawa et al., 1996). Furthermore, the $CD34^-$LSK population consistently repopulated secondary mouse recipients, while the $CD34^+$LSK cells were observed to have much lower and less consistent secondary engraftment, implying the $CD34^-$ possessed superior long-term self-renewal capacity (Tajima et al., 2000).

In contrast, clinical studies in humans using CD34-selection before transplants were asserting the phenotype of the HSC as $CD34^+$, despite independent reports suggesting the $CD34^-$ compartment should not be ignored in the search for the human HSC (Bensinger et al., 1996; Hassan et al., 1996; Link et al., 1996). After Goodell and colleagues demonstrated that murine HSCs are highly enriched in a "side population" (SP) characterised by its efficiency to efflux Hoechst dye, they observed a similar SP^+ HSC in humans and rhesus monkeys (Goodell et al., 1996, 1997b). A large proportion of the SP^+ cells were $CD34^-$, and rhesus $CD34^-$ SP^+ cells expanded in long-term culture with stromal support and were reported to acquire CD34 expression over time (Goodell et al., 1997b). Similarly, umbilical cord blood sorted Lin^-CD34^- cells were shown to have SCID repopulating capacity, while bone marrow-derived $CD34^-$ cells generated multilineage engraftment in foetal sheep (Bhatia et al., 1999; Zanjani et al., 1998). These early experiments were already alluding to the challenges the $CD34^-$ population would pose. The rhesus $CD34^-SP^+$ cells could not generate colonies without expansion on stromal cells first, suggesting the colony-forming cells within the population were refractory to the signalling molecules in standard CFC assays which successfully push $CD34^+/CD38^-$ cells to generate colonies (Goodell et al., 1997b). Hundreds of thousands $Lin^-CD34^-CD38^-$ cells had to be injected in NOD/SCID mice to achieve human reconstitution, implying that to study $CD34^-$ SCID repopulating cells at population level, purity had to be improved (Bhatia et al., 1997; Goodell et al., 1997b). Drawing on findings in mice, CD93 was later identified as a positive selection marker for $CD34^-$ repopulating cells derived from both from umbilical cord blood and adult human bone marrow (Danet et al., 2002; Petrenko et al., 1999). The repopulating cell frequency in the $Lin^-/CD34^-/CD38^-/CD93^+$ (-/-/+) was estimated at 1/7500 cells in the enhanced NOD/SCID mouse (Danet et al., 2002). In a different study, the -/-/+ population was shown to reliably

repopulate NOD/SCID mice, but with a 6-week delay compared to $CD34^+/CD38^-$ cells, potentially explaining why the population has been largely overlooked as a source of HSCs, and also alludes to its quiescent nature (Anjos-Afonso et al., 2013). Furthermore, the -/-/+ population was able to give rise to functional repopulating $CD34^+$ cells, while the $CD34^+/CD38^-$ population could not give rise to $CD34^-$ cells that could repopulate secondary mouse recipients (Anjos-Afonso et al., 2013). All this evidence generated over almost three decades has asserted the importance of the $CD34^-$ HSCs in mouse, human and other species. However, CD34 remains the *bona fide* human HSC marker with very little new data coming in for the elusive $CD34^-$ HSC.

Despite being widely used for purification and clinical assessment of haematopoietic stem cell transplants (HSCT), the significance and function of CD34, a cell surface glycoprotein, still isn't clearly understood. The diminished capacity of murine and human $CD34^-$ HSCs to expand on stroma and form colonies *in vitro* has been reported by multiple groups. Together with their delayed engraftment in NOD/SCID mice, this gives a strong indication that while the $CD34^+$ cells contain a mix of short-term and long-term repopulating cells, the repopulating $CD34^-$ SRC fraction largely consists of long-term quiescent HSCs (Anjos-Afonso et al., 2013; Danet et al., 2002). How much does the diminished CD34 expression play a role in maintaining the population at a latent state is an open question. Using the BNX (beige/nude/X-linked immunodeficient) mouse strain, Dao et al. reported that the injection of $CD34^+/CD38^-$ cells results in long-term multilineage engraftment, but interestingly, in the bone marrow of these mice CD34 was downregulated on the cell surface and also at the transcript level (Dao et al., 2003). When BNX-derived human $CD34^+/CD38^-$ xenografts were transplanted in a secondary NOD/SCID mouse recipient they successfully generated secondary multilineage engraftment with detectable CD34 expression. The authors concluded that CD34 expression might be reversible. Along with the observations that the xenografts generated from BNX mice performed better in LTC-IC and colony assays than the xenografts from NOD/SCID mice, this is indicative that CD34 expression may be an "activity switch" for HSCs.

Very little had been done to assess the molecular mechanisms of the $CD34^-$ HSCs until it was reported that the -/-/+ population expresses high levels of cell cycle regulators such as CDKN2C, CDKN1B, and CDKN1A that maintain the population in G0 state (Anjos-Afonso et al., 2013). In addition, following culture with TGFB1, these cell cycle regulators were upregulated in -/-/+ cells, indicating that the quiescent state of the cells may be at least in part maintained by TGFβ signalling (Anjos-Afonso et al., 2013). Interestingly, the -/-/+ population is refractory to Wnt pathway activation as shown by addition of Wnt3A and Wnt10b with little or no response on the population (Anjos-Afonso et al., 2013). These findings finally succeeded at shedding some light on the molecular mechanisms maintaining the $CD34^-$ population, but did not definitively establish how it is maintained in steady-state haematopoiesis as the frequency of SRC is only 1 in 6000 cells in the -/-/+ fraction. In order to zoom in on the actual $CD34^-$ HSC close to single-cell purity was required, which scientists looking for the human HSC within the $CD34^+$ compartment were far more successful at achieving.

The inability to isolate a pure human HSC resulted in the human haematopoietic hierarchy being modelled after findings in mouse. But with the advancement of sequencing technology and methods in large-scale gene expression profiling, the understanding of the human haematopoietic hierarchy began to change rapidly.

1.2.4 Cell fate decision – single cells

The standard model of haematopoiesis pertains a hierarchy maintained by self-renewing multipotent HSCs, which pass through stages of differentiation leading to oligopotent and eventually, unipotent progenitors. All conclusions were drawn from population-based studies under the assumption that marker-pure subsets would reflect functionally homogenous populations. The improvement of purification strategies and the development of high-throughput single cell-based assays allowed to revisit the hierarchical structure of haematopoiesis.

In experiments relying on engineering mouse strains with inducible transposon tagging, the lineage output of single HSCs and progenitors could be

tracked in unperturbed steady-state haematopoiesis (Sun et al., 2014). Analysis of the transposon tags of differentiated cell, stem cells and progenitors revealed that LT-HSCs had very limited contribution to the lineage output, compared to MPPs (Sun et al., 2014). Furthermore, only a small number of myeloid and lymphoid tags were shared, indicating an early fate specification at the level of ST-HSCs and MPPs at the time of the induction of the tags (Sun et al., 2014). While contribution of LT-HSCs cannot be ruled out by this study, it has implications that have been previously suggested by the poor reconstitution capacity previously demonstrated in transplantation studies using quiescent HSCs (Anjos-Afonso et al., 2013; Bhatia et al., 1997; Osawa et al., 1996). Using different approaches to trace clonal dynamics, it was actually reported that LT-HSCs do contribute eventually to the different lineages over periods longer than an year, but confirmed that within shorter windows of sampling it is ST-HSCs and MPPs that make major contributions to steady state haematopoiesis (Busch et al., 2015; Rodriguez-Fraticelli et al., 2018). Another observation that holds implications for human xenograft studies and HSCTs is that transplanted HSCs were heterogeneously reconstituting donors, with some replicates resulting in disproportionately high contribution to mature haematopoietic cells and others failing to contribute at all (Busch et al., 2015). This hints at a significant pitfall of transplantation experiments, where engraftment kinetics and lineage reconstitution of few HSCs may not necessarily represent the steady-state haematopoiesis. This is further corroborated by ST-HSCs being shown to exhaust fairly quickly when transplanted, while in unperturbed conditions appear to be steadily maintained (Busch et al., 2015).

In humans, single cell assays began also boosting our understanding of the haematopoietic hierarchy. With improved assays that could reliably readout Mk differentiation potential *in vitro*, one of the first studies investigating clonal output of single cells revealed that the highest number of multipotent cells resides within the HSC and MPP compartment, with CMPs mainly consisting of a mix of unipotent and with the majority of multipotent clones restricted only to two lineages (i.e. Mk-myeloid or erythroid-myeloid) (Notta et al., 2016). Cells with robust Mk-erythroid potential could be isolated by sorting for CMPs, but also in MPPs and MEPs, with the unifying phenotypic feature being a combination of negative CD110 (TPO receptor) and positive CD71 expression (Notta et al., 2016). This was in line with previous theories that proposed Mks and red blood cells could emerge from several branching points

of the haematopoietic hierarchy to satisfy to high demand for these cells in a healthy individual. In another study, focusing more closely on the MEP compartment, single cell functional assays revealed through enhanced cell surface marker isolation that significant heterogeneity existed within the compartment (Psaila et al., 2016). Mk-biased MEPs were shown to be largely $CD71^{+}/CD41^{+}$, Erythroid-MEPs – $CD71^{+}/CD41^{-}$, and the bipotent MEPs were negative for both markers but positive for CD44 (Psaila et al., 2016). This was confirmed independently by study looking into the differentiation patterns of single $CD41^{+}$ CMPs, which gave predominantly rise to Mk cells (Miyawaki et al., 2017). Interestingly, these versatile "pre-MEPs" also generated granulocyte/macrophage (GM) colonies and immature colonies of mixed lineages (GEMM), suggesting they possessed residual myeloid potential despite being largely MkE-restricted (Psaila et al., 2016).

Another tracing of the progenies from single sorted HSCs merged clone size and clonal lineage output data into multi-dimensional analysis suggesting cell fate may not only be determined at the stem cell stage through priming, but was also an inheritable feature observed in daughter and granddaughter cells (Lee et al., 2017). Daughter cells that were deviating from their ancestral lineage bias frequently produced myeloid cells, which may be explained by observations that a myeloid bias emerging with advancing age of an organism (Pang et al., 2011; D. J. Rossi et al., 2005; Sudo et al., 2000). The authors also showed that increasing number of divisions reduce the clonal potency of HSCs, but this effect was weaker in *in vivo* experiments compared to *in vitro* cultures (Lee et al., 2017). Most importantly, the *in vitro* experiments entirely excluded readouts of Mk and erythroid output, without which any conclusions about lineage bias and priming are incomplete.

In attempts to define the bifurcation of the myeloid and lymphoid lineages, single-cell lineage output assays have indicated compartmentalized cell fate commitment is not as well restricted to specific populations as previously thought. In single cell differentiation assays, while MLPs are exclusively lymphoid primed, LMPPs and GMPs have been reported to generate both lymphoid and myeloid cells, with small number of bipotent progenitors contributing to both lineages simultaneously (Karamitros et al., 2018). Refinement of sorting strategy resulted in improved myeloid or lymphoid restricted lineage output, it still failed to completely eliminate bipotent and multipotent single cells, which may be due to limitations in

FACS technology and the challenge of sorting cells to the level of functional homogeneity solely based on cell surface markers (Karamitros et al., 2018).

While there were strong indications that cell fate priming may be occurring at the stem cell level in humans, little was done to investigate single cell lineage outputs within the purest reported HSCs, the $CD49f^{+}/CD90^{+}$ population with the majority of reports focusing further down the hierarchy or by failing to incorporate CD49f in the HSC sorting strategy. Belluschi and colleagues finally assayed 5000 single $CD49f^{+}$ HSCs and correlated the clonal lineage outputs with the expression of a panel of cell surface markers, finding that Clec9a and CD34 expression drove most of the variance in multilineage colony output (Belluschi et al., 2018). Consequently, the $Clec9a^{+}/CD34^{Low}$ subset was shown to be primed for cells with myeloid and erythroid potential, while the $Clec9a^{-}/CD34^{High}$ subset for myeloid and lymphoid (Belluschi et al., 2018). In addition, the $Clec9a^{+}/CD34^{Low}$ subset was capable of producing $Clec9a^{-}/CD34^{High}$ in culture, but not the other way around, implying that the former is hierarchically superior to the latter *in vitro* (Belluschi et al., 2018).

While all these single cells studies of human haematopoietic output have indicated important features of the haematopoietic hierarchy, they have been conducted either *in vitro* or in transplantation studies in immunodeficient mouse recipients, which are limited in their capacity to mimic normal human physiology and steady state haematopoiesis. But advancements in next generation sequencing technology unveiled age-related patterns of certain mutation signatures and how they can be exploited to track clonal dynamics (Alexandrov et al., 2015; Jaiswal et al., 2014). In a study conducted on samples from single individual, somatic mutation clones revealed that HSCs defined as $CD34^{+}/CD38^{-}/CD45RA^{-}/CD90^{+}$ contribute continuously to mature lineages, which contradicted similar studies in mouse, where LT-HSCs are largely quiescent and SH-HSCs largely maintaining the system (Busch et al., 2015; Lee-Six et al., 2018; Sun et al., 2014). These results were in agreement with a long-term study on the haematopoietic reconstitution of patients that received a transplant of engineered $CD34^{+}$ cells, which showed that the transplant was contributing steadily to mature cells over a four-year period (Biasco et al., 2016). While the latter of these studies is tracking reconstitution by transplanted cells, they allude to a fundamental interspecies difference in that human HSCs appear to act more similarly to murine SH-HSCs, contributing steadily to the haematopoietic system in healthy individual, and mice retaining an inactive LT-HSC population.

While this may be an artefact of evolutionary adaption, another overlooked explanation is that human clonal dynamics studies have overlooked the $CD34^-$ HSCs, which may possibly be the counterpart of the murine LT-HSC and if incorporated reveal that the haematopoietic clonal dynamics between the two species is more similar than these incomplete studies suggest.

Attempts to reconcile lineage potential with specific gene expression were beginning to be made as qPCR and microarray technologies were developed. But with the rapid advances in bulk and RNA-sequencing technologies, quickly amassing data was beginning to shed light on the limitations of population-based approaches to studying haematopoiesis and, it was becoming evident that haematopoietic cell differentiation could hardly be characterized as a clear-cut step-wise process.

1.2.5 The molecular landscape of human haematopoiesis

Population heterogeneity was being appreciated early on in the days of haematopoietic research and low-throughput methods to assay the gene expression profiles of single cells were undertaken, but only could yield information for few genes and cells at a time (Adolfsson et al., 2005). It was, however, clear even on population level that different lineages exhibited distinct gene expression patterns. With the development of microfluidics systems and chip arrays, small panels of genes were being investigated in single sorted cells from different haematopoietic compartments, revealing how known myeloid regulator, SPI1, for example, had higher transcripts in CMPs with Flt3 expression than those negative for Flt3 (Warren et al., 2006). With improving cell capture rates and growing gene panels, previously established expression patterns, such as GATA1 (megakaryocyte and erythroid cells), GATA2 (HSCs), GFI1 (CMPs, LMPPs, CLPs), could be used to validate single cell transcriptomic data (Hamlett et al., 2008; Huang et al., 2009; Mouthon, 1993; Tipping et al., 2009). Previous theories that the megakaryocyte and erythroid lineages were more comparable to HSCs were being confirmed, while also presenting the opportunity to map gene regulatory networks by identifying correlated expression (Moignard et al., 2013). By sorting individual cells by cell-surface markers and assigning each an index, individual cell transcriptomes can be mapped to a particular phenotype and allow the linking of gene expression to a particular population. In such

studies, genes with different expression patterns can give clues about their roles in haematopoiesis: RUNX1 and FLI1 were similarly expressed by different populations, while others had more variable expressions (Moignard et al., 2013). This indicates that some genes may be essential for normal functioning of a large number of haematopoietic cells, while others are probably required for dynamic transitions from quiescence to proliferation or commitment from a multipotent to a unipotent progenitor. With increasing number of transcripts that could be sequenced, it became evident that through hierarchical clustering algorithms, gene expression profiles were sufficient to delineate HSCs and the different progenitor population and reveal correlations between gene expression modules and particular cellular phenotypes (Guo et al., 2013). Methodologies coupled with single-cell functional assays to validate bioinformatics findings were also proving useful in generating lists of "signature" genes (Wilson et al., 2015). With the growing number of cells and genes that could be sequenced, an argument was being made that the transcriptomes of thousands of single cells could give a snapshot of the entire haematopoietic system and resolve the hierarchy in full. But a real challenge existed in drawing meaningful conclusions from data representing more than two dimensions. For that many sought answers by leaning onto dimensionality reduction algorithms that work by reducing high-dimensional data to low-dimensional representation where individual cells are represented by map points. Principal component analysis (PCA), stochastic neighbour embedding (t-SNE), and uniform manifold approximation and projection (UMAP) became some of the most commonly used dimensionality reduction algorithms, the role of which is to ensure maximal preservation of the high-dimensional data structure to the low-dimension representation (Hotelling, 1933; McInnes et al., 2018; Van Der Maaten & Hinton, 2008). While PCA and t-SNE could preserve local data structures, they didn't necessarily preserve the global data structure, which meant that closely related cells within a "neighbourhood" would be plotted together, but the degree of similarity between cells in different parts of the plot could not be inferred (Haghverdi et al., 2015; Hotelling, 1933; Van Der Maaten & Hinton, 2008). UMAP was developed precisely to address this and can be used to interpret the global relationship of mapped cells (McInnes et al., 2018). To address more complex questions such as building maps of differentiation kinetics, new computational analyses like pseudotemporal ranking were developed that could represent the cell differentiation continuum and infer the timing of gene expression

(Trapnell et al., 2014). To precisely identify intermediate populations during differentiation, iterative clustering coupled with known gene expression data was also reported as a powerful predictive tool (Grün et al., 2016; Olsson et al., 2016).

As technologies became more widely available, an onslaught of data on single-cell haematopoietic cells was published. Much of it was generated from mouse and used to validate the algorithms and computational analysis methods discussed above. These studies were making fascinating suggestions about haematopoiesis. For one, some data was supporting results from single-cell clonal assays indicating early priming within the HSC compartment (Grün et al., 2015). While differentiated populations would cluster tightly together, higher entropy appeared to exist among HSCs, which were also expressing transcription factor modules associated with specific lineages (Grün et al., 2016). Unfortunately, these methods used in mouse have not been applied to a human dataset, but a study solely based on index-sorted haematopoietic cells derived from an adult human bone marrow allude to a continuous process of differentiation from a stem cell pool to unipotent progenitors with progenitors such as MLPs and MPPs representing transitory populations that cluster within the HSC population (Velten et al., 2017). Index-sorting also allowed coupling cell surface marker expression data with clusters based on transcriptomic data, and showed that the majority of $CD34^{+}/CD38^{-}$ cells clustered together, while $CD34^{+}/CD38^{+}$ cells were clustering into distinct populations (Velten et al., 2017). This indicated that gene expression variability was insufficient to separate the phenotypic HSC compartment, while the population known for containing oligopotent progenitors resulted in a map comprised of disconnected cell types (Velten et al., 2017). What is more, the predefined CMP population that had been previously shown to occupy a major branching point within the haematopoietic hierarchy was turning out to be composed of largely heterogeneous cells with unipotent potential (Notta et al., 2016; Velten et al., 2017). With these human studies confirming what had been previously reported in mouse - transcription factors associated with distinct lineages were expressed ubiquitously at low levels by HSCs - the regulation of lineage commitment became again a hot topic of investigation (Notta et al., 2016; Sanjuan-Pla et al., 2013; Velten et al., 2017). While this new data can be reconciled with stochastic processes that determine cell fate, this view does not comply with the fact that cell surface markers such as CD34 and CD38 appear deterministic in delineating different levels of commitment within the haematopoietic hierarchy. Another challenge for the study

of lineage priming within HSCs is with the limitation of scRNA-seq, itself. While the technology has rapidly improved, low expressed genes are unlikely to be picked up. This is a major problem for transcription factors which are known to be expressed at low levels. This also means that closely related cell types may not be distinguishable at the level of current technology and analysis methods. Another major challenge with these recent human studies is that none of them look within the $CD34^-$ compartment, therefore omitting an HSC population from their analysis. Previously shown to be highly latent *in vitro* and *in vivo,* this population may be the equivalent of the murine LT-HSC, which has been shown to have low contribution to steady-state haematopoiesis. In this scenario the human $CD34^+/CD38^-$ HSC may represent the murine ST-HSC, and consequently make the human haematopoiesis model match the murine. A single attempt to conduct scRNAseq on $CD34^-$ HSCs only yielded data for about 20 cells, which was insufficient to properly address the relationship of these cells to the $CD34^+$ HSCs (Sumide et al., 2018). A major caveat is that $CD34^-$ HSCs are quite rare and would be difficult to capture in sufficient numbers from whole bone marrow samples, which so far are reported to yield datasets consisting over several thousand haematopoietic cells. Pooling of samples or use of umbilical cord blood, where this population is more abundant may be necessary to allow proper single cell analysis (Bhatia et al., 1998; Danet et al., 2002).

While it is still difficult to determine whether the haematopoietic hierarchy consists of a continuum of gradually maturing cells or select compartments, it is clear that adult bone marrow is far more enriched for unipotent progenitors rather than progenitors that can yield mixed-lineages (Notta et al., 2016). But such oligopotent and multipotent progenitors have been reported to be enriched in human with both functional and single cell gene expression assays (Notta et al., 2016). However, they exist at earlier developmental stages in the foetal liver (FL), foetal bone marrow (FBM) and the umbilical cord blood.

1.3 Haematopoiesis during ontogeny

It was initially in fertilized chick eggs that the sources of haematopoietic cells were first identified (P. Murray, 1932). In mouse, primitive haematopoietic cells were shown to exist in the yolk sack, but explant cultures derived from murine embryos

were highlighting the existence of a functionally distinct haematopoietic population that would later become known as definitive (K. Takahashi et al., 1989; P. Wong et al., 1986). The intraembryonic source of definitive haematopoiesis was confirmed with grafts of 9-day splanchnopleura that repopulated SCID mice and gave rise to multilineage haematopoietic colonies (Godin et al., 1993, 1995). Finally, the aorta-gonad-mesonephrous region of the dorsal aorta was shown to be the source of cells with potent CFU-S activity (Medvinsky et al., 1993). Studies in human embryos followed, reporting that $CD34^+$ cells isolated from the ventral region of the aorta could produce functional haematopoietic cells *in vitro* (Tavian et al., 1996; Tavian, Hallais, & Peault, 1999). What is more interesting is that these haematopoietic cells were being isolated from embryos prior to the formation of the liver that had previously been shown to be the site of early embryonic haematopoiesis (Tavian, Hallais, & Peault, 1999). Limited not solely to the aorta, but the vitelline artery and the aorta-gonad-mesoneprhous (AGM), it was becoming accepted that definitive HSCs capable of generating the myelo-lymphoid lineages do not emerge from extra-embryonic sites like the yolk sac, but the newly formed endothelium (de Bruijn et al., 2000; Ivanovs et al., 2011; Medvinsky & Dzierzak, 1996; Tavian et al., 2001).

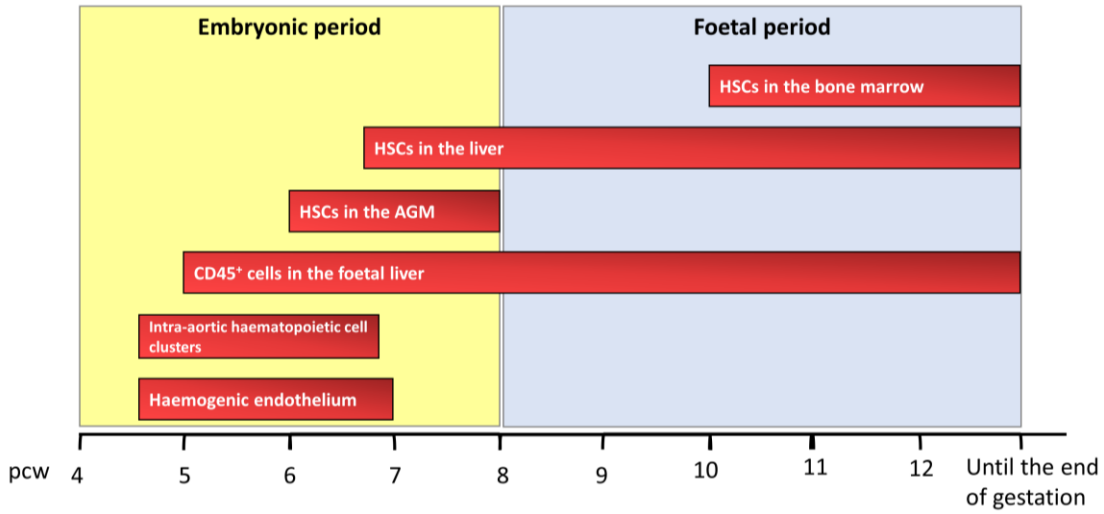

Figure 1.2. Development of human haematopoiesis

Chronology of human haematopoietic development during ontogeny. Embryonic period ends at 8 pcw (post-conception weeks) and is usually annotated in Carnegie stages (CS). These have been converted to pcw for simplicity here. Red rectangles depict processes in the development of definitive haematopoiesis. Beginning and ends of rectangles correspond to start and extinction of a process.

1.3.1 Characteristics of foetal HSCs

The AGM HSCs in both mouse and human express genes well-associated with adult HSCs such as RUNX1, TAL-1, GATA-2, c-myb and LMO2 prior to colonizing the FL (Bertrand et al., 2004; Labastie et al., 1998; Tavian et al., 1996). While it appears, the first definitive cells acquire CD34 expression as they emerge from a transient population of mesoderm cells that are $CD34^-$/VEGFR-2^+, these $CD34^-$ precursors have not been properly assayed for haematopoietic potential (Cortés et al., 1999). It is, however, likely that in order for the first definitive HSCs to emerge at around 4 post-conception weeks (pcw), they first need to acquire CD34 expression before entering circulation and homing to the foetal liver (FL). As definitive murine HSCs emerge, they exhibit dynamic expression of cell surface markers. One of the first to be expressed by nascent murine HSCs emerging from the AGM is CD41, which is gradually downregulated as the HSCs mature (Bertrand et al., 2004; Corbel & Salaün, 2002). In both murine and human early HSCs CD43 and CD45 expression can be detected by HSCs later, but cannot be isolated from the foetal liver prior to colonization of cells generated from the AGM (North et al., 2002; Rybtsov et al., 2014; Vodyanik et al., 2006).

Murine AGM HSCs capable of *in vitro* expansion have also been shown to be positive for other well-characterised markers like C-kit and CD93, but also endothelial marker CD31 (Bertrand et al., 2004; Tavian et al., 1996). Both human and murine AGM-HSCs gradually acquire expression for the pan-haematopoietic marker CD45, and primitive human $CD41^+$/$CD31^+$/$CD93^+$/Kit^+/$CD45^-$ AGM-derived cells have been reported to repopulated mouse recipients (Ivanovs, Rybtsov, Anderson, Turner, et al., 2014). Post the onset of circulation, HSCs begin colonising the FL, which is the primary site of early embryonic haematopoiesis in both mouse and human before the development of the bone marrow niches. One key HSC marker, CD90, was first identified in the human foetal liver, but over time it has become apparent that purifying markers for foetal HSCs do not necessarily match ones used for adult or UCB-derived HSCs. A number of markers have been reported in both mouse and human over the years, including CD143, EPCR, CD144, GPI-80, CD110, but no consensus has been reached for defining cell surface phenotype for foetal HSCs (H. Iwasaki et al., 2010; Jokubaitis et al., 2008; Kobayashi et al., 1996; Prashad et al., 2015; Vodyanik et al., n.d.). It is also possible that in the developing embryo,

physiological demands are quickly changing and requiring of foetal haematopoiesis to adapt accordingly. Some of these adaptations occur in one vital foetal haematopoietic organ – the foetal liver (FL).

1.3.2 The foetal liver

In the FL, HSCs have been shown to rapidly expand through repopulation studies in mouse (Ema & Nakauchi, 2000). The peak of HSC numbers in the mouse embryo has been estimated to be at around 16 post-conception days, which is the equivalent of 12 pcw in human (Ema & Nakauchi, 2000; Otis & Brent, 1954). Murine HSCs derived from the FL have also been shown to be highly proliferative and capable of significantly higher reconstitution capacity than adult bone marrow HSCs, which alludes to significant functional differences between adult and foetal HSCs (Harrison et al., 1997; Morrison et al., 1995; Rebel et al., 1996). The murine FL haematopoietic output also goes through waves of changes with the majority of generated cells being of erythroid lineage in early development, and lymphoid and myeloid lineages gradually accumulating with increasing development, which may be to align with the changing needs of the growing foetus (Morrison et al., 1995). Single-cell RNAseq of different stages of foetal liver has confirmed similar findings in human with the contribution of erythroid lineages declining and gradual increase in lympho-myeloid outputs with increasing gestational age (Popescu et al., 2019). A transient population of murine foetal HSC with FLT3 expression capable of producing developmentally restricted lymphoid cells have also been reported in mouse embryos (Beaudin et al., 2016). Identifying such HSC populations that appear only briefly during development has significant implications for understanding the establishment of the innate and adaptive immune systems.

Peak of foetal HSC expansion correlates with the onset of migration to the newly formed bone marrow niches. Little is known about the mechanism responsible for the colonization of both the liver and bone marrow, but factors such as SDF-1a and its receptor, CXCR4, have been implicated in mouse (Kawabata et al., 1999; Nagasawa et al., 1996). The organisation of haematopoietic cell migration from the FL to the bone marrow and other haematopoietic organs has not been well-defined,

but short-term repopulating progenitors have been isolated from embryonic murine bone marrows prior to seeding with HSCs (Christensen et al., 2004).

Microscopy and gene expression studies looking at the murine foetal liver have reported observations that expanding HSCs associate with $Nestin^{+}/NG2^{+}$ pericytes of the portal vessels (Khan et al., 2016). Interestingly, there is a synchronicity between the branching of the portal vessels and the expansion of HSCs, suggesting the existence of a potential interplay between the two processes. In contrast, $Nestin^{+}/NG2^{+}$ pericytes in the bone marrow have been shown to be important in maintaining the quiescence of HSC (Kunisaki et al., 2013). Comparing both pericyte populations from the foetal liver and adult bone marrow may give interesting insights into the potential mechanisms that regulate this switch between active cycling and long-term quiescence.

1.4 Immunodeficient mouse models and their limitations

Evidence of the limitations of immune-deficient mouse strains can be seen in some of the earliest reports relying on SCID mice. Baum and colleagues demonstrated in the early 90s that human $CD34^{+}/CD90^{+}$ cells could repopulate SCID mice, while the $CD34^{+}/CD90^{-}$ had very little SRC capacity. A few decades later Notta and Laurenti, demonstrated in the superior NSG that $CD90^{-}$ cells could yield long-term engraftment *in vivo.*

Even the NSG mouse model, widely acknowledged as the most permissive for human xenografts requires irradiation conditioning to allow small numbers of human repopulating cells to colonise the bone marrow. Irradiation of NSG mice is a necessary step to free up niche space for donor human haematopoietic stem cells, and it can be used to assess SRC potential under physiological conditions and uncover populations with stem cell properties that would otherwise be overlooked in traditional *in vitro* assays. The irradiation, inevitably, causes global long-lasting damage to both the haematopoietic and stromal cells in the bone marrow niche, the effects of which can last up to weeks after the procedure (Cao et al., 2011). Several studies have shown that irradiation has a negative impact not only on the haematopoietic compartment, but also the components of the

niche. In mice, 1-day post-irradiation, significant decrease in the number mesenchymal stromal cells (MSCs), endothelial cells, osteoblasts and other cell types have been shown (Cao et al., 2011; Severe et al., 2019). While, MSCs appear to be more resistant to irradiation damage than haematopoietic and other niche cells, it is easy to speculate that a compromised bone marrow microenvironment, containing lower numbers of some cell types and altered cytokine profile, may provide a niche that is only suitable for some repopulating cells (Severe et al., 2019; Sugrue et al., 2014).

A widely-acknowledged challenge with NSG mouse model was that its lineage readout for human HSCs was skewed toward the production of B cells and not really representative of normal haematopoiesis (Brehm et al., 2010; Covassin et al., 2013). Presumed to be due to the lack of the cytokines required for the development of other lineages, an NSG mouse model was created with overexpression of human SCF, IL-3 and GM-CSF (Nicolini et al., 2004). This model, NSG-SGM3 (NSG-S), was shown to be more successful at producing a more balanced lineage output, but would also result in increased mobilisation of HSCs and consequently lower and shorter reconstitution (Jangalwe et al., 2016; Nicolini et al., 2004).

Another enhanced immune-deficient mouse model created to address the challenges on irradiation conditioning was the NSG$^{W41/W41}$ strain, which harboured mutations in the KIT (CD117) gene (Cosgun et al., 2014; McIntosh et al., 2018). The native HSCs are functional but have impaired self-renewal and repopulation capacity, which makes injected human HSCs more competitive and successful at occupying niche space in the bone marrow (Cosgun et al., 2014). While this model provides an undamaged niche for the homing human HSCs, it still does not resolve the issue that niche cells will be producing only murine cytokines and growth factors that wouldn't necessarily bind human receptors effectively.

Without a doubt, much advancements have been made since the first attempts to test the reconstitution capacity of human HSPCs *in vivo*. New enhanced NSG mouse models allow the engraftment and tracking of lineage output for populations that were previously undetectable. However, it is important

to note that there is still much to be desired and no murine mouse model has been yet generated that can mimic the human bone microenvironment. Ultimately, the readouts of these *in vivo* assays represent only the HSCs capable of homing and proliferating, but not necessarily of all HSCs being tested.

1.5 Project objectives

1.5.1 CD34$^-$ HSCs

Twenty years after the self-renewal and differentiation potential of CD34$^-$ haematopoietic cells was first reported, their role in human haematopoiesis still remains elusive. Considering their highly quiescent nature, yet robust repopulation capacity, it is easy to speculate that the CD34$^-$/CD38$^-$/CD93$^+$ population serves as an HSC reservoir contributing to blood regeneration only under conditions of stress and injury rather than steady state haematopoiesis.

With estimated SRC frequency significantly lower than other defined HSCs, the first goal of this project was to purify the CD34$^-$/CD38$^-$/CD93$^+$ population. Improving the SRC frequency is crucial for the meaningful interpretation of the functionality of the HSCs residing within the CD34$^-$ compartment, where only a single cell out of 6000 cells represents a potential HSC.

By identifying CD117 as new positive selection marker we demonstrate a significantly improved frequency of repopulating cells within the CD34$^-$ population and reassert its distinct functional output through *in vivo* assays compared to CD34$^+$ HSCs.

We next endeavoured to elucidate the gene expression profile of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$. Using single cell RNA-seq we report potential new markers that can be used to enhance the purity of this population and molecular features suggestive of early lineage priming within subfractions of the CD34$^-$ HSC population.

1.5.2 Haematopoietic stem cell heterogeneity during ontogeny

Decades of active research have established with certainty that foetal haematopoiesis is distinct from the adult one, both on the functional and phenotypic

level. Multiple reports on different HSC purifying cell surface markers during early human development have asserted the disparity between adult and foetal HSCs, but little has been done to unify existing data into a map of phenotypic and functional foetal HSC dynamics during different gestational stages and transition of haematopoietic cells from the foetal liver to the bone marrow.

To address these questions, we optimised a high-parameter flow cytometry panel of previously described foetal HSC markers and applied it to mononuclear cells derived from foetal liver and bone marrow samples derived from late first and early second trimester. Using dimensionality reduction and self-organizing map clustering we analysed phenotypic dynamics across tissues and developmental stages.

Chapter 2. Materials and Methods

2.1 Primary human material processing

2.1.1 Umbilical cord blood

Umbilical cord blood (UCB) was obtained after informed consent at the Royal London Hospital according to approved protocols by the East London and Research Ethics Committee. Two or more UCB samples were pooled for each experiment, and mononuclear cells (MNCs) were isolated by density centrifugation using Ficoll-Plaque (GE Healthcare). Red blood cells were lysed with ammonium chloride at room temperature for 10 minutes.

Depletion of lineage cells was performed by immunomagnetic removal of cells positive for mature haematopoietic markers using StemSep Human Hematopoietic Progenitor Cell Enrichment Cocktail (STEMCELL Technologies) according to the manufacturer's instructions. All experiments were carried out with MNCs in PBS containing 2% FBS (2% FBS/PBS). When frozen, cells were resuspended in FBS supplemented with 10% dimethyl sulfoxide (DMSO) and stored at -80°C for short-term storage or liquid nitrogen for long-term storage. All mononuclear cells and resulting xenografts were stored in accordance to the Human Tissue Act.

2.1.2 Foetal liver and foetal bone marrow

Human foetal samples were obtained from the Human Developmental Biology Resource (HDBR), approved by the National Research Ethics Service. Written consent for tissue donation is obtained from women post decision to undergo termination of pregnancy in HDBR-collaborating clinics.

Dissected foetal liver and foetal bones samples were obtained from the Institute of Child Health (London) and transferred to the Francis Crick Institute in serum-free medium. Bones were cleaned and flushed with 2% FBS/PBS, and red cells were lysed with ammonium chloride at room temperature for 10 minutes. Cells were frozen in FBS with 10% DMSO.

Foetal liver samples were passed through a 60 µm filter into 2% FBS/PBS and red cells were lysed with ammonium chloride at room temperature for 10

minutes. Samples from stages of 12 weeks post conception or over were subjected to density gradient separation with Ficol-Plaque to obtain mononuclear cells and frozen with FBS with 10% DMSO. Cells from stages earlier than 12 weeks post conception were frozen directly after red blood cell lysis.

All samples were stored in accordance to the Human Tissue Act.

2.2 Cell culture

2.2.1 Adherent cell culture

MS-5 cells were obtained from the Crick Cell Services STP and maintained in culture with DMEM supplied with 10% FBS and 1% Pen/Strep up to passage six. Embryonic mouse stromal cells lines were a gift from Elaine Dzierzak (Edinburgh) and maintained according to supplied instructions and used for experiments up to passage six.

2.2.2 Long-term haematopoietic stem cell culture and limiting dilution assays

Plates and dishes were coated with collagen solution (0.3 mg/ml) and washed. MS-5 or other stromal support cells were plated and left to reach 90% confluency before receiving irradiation dose of 7.5 Gy from a ^{137}Caesium source. Media was substituted for Myelocult™ H5100 (StemCell Technologies) and human haematopoietic cells were added to the culture. Every week by half of the well volume and replenished with fresh H5100. After six weeks of culture, haematopoietic cells were separated from stromal cells by removing Sca-1-antibibody labelled cells with a PE-positive Selection Kit (StemCell Tehnologies). For limiting dilution assays, cells were replated in limiting dilutions on stroma in a 96-well plate, pre-coated with collagen and cultured for further 2 weeks. Then media was removed and replaced MethoCult™ 4435 was added. Wells with positive colony growth were scored after 2 weeks of culture.

2.2.3 Colony-forming assays

Sorted cells were spun down and resuspended in MethoCult™ 4434 (StemCell Technologies) and cultured at 37°C at 20% oxygen and 5% carbon dioxide. After two weeks, wells were assessed on an inverted microscope and different colony types were scored according to phenotype. Representative images of the different colony types were taken for reference.

2.3 Immunophenotyping and flow cytometry

2.3.1 Antibody staining

Cells were incubated with antibodies at room temperature for 15 minutes, or 4°C for 30 minutes, in the dark in 50 µl 2% FBS/PBS. Samples were subsequently washed and suspended in DAPI (4', 6'-diamidino-2-phenylindole) at 0.5 µg/ml or Sytox Green™ (ThermoFischer Scientific) according to manufacturer's instruction in 2% FBS/PBS.

For multicolour experiments single colour controls of stained OneComp compensation beads (BD) were used for compensation. Gates were set to exclude non-viable cells, debris and doublets during analysis.

2.3.2 Ki67 antibody staining

For cell cycle analysis cells were fixed and permeabilised with Fixation/Permebilisation Solution Kit (BD) according to manufacturer's instructions stained with 4 µl of anti-Ki67- APC. Staining was performed for 1 hour at 4°C. Cells were washed in 2% FBS/PBS and resuspended in 2% FBS/PBS with DAPI 0.5 µg/ml.

2.3.3 Flow cytometry and fluorescence-activated cell sorting

Flow cytometry was conducted on Fortessa analysers maintained by the Crick Flow Cytometry STP. FACS sessions were conducted on Aria Fusion (BD) cell

sorters for UCB-derived tissues and on Aria III (BD) by members of the Crick Flow Cytometry STP.

Cells were passed through a 70 µm filter prior to analysis and sorting to prevent blockages. At the beginning of every analysis or cell sorting voltage thresholds were set using Fluorescence Minus One (FMO) controls and unstained controls. Manual compensations were run unless more than 10 distinctly conjugated antibodies were used. Gating thresholds were set up according to FMO controls with the exclusion of debris, doublets and dead cells.

For FACS, the 100 µm nozzle was used. Samples were sorted at 4°C into 2% FBS/PBS. Purity checks were routinely, and samples were resorted if purity was less than 95%.

2.4 Transcriptomics

2.4.1 RNA extraction

Cells were pelleted after sorting and supernatant removed. RNA was extracted using the RNeasy Microkit (Qiagen) according to manufacturer's instructions. RNA was eluted in RNase-free water. The RNA concentration was quantified on a Nanodrop Spectrophotometer (Thermo Scientific). RNA quality was quantified by BioAnalyzer. Samples were stored at -80^{o}C.

2.4.2 Bulk RNA-sequencing

RNA-seq was conducted by the Advanced Sequencing Facility at the Francis Crick Institute. Sequencing was performed on biological triplicates for each condition tested, and it generated libraries ranging 20-30 million, 100 bp paired end reads.

Raw file processing was performed by the Crick bioinformatics team. Read trimming and adapter removal were performed using Trimmomatic version 0.36. Reads were aligned using STAR version 2.5.2a to the human reference genome, Ensembl GRCh38, release 89 and RSEM package version 1.2.31 was used to obtain gene level counts.

2.4.3 Single-cell RNA-sequencing

Single cell RNA-seq was performed by the Crick Advance Sequencing Facility. Single cells were sorted and counted to ensure optimum concentration for cell capture. Cell capture and libraries preparation were performed on the Chromium-10x instrument according to manufacturer's instructions using the 10x v.3' kit. Single cell libraries of 100 bp paired-end reads were pooled and sequenced on HiSeq 4000.

Raw file processing was performed by the Crick bioinformatics team. Read trimming and adapter removal were performed using Trimmomatic version 0.36. Reads were aligned using STAR version 2.5.2a to the human reference genome, Ensembl GRCh38, release 89 and RSEM package version 1.2.31 was used to obtain gene level counts.

2.5 Bioinformatics

2.5.1 Single Cell RNA-seq analysis

Downstream analysis was conducted in collaboration with Manuel Garcia-Albornoz of the Bonnet Group. Quality control, normalization and scaling of data was performed using the Seurat (version 4.0.0) within R version 3.6.2. Dimensionality reduction and principal components were generated and visualized using Seurat and the ggplot2 package (version 3.3.0). Annotated gene lists for cell cycle analysis were pulled from ensemble using the ensembledb (version 2.14.0) and AnnotationHub (version 2.22.0) packages. Gene ontology analysis was performed with the packages ClusterProfiler (version 3.18.1) and EnrichPlot (version 1.10.2). GSE analysis was conducted using the fgsea (version 1.16.0) package. Differentially expressed gene lists were obtained with the Seurat package and filtered for adjusted p values less than 0.05.

2.5.2 High-parameter flow cytometry analysis

Compensated flow cytometry files were loaded in R using the FlowCore (version 2.2.0) package. The data was transformed using hyperbolic arsin with a cofactor of 250. Dimensionality reduction was performed using the CATALYST (version 1.14.0) package and clusters were calculated using the SOM algorithm and projected onto UMAP plots. For each sample, 2 x 10^4 cells were analysed.

2.6 *In vivo* work

2.6.1 Mice

All work was performed in accordance with Home Office regulations and with the approval the Francis Crick Institute ethics committee. NOD/SCID/IL2Rγ^{Null} (NSG), NSG-SGM3 (NSG-S) and NSG-Kit$^{W41/W41}$ (C-Kit^{Mutant}) were bred at the Francis Crick Institute and housed under sterile conditions.

2.6.2 Primary transplantations

NSG and NSG-S mice (between 8 and 12 weeks of age) were irradiated with 3.5 Gy from a ^{137}Cesium source up to 24 hours prior to transplantation and administered Baytril for two weeks post-irradiation. C-Kit mice were not irradiated or administered Baytril.

Tail vein injection were conducted under sterile condition and the cell suspension was a delivered in a maximum volume of 150 µl using a 30G needle.

2.6.3 Intrabone sampling and engraftment analysis

Analysis of human engraftment was conducted through bone marrow aspirate or total bone marrow analysis. Prior to aspirations, mice were anaesthetized though inhalation of isoflurane and a 30G needle was used to puncture the tibia at the tibiofemoral joint. A maximum of 15 µl of BM was aspirated.

For total bone marrow analysis, mice were culled through cervical dislocation and tibia, femur and ileac crest were dissected bilaterally. Bones were cleaned and bone marrow was removed through centrifugation. Red blood cell lysis was performed. Cells were resuspended in 2% FBS/PBS for immunophenotyping.

2.6.4 Secondary transplantation

For secondary transplantation primary xenografts were obtained as described. For transplants of human $CD45^+$, mouse cells were depleted using the EasySep™ Mouse/Human Chimera Isolation Kit (StemCell Tehcnologies). Human cell purity was assessed with flow cytometry. For secondary transplantations of specific populations, cells were FACS-sorted as described. Secondary mouse recipients were injected intravenously as described.

Chapter 3. Haematopoietic Stem Cell Heterogeneity in Postnatal Haematopoiesis

3.1 Introduction

Despite evidence that the human $CD34^-$ haematopoietic compartment contains cells with potent long-term SRC activity, the population remains largely ignored (Anjos-Afonso et al., 2013; Bhatia et al., 1998; Danet et al., 2002; Goodell et al., 1997a; Zanjani et al., 1998). The $CD34^+/CD38^-$ population was the prime candidate of the search for the human HSC with its well-reported capacity to self-renew and differentiate using LTC-IC and CFC assays. On the other hand, the *in vitro* self-renewal and differentiation capacity of $CD34^-$ cells is very limited, making the study of $CD34^-$ HSCs possible only through *in vivo* repopulation assays (Anjos-Afonso et al., 2013; Bhatia et al., 1999; Danet et al., 2002). Furthermore, the SRC frequency in the $CD34^-/CD38^-$ population was estimated in the range of one per hundreds of thousands of cells (Bhatia et al., 1998). This frequency was significantly improved with the discovery that CD93 positively selected for $CD34^-$ SRCs, but the new frequency was still estimated at 1/6000 cells (Danet et al., 2002). The low purity of the population was hindering efforts to identify the pathways and molecular processes maintaining it, while the fact it could only be properly assayed through *in vivo* repopulation of mouse recipients was adding another challenge to studying $CD34^-/CD38^-/CD93^+$ HSCs.

Identifying better isolation strategies to facilitate the study of $CD34^-$ HSCs was the first task undertaken by groups interested in this elusive population. A Japanese group investigated $CD34^-$ expressing CD110 (receptor for thrombopoietin), CD133 (PROM1), and GPI-80, demonstrating in separate studies that all of these markers can enrich for $CD34^-$ SRCs (Abe et al., 2017; Matsuoka et al., 2016, 2017; M. Takahashi et al., 2014). By sorting $Lin^-/CD34^-/CD133^+/GPI\text{-}80^+$ cells they demonstrated the SRC frequency in the population to be between the astounding 1/4.9 and 1/8.1 (Abe et al., 2017). Single-cell gene expression analysis highlighted that this population is staggeringly similar to the $CD34^+$ HSCs, but gene-set enrichment analysis (GSEA) indicated processes such as hypoxia, angiogenesis, Wnt signalling and pathways involved in megakaryocyte

and erythroid differentiation to be enriched in $Lin^-/CD34^-/CD133^+/GPI\text{-}80^+$ (Sumide et al., 2018). While these findings were encouraging and of significant interest to the work conducted in this thesis, Sumide and colleagues have highlighted that only around 1% of the $Lin^-/CD34^-/CD133^+/GPI\text{-}80^+$ population express CD93, suggesting that they might be investigating a distinct $CD34^-$ population.

While this is possible, *in vivo* transplantation of as many as 840, 000 $CD34^-/CD38^-/CD93^-$ cells in conditioned NOD/SCID mice have failed to yield engraftment (Danet et al., 2002). This discrepancy could be attributed to the fact that the CD93 antibody used by Sumide and colleagues is different from the one used in this thesis. In the past decade, our group has identified a CD93 antibody clone (VIMD2) that stains $CD34^-$ SRCs with high degree of reproducibility (Anjos-Afonso et al., 2013). The antibody clone used by Sumide and colleagues (X-2, Origene) hasn't been previously tested by us, and thus we cannot comment on its specificity. However, published data from other groups has shown that some CD93 antibody clones can produce different staining patterns. A study concluding CD93 is a functional marker of leukaemia stem cells (LSCs) in MLL-rearranged acute myeloid leukaemia (AML) was making use of CD93 antibody that marked only 0.31% of normal UCB-derived $CD34^+/CD38^-$ cells, while it has been previously reported that CD93 is expressed by the majority of cells in this population (Danet et al., 2002; M. Iwasaki et al., 2015). By using the exact same antibody for GPI-80, we attempted to investigate the expression patterns of GPI-80 within the $CD34^-$ and $CD34^+$ populations (Fig. 3.1). We conducted seven independent experiments using pools of mononuclear cells derived from umbilical cord blood. Prior to staining every sample underwent immunomagnetic lineage depletion to eliminate mature haematopoietic cells expressing markers CD2, CD3, CD10, CD14, CD16, CD19, CD20, CD24, CD56, CD66b, CD235a. In addition, our antibody staining panel included a commercially-available lineage cocktail (CD2, CD3, CD10, CD14, CD19, CD20, CD56) as well as CD7 (naïve T cells), CD11b (myeloid cells), and CD45RA, which we used to ensure exclusion of lineage-committed cells. We expected to find the majority of $GPI\text{-}80^+$ cells within the $CD34^-/CD38^-/CD93^{Low/-}$ population as Sumide and colleagues highlight

that only about 1% of cells within their population of interest defined as CD34$^-$/CD133$^+$/GPI-80$^+$ express CD93 (Sumide et al., 2018). We observed that the proportions of GPI-80$^+$ cells were indeed higher in CD34$^-$/CD38$^-$/CD93$^{Low/-}$ (Mean=6.6, SD=4.8, Range: 0.6 to 13.8, n=7) compared to the CD34$^-$/CD38$^-$/CD93$^+$ population (Mean=3.7, SD=2.1, Range: 1.5 to 7.4, n=7) (Fig. 3.1E-G). We observed lower proportion of cells expressing GPI-80 within the CD34$^+$/CD38$^-$ fraction (Mean=1.8, SD=2.0, Range: 0.2 to 6.0, n=7) (Fig. 3.1D). However, if only about 1% of Sumide's CD34$^-$/CD133$^+$/GPI-80$^+$ are CD93$^+$, 99% would be CD93$^-$. In terms of absolute cell numbers, a larger number of GPI-80$^+$ cells are present in the CD34$^-$/CD38$^-$/CD93$^{Low/-}$ population (Fig. 3.1F) compared to CD34$^-$/CD38$^-$/CD93$^+$ (Fig. 3.1E), however we did also detect a significant number of GPI-80-expressing cells in the latter.

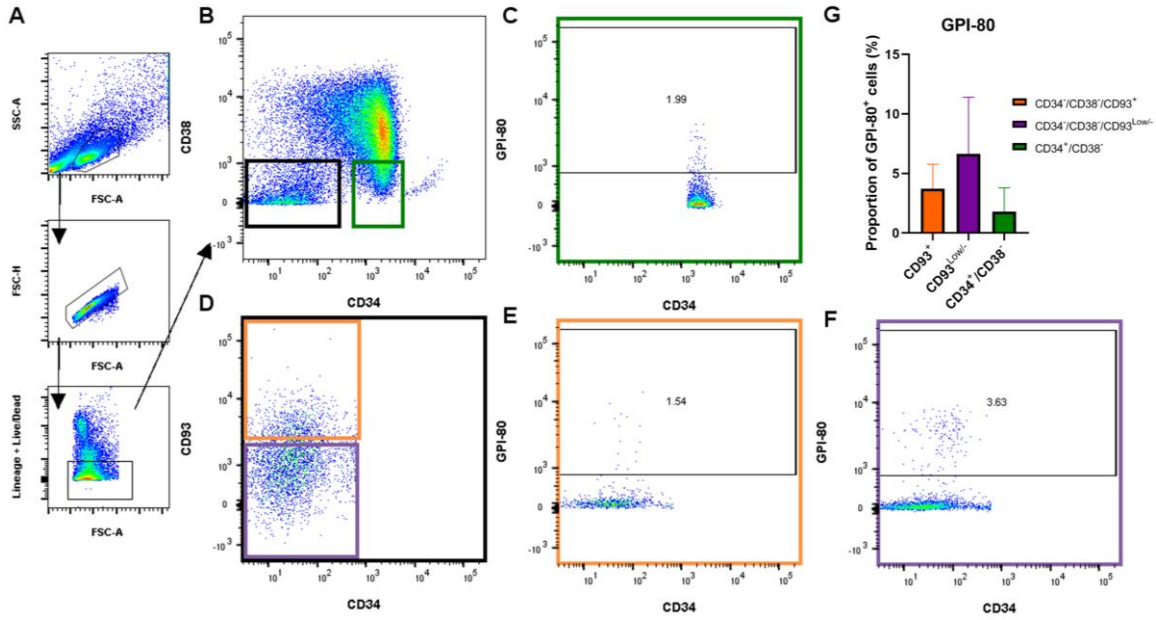

Figure 3.1. GPI-80 expression within the CD34+ and CD34- fractions in umbilical cord blood mononuclear cells

A. Representative flow cytometry plots for gating strategy of mononuclear cell population, singlets, and CD45RA$^-$/Lineage$^-$/Live cells.
B. Representative flow cytometry plot for CD34 versus CD38 expression and gating strategy for CD34$^-$/CD38$^-$ and CD34$^+$/CD38$^-$ populations (black and green gates, respectively).
C. Representative plot for CD93 expression by CD34$^-$/CD38$^-$ population and gating strategy for CD93$^+$ and CD93$^{Low/-}$ populations (orange and purple gates, respectively).

D. Representative flow cytometry plot of the expression pattern of GPI-80 by $CD34^+/CD38^-$ cells.
E. Representative flow cytometry plot of the expression pattern of GPI-80 by $CD34^-/CD38^-/CD93^+$.
F. Representative flow cytometry plot of the expression pattern of GPI-80 by $CD34^-/CD38^-/CD93^{Low/-}$.
G. Quantification of GPI-80^+ cells as percentage values of parent populations $CD34^-/CD38^-/CD93^+$, $CD34^-/CD38^-/CD93^{Low/-}$, and $CD34^+/CD38^-$ (orange, purple and green bars, respectively (n=7, error bars represent standard deviations).

As the reported gating strategy for the $CD34^-/CD133^+$/GPI-80^+ population does not exclude $CD38^+$ cells, we investigated whether GPI-80 is also expressed by cells within the $CD34^-/CD38^+$ as well as the $CD34^+/CD38^+$ $CD34^+/CD38^-$, and $CD34^-/CD38^-$ populations. We observed the highest proportion of GPI-80-expressing cells within the $CD34^-/CD38^-$ fraction (Mean=6.01%), (Fig. 3.2B & F). In contrast, less than 1% of $CD34^+/CD38^+$ cells were GPI-80^+ (Fig. 3.2C & F). Around 3.6% of $CD34^-/CD38^+$ cells were GPI-80^+, a lower frequency of expression compared to the $CD34^-/CD38^-$ population (Fig. 3.2B, E & F). While at smaller proportion, it appears that the $CD34^-/CD133^+$/GPI-80^+ population could contain $CD38^+$ cells. We attempted to incorporate CD133 into the gating strategy to fully evaluate the frequencies of $CD34^-/CD133^+$/GPI-80^+ cells in the $CD34^-$ compartment, but despite using the same CD133 antibody as Sumide and colleagues we never succeeded in detecting the population due to the fact we observed less than 1% of $CD34^-$ cells expressing CD133 and almost no overlap with GPI-80 expression. $CD133^+$/GPI-80^+ have been reported to constitute an average of 6% of the $CD34^-$ population, but we could not reproduce these results despite using mononuclear cells derived from UCB, similar strategy to eliminating lineage-committed cells, and using the same CD133 and GPI-80 antibody clones (Sumide et al., 2018). The purpose of these initial analyses was to validate the existence of the $CD34^-/CD133^+$/GPI-80^+ population and confirm whether it overlaps with the $CD34^-/CD38^-/CD93^+$, but we have been unable to detect the $CD34^-/CD133^+$/GPI-80^+ population. We nevertheless, confirmed the existence of GPI-80^+ cells in both the $CD34^-/CD38^-/CD93^+$ and $CD34^-/CD38^-/CD93^-$ compartments. However, the reported enrichment of GPI-80^+ cells as a percentage of the $CD34^-$ compartment was an average of 14%, which is

significantly higher than the enrichments we observed in the CD34$^-$/CD38$^-$ and CD34$^-$/CD38$^+$ fractions.

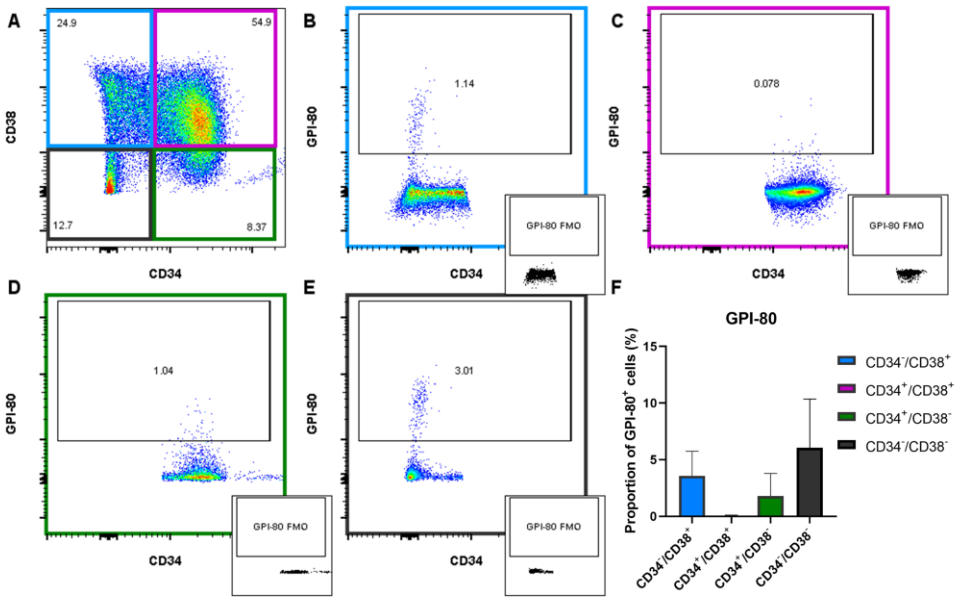

Figure 3.2. GPI-80 expression within the CD34+/-/CD38+/- populations

A. Representative flow cytometry plot for CD34 versus CD38 expression and gating strategy for CD34$^-$/CD38$^+$, CD34$^+$/CD38$^+$, CD34$^+$/CD38$^-$, and CD34$^-$/CD38$^-$ populations (blue, pink, green, and black, and green gates, respectively).
B – E. Representative flow cytometry plots of GPI-80 versus CD34 expression and gating of GPI-80$^+$ cells in (B) CD34$^-$/CD38$^+$, (C) CD34$^+$/CD38$^+$, (D) CD34$^+$/CD38$^-$, and (E) CD34$^-$/CD38$^-$ populations and corresponding FMO controls.
F. Quantification of GPI-80$^+$ cells as percentage values of parent populations CD34$^-$/CD38$^+$, CD34$^+$/CD38$^+$, CD34$^+$/CD38$^-$, and CD34$^-$/CD38$^-$ (n=7, error bars represent standard deviations).

Seeing that the major caveat of studying CD34$^-$ HSCs is reaching sufficient population purity, to pursue an in depth molecular characterisation at the single cell level, our first aim was to identify new markers or validate the ones reported in recent years, to allow us to further purify CD34$^-$ HSCs. We first performed FACS analysis and then validated *in vivo* the functionality of the putative marker we identified, CD117. Lastly using limiting dilution analysis, we evaluated the level of purity being achieved via the addition of positive CD117 selection.

Another significant difficulty in the study of CD34$^-$ SRCs has been not only that they engraft slower than CD34$^+$ SRCs, but also their low repopulation frequency (Anjos-Afonso et al., 2013). Injection of human SCF and G-CSF in

mice transplanted with $CD34^-$ SRCs has been previously used to enhance the level of engraftment of this population in NOD/SCID/β_2-microglobulin^{Null} and NSG mice (Anjos-Afonso et al., 2013). With the significant advancements in the generation of immunodeficient mouse strains, we also undertook comparisons of engraftment kinetics in several enhanced NSG mouse models. We tested engraftment in NSG-S, which overexpresses three human cytokines, and NBSGW (henceforth referred to as C-Kit^{Mutant}), which does not require irradiation to support engraftment of human cells.

In the onset of this project, we first undertook protein expression analysis of potential cell surface markers that could aid the isolation of $CD34^-$ SRCs at higher purity level. Then, after we had identified a candidate marker, we tested the engraftment ability in three different mouse strains in search for an *in vivo* model that could support the self-renewal of the population, while promoting higher engraftment.

3.2 Cell surface marker analysis

Drawing from studies in mouse, where CD93 was first identified as a marker of HSCs in foetal haematopoiesis, we looked into cell surface antigens reported to enrich for HSC fractions in both mouse and human foetal haematopoiesis (McKearn et al., 1985). CD144 (VE-cadherin) had been described as an HSC marker during ontogeny both in mouse and human (Kim et al., 2005; Oberlin et al., 2010). CD143 (ACE) has been similarly confirmed as a positive selection marker for HSCs during ontogeny as well as a $CD34^+$ HSC-associated gene in single-cell RNAseq datasets (Jokubaitis et al., 2008; Sinka et al., 2012; Velten et al., 2017). EPCR (CD201) and THBD (CD141) have been reported to play a functional role in maintaining HSC quiescence, while EPCR expression has also been confirmed in murine foetal liver cells with differentiation and self-renewal potential (Basu et al., 2020; H. Iwasaki et al., 2010; Subramaniam et al., 2019). CD133 (PROM1) has a long history as a stem cell marker in multiple tissues (Holmberg Olausson et al., 2014; Snippert et al., 2009; M. Takahashi et al., 2014). To this panel of markers, we added GPI-80, which had been recently reported as a $CD34^-$ HSC marker and CD117 (Kit), which previous

work from the group had indicated might be expressed by a fraction of the $CD34^-/CD38^-/CD93^+$ population (Danet et al., 2002).

To test for the expression of these markers we derived mononuclear cells from UCB and used immunomagnetic depletion to remove all cells expressing markers: CD2, CD3, CD10, CD14, CD19, CD20, CD24, CD56, CD66b, and CD235a. Cells were then stained with antibodies for flow cytometry analysis. Lineage markers CD7 and CD11b were further gated out, along with $CD45RA^+$ cells, doublets and dead cells (Fig. 3.3A).

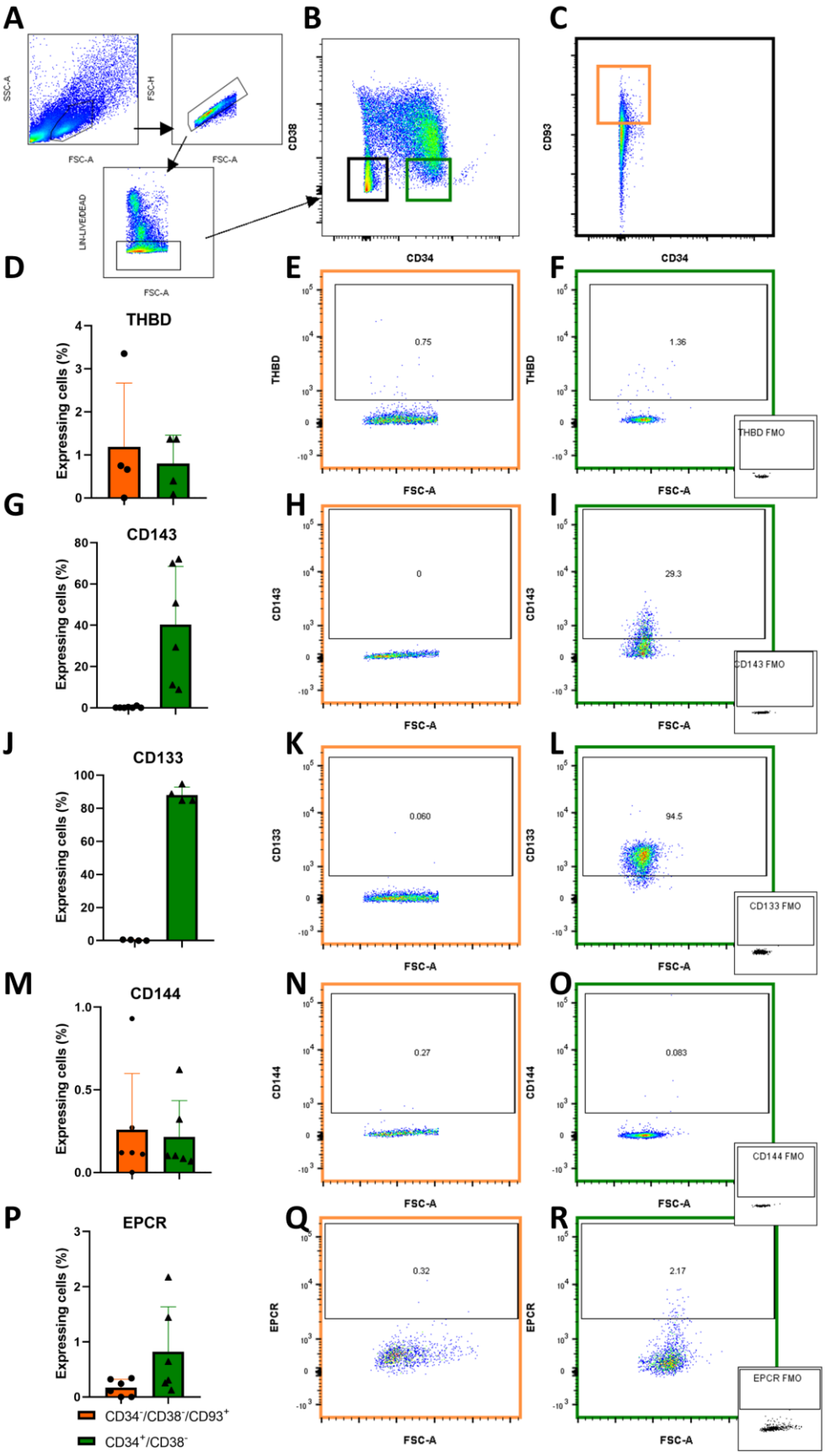
A
SSC-A
FSC-A
FSC-H
FSC-A
LIN-/LIVE/DEAD
FSC-A
B
CD38
CD34
C
CD93
CD34
D
THBD
Expressing cells (%)
E
THBD
0.75
FSC-A
F
1.36
THBD FMO
FSC-A
G
CD143
Expressing cells (%)
H
CD143
0
FSC-A
I
29.3
CD143 FMO
FSC-A
J
CD133
Expressing cells (%)
K
CD133
0.060
FSC-A
L
94.5
CD133 FMO
FSC-A
M
CD144
Expressing cells (%)
N
CD144
0.27
FSC-A
O
0.083
CD144 FMO
FSC-A
P
EPCR
Expressing cells (%)
CD34-/CD38-/CD93+
CD34+/CD38-
Q
EPCR
0.32
FSC-A
R
2.17
EPCR FMO
FSC-A

Figure 3.3. HSC-associated cell surface markers' expression by CD34-/CD38-/CD93+ and CD34+/CD38- populations

A-C. Representative flow cytometry plots for gating strategy of (A) "blast" population, single cells, Lineage-negative, and live cells; (B) $CD34^-/CD38^-$ (black gate) and $CD34^+/CD38^-$ (green gate) populations; (C) CD93-expressing $CD34^-/CD38^-$ cells (orange gate).
D-F. (D) Quantification of THBD-expressing cells as percentage values of parent populations $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ (n=4; orange and green bars, respectively); (E & F) Representative flow cytometry plots and gating of THBD-expressing cells in $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ populations (E and F, respectively).
G-I. (G) Quantification of CD143-expressing cells as percentage values of parent populations $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ (n=6; orange and green bars, respectively); (H & I) Representative flow cytometry plots and gating of CD143-expressing cells in $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ populations (H and I, respectively).
J-L. (J) Quantification of CD133-expressing cells as percentage values of parent populations $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ (n=4; orange and green bars, respectively); (K & L) Representative flow cytometry plots and gating of CD133-expressing cells in $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ populations (K and L, respectively).
M-O. (M) Quantification of CD144-expressing cells as percentage values of parent populations $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ (n=6; orange and green bars, respectively); (N & O) Representative flow cytometry plots and gating of CD144-expressing cells in $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ populations (N and O, respectively).
P-R. (P) Quantification of EPCR-expressing cells as percentage values of parent populations $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ (n=6; orange and green bars, respectively); (Q & R) Representative flow cytometry plots and gating of EPCR-expressing cells in $CD34^-/CD38^-/CD93^+$ and $CD34^+/CD38^-$ populations (Q and R, respectively).

Overall, we observed low expression of all investigated markers within the $CD34^-/CD38^-/CD93^+$ compartment. In comparison, CD133 and CD143 were expressed by considerable numbers of cells within the $CD34^+/CD38^-$ population at an average 88.1% and 40.3%, respectively (Fig. 3.3G, I, J, L). While the staining for CD133 was highly reproducible with standard deviation of 4.6, the numbers of $CD143^+$ cells within the $CD34^+/CD38^-$ compartment varied significantly between experiments with a standard deviation of 28.2% (Fig. 3.3G & J). Neither of these cell surface markers was expressed in the $CD34^-/CD38^-/CD93^+$ population (Fig. 3.3G, H, J, K). We observed an average of 1.19% of

$CD34^-/CD38^-/CD93^+$ cells and 0.8% of $CD34^+/CD38^-$ expressed THBD (Fig. 3.3D-F). A single outlier sample pulled the average number of THBD-expressing $CD34^-/CD38^-/CD93^+$ cells higher than the number of THBD-expressing $CD34^+/CD38^-$ cells. In addition, the small number of cells detected in the $THBD^+$ gate made it challenging to determine whether the marker separated cells into a distinct population. EPCR and CD144 were expressed by less than 0.5% of $CD34^-/CD38^-/CD93^+$ (Fig. 3.3M, N, P, Q). CD144 was expressed by a similar number of $CD34^+/CD38^-$ cells, an average of 0.22% (Fig. 3.3M & O). The average number of EPCR-expressing cells within the $CD34^+/CD38^-$ population was slightly higher at an average of 0.82% (Fig. 3.3P & R).

In summary, we observed only a small number of $CD34^-/CD38^-/CD93^+$ cells expressed the investigated markers. CD143 and CD133 were expressed by a high proportion of the $CD34^+/CD38^-$ population but not in the $CD34^-/CD38^-/CD93^+$, despite previous reports suggesting CD133 being expressed in the $CD34^-$ compartments. THBD was the only marker expressed by an average of over 1% of $CD34^-/CD38^-/CD93^+$ cells, but with high variation between independent experiments.

Another marker of interest, CD117, distinguished itself from the markers above with more reproducible staining pattern. Flow cytometry analysis revealed that CD117 is expressed by an average of 5.8% of $CD34^-/CD38^-/CD93^+$, while 96.2% of the $CD34^+/CD38^-$ fraction expressed CD117 (Fig. 3.4B). CD117-expressing cells within the $CD34^-/CD38^-/CD93^+$ population was in the range of 1.1-10.4% with a standard deviation of 3.6% (Fig. 3.4B). We also investigated the CD117 expression pattern within different HSPC subfractions defined by the expression of surface markers CD90 and CD49f. Within the population reported to be most enriched for long-term HSCs, $CD34^+/CD38^-/CD49f^+/CD90^+$, an average 97.0% of cells expressed CD117 (Fig 3.4B). The average percentage of $CD117^+$ cells within the $CD49f^+/CD90^-$, $CD49f^-/CD90^+$, and $CD49f^-/CD90^-$ populations were observed to be lower with 92.0%, 93.6%, and 82.1%, respectively (Fig 3.4B).

By looking into CD117 expression versus the forward scatter area (FSC-A), we also observed that $CD34^+/CD38^-/CD117^+$ cells were of similar size and

congregated into a distinct well-defined population. Interestingly, we observed that the $CD117^+$ cells within the $CD34^-/CD38^-/CD93^+$ fraction formed a similar population to that observed in $CD34^+/CD38^-$ cells, in addition to a number of cells that expressed lower CD117 levels. The relatively low but reproducible frequency of $CD34^-/CD38^-/CD93^+/CD117^+$ cells coupled with the observation that CD117 expression demarcated a distinct population within the compartment urged us to test $CD34^-/CD38^-/CD93^+/CD117^{+/-}$ for *in vivo* repopulation capacity.

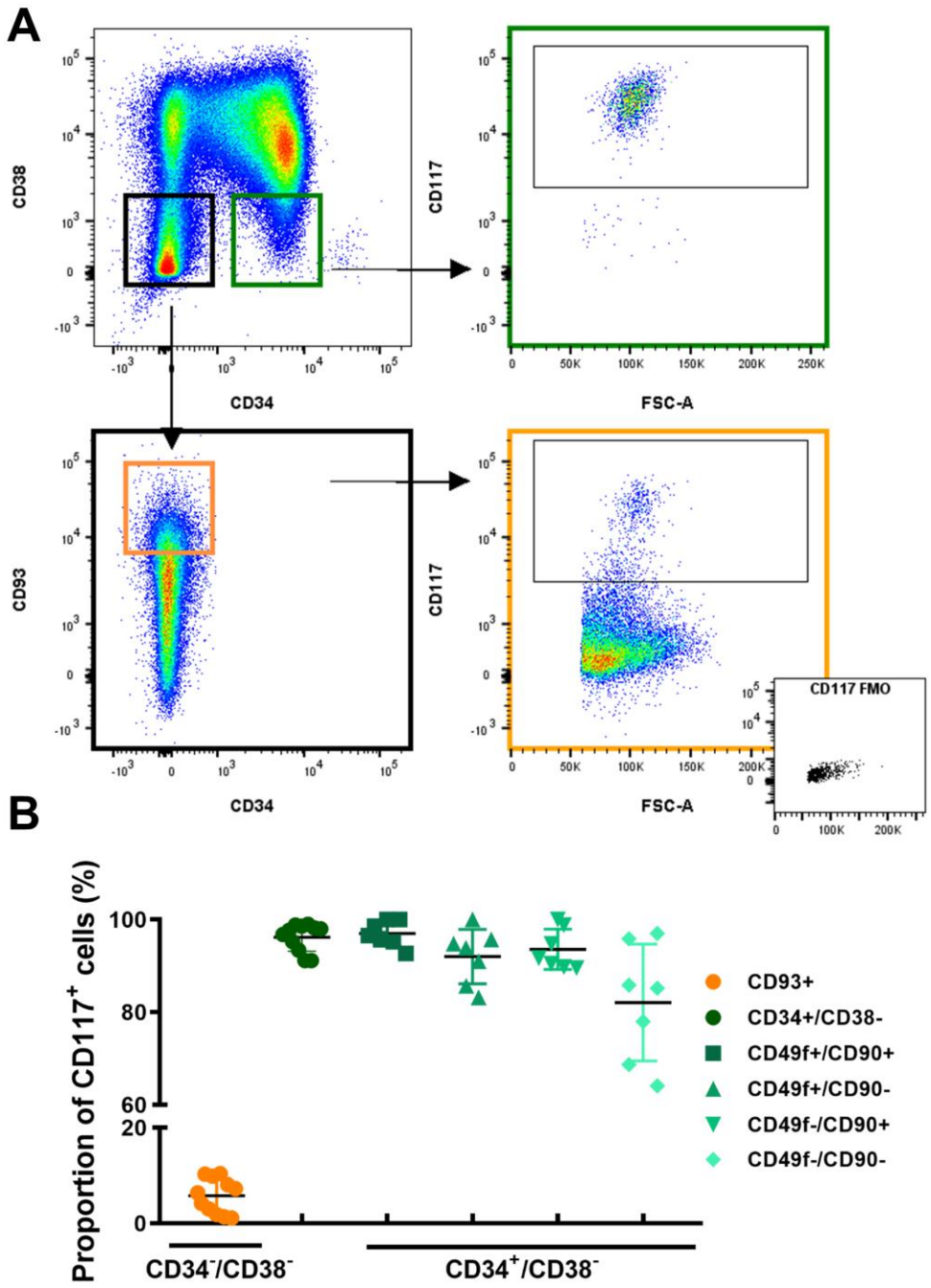

Figure 3.4. CD117 expression in CD34+ and CD34- subfractions

A. Representative flow cytometry plot for gating strategy of $CD34^{+}/CD38^{-}$ and $CD34^{-}/CD38^{-}/CD93^{+}$ and CD117 versus forward scatter-area phenotype plots with corresponding FMO control.
B. Positive cell counts for expression of CD117 as percentage values of the $CD34^{-}/CD38^{-}/CD93^{+}$ (n=11; Mean=5.82; SD=3.64), $CD34^{+}/CD38^{-}$ populations (n=11; Mean=96.20; SD=2.81), $CD34^{+}/CD38^{-}/CD49f^{+}/CD90^{+}$ (n=7; Mean=97.0; SD=2.7), $CD34^{+}/CD38^{-}/CD49f^{+}/CD90^{-}$ (n=7; Mean=92.0; SD=5.9), $CD34^{+}/CD38^{-}/CD49f^{-}/CD90^{+}$ (n=7; Mean=93.6; SD=4.34), and $CD34^{+}/CD38^{-}/CD49f^{-}/CD90^{-}$ (n=7; Mean=82.11; SD=12.63).

3.3 CD117 demarcates a $CD34^-/CD38^-/CD93^+$ subpopulation with a potent *in vivo* repopulation kinetics

To test the *in vivo* repopulation capacity of $CD34^-/CD38^-/CD93^+/CD117^{+/-}$ we sorted these two populations and carried our intravenous injections in sublethally irradiated NSG mouse recipients (Fig. 3.5A). At 24 weeks, the mice were culled and bone marrow harvested from the femur, tibia and ileac crest. Human engraftment and distribution of mature lineages was checked via flow cytometry. In the initial experiment, we detected human engraftment in 1 of 2 mice injected with 1, 500 $CD34^-/CD38^-/CD93^+/CD117^+$ accompanied by presence of human $CD33^+$, $CD19^+$, and $CD3^+$ cells (Fig. 3.5B-D). When we repeated the experiment with lower cell doses of 1, 000 and 600 cells, we also detected multi-lineage engraftment in 1 out of 3 mice and 2 out of 3 mice, respectively. The average engraftment was 1.1% with the highest engraftment estimated at 5.7% in a mouse injected with 1, 000 cells. The average normalized distribution of lineages was 43.8% $CD33^+$ cells, 54.5% of $CD19^+$ B cells, and 1.5% of $CD3^+$ T cells. Overall, the lineage distribution of engrafted mice indicated slight bias toward the B cell lineage. One mouse injected with 1, 000 $CD34^-/CD38^-/CD93^+/CD117^+$ exhibited a more pronounced bias toward the B cell production with approximately 98% of $hCD45^+$ cells being $CD19^+$ (Fig. 3.5D).

In contrast, injection of as many as 10, 000 $CD34^-/CD38^-/CD93^+/CD117^-$ cells failed to generate engraftment by 24 weeks (Fig. 3.5B).

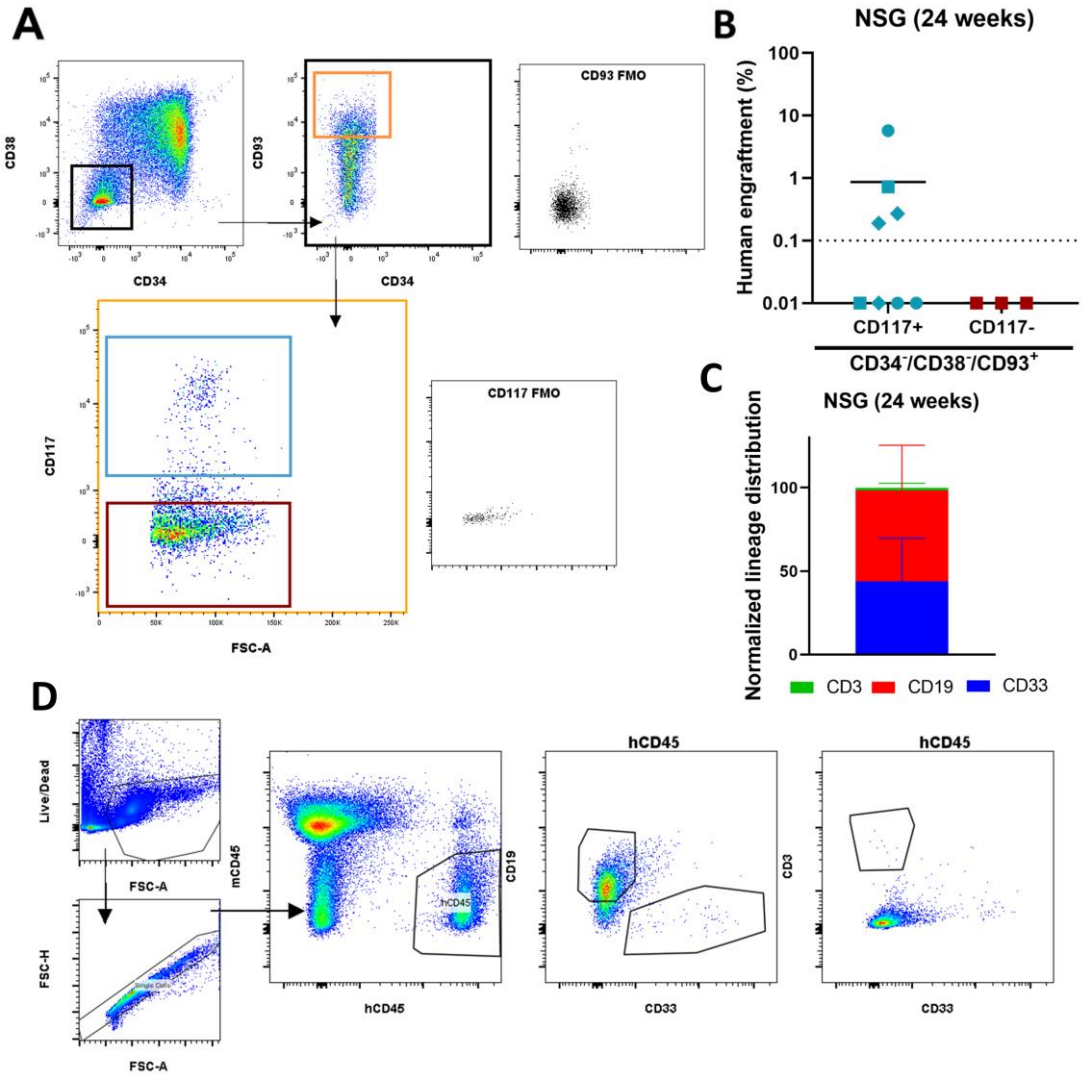

Figure 3.5. Human engraftment of CD34-/CD38-/CD93+/CD117+/- in NSG mice

A. Representative flow cytometry plots for gating strategy for $CD34^-/CD38^-$ cells (black gate), $CD34^-/CD38^-/CD93^+$ (orange gate),$CD34^-/CD38^-/CD93^+/CD117^+$ (blue gate) and $CD34^-/CD38^-/CD93^+/CD117^-$ (red gate). Corresponding CD93 and CD117 FMO controls shown on the right.
B. Quantification of human chimerism as percentage of human $CD45^+$ ($hCD45^+$) cells of all single live cells. NSG mice were injected with $CD34^-/CD38^-/CD93^+/CD117^+$ (n=7, blue) or $CD34^-/CD38^-/CD93^+/CD117^-$ (n=3, red). Different point shapes represent independent experiments, where mice were injected with different cells numbers (blue squares =1, 500; blue circles = 1, 000; blue diamonds= 600; red squares = 10, 000).
C. Quantification of lineage distribution normalised to total lineage-positive cells in NSG mice 24 weeks post-injection with $CD34^-/CD38^-/CD93^+/CD117^+$ (n=5).
D. Representative flow cytometry plots for gating strategy for assaying human engraftment ($hCD45^+$ cells), myeloid ($CD33^+$), B cell ($CD19^+$) and T cell ($CD3^+$) engraftment.

We next proceeded to investigate the repopulation kinetics of $CD34^-/CD38^-/CD93^+/CD117^{+/-}$ in NSG, NSG-S, and C-Kit^{Mutant}. NSG and NSG-S mice were conditioned with radiation, while C-Kit^{Mutant} mice received transplants without radiation conditioning. We sampled bone marrow from mice every 6 weeks by bone marrow aspiration of the tibia and sacrificed 24 weeks post-injection. Human chimerism and lineage output were quantified using flow cytometry (Fig. 3.6A).

Engraftment of $CD34^+$ SRCs can be usually detected 6 weeks post-transplantation, but $CD34^-$ SRCs engraftment is detected later at around 12 weeks post-transplantation (Anjos-Afonso et al., 2013). We observed few to no human cells in NSG and C-Kit^{Mutant} mice injected with $CD34^-/CD38^-/CD93^+/CD117^+$ at 6 weeks, but detected robust human engraftment in NSG-S mice at the same time-point (Fig 3.6A). The average engraftment in NSG-S mice at 6 weeks was estimated at 8.4% within a range of 1.3-34.5% (Fig. 3.6A). In contrast, the highest engraftment in NSG and C-Kit^{Mutant} mice injected with $CD34^-/CD38^-/CD93^+/CD117^+$ was 0.6% and 0.5%, respectively, at the same time point. 8 out of 8 NSG-S mice were engrafted at 6 weeks, while only 2 out of 6 NSG and 2 out of 9 C-Kit^{Mutant} mice showed low engraftment (Fig. 3.6A-C). More robust human engraftment could be observed in NSG and C-Kit mice injected with $CD34^-/CD38^-/CD93^+/CD117^+$ from 12 weeks onwards. Engraftment in the NSG strain remained relatively low throughout the experiment, ranging between 0.1 and 5% at 24 weeks (Fig. 3.6A). The multi-lineage output of lymphoid and myeloid cells could be detected between 6 and 24 weeks, with $CD19^+$ B cells being the most represented lineage (Fig. 3.6D). Engraftment of $CD34^-/CD38^-/CD93^+/CD117^+$ was highest in C-Kit^{Mutant} mice, between 67% and 86% at week 24 (Fig. 3.6C). Lineage distribution was similar between C-Kit and NSG with $CD19^+$ B cells being most abundant differentiated lineage (Fig. 3.6F). In NSG-S $CD34^-/CD38^-/CD93^+/CD117^+$ cells not only engrafted earlier than in NSG and C-Kit^{Mutant} mice, but also generated a lineage output that was predominated by $CD33^+$ myeloid cells (Fig. 3.6B, E, H).

In comparison, long-term human engraftment was not detected in mice injected with $CD34^-/CD38^-/CD93^+/CD117^-$ (Fig. 3.6A-C). Small transient engraftment was observed in one NSG-S mouse at 6 weeks and one C-Kit^{Mutant} mouse at 12 weeks, both injected with 10, 000 $CD34^-/CD38^-/CD93^+/CD117^-$ cells (Fig. 3.6B-C). However, no differentiated lineages could be detected. No human cells were

detected in NSG mice injected with as many as 10, 000 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ cells (Fig. 3.6A). These experiments indicated that the CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ population is not highly enriched with SRCs capable of generating multi-lineage engraftment in NSG, NSG-S and C-Kit^{Mutant} mouse stains. To completely dismiss the existence of competent SRCs within this population, further validations with larger numbers of cells would be required.

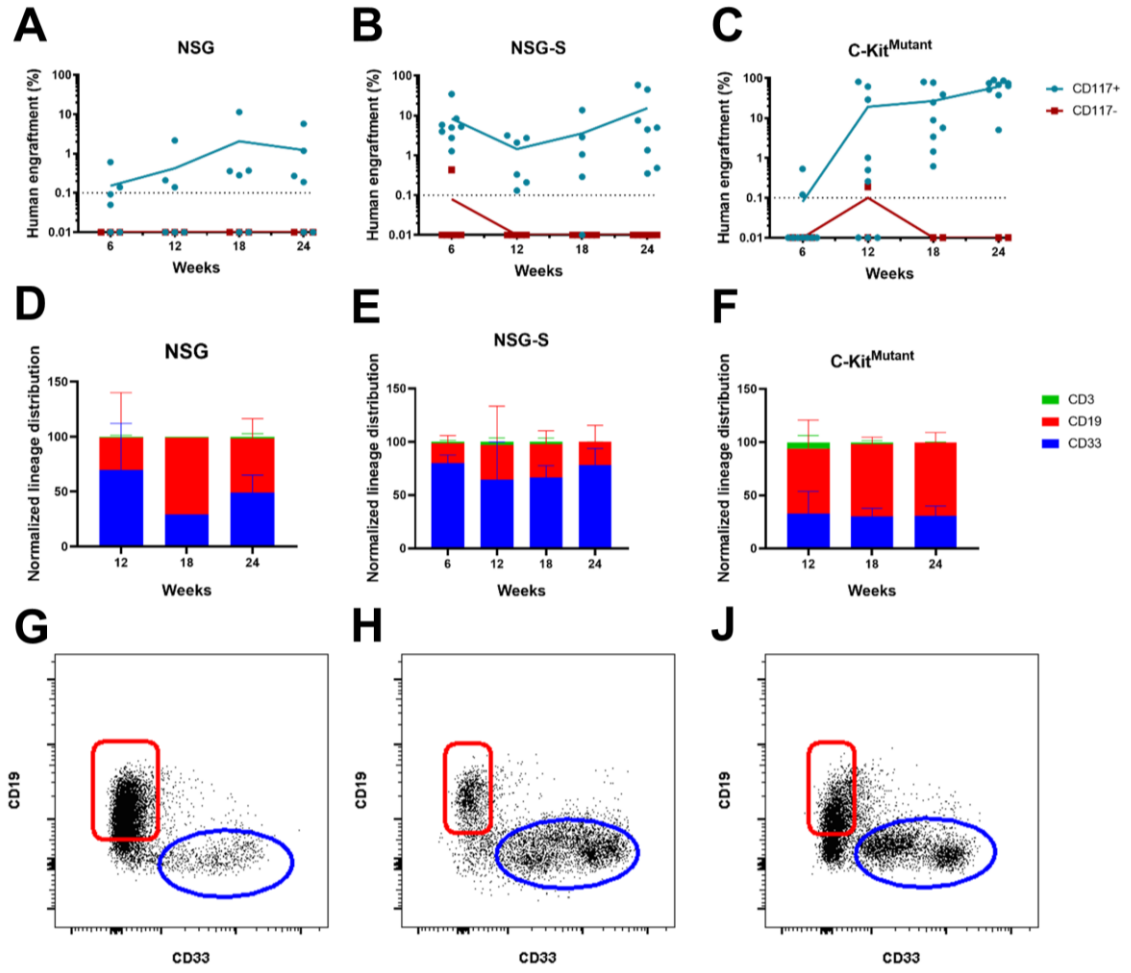

Figure 3.6. Human engraftment kinetics and lineage output of CD34-/CD38-/CD93+/CD117+ in NSG and enhanced NSG mouse strains

A-C. Quantification of human chimerism as percentage of human CD45$^+$ cells of all single live cells in bone marrow aspirates (weeks 6-18) and total bone marrow harvested (week 24). Mice were injected with CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ (blue circles) or CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ (red squares). (A) NSG mice were injected intravenously with 1,500-600 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells (n=6) and 10, 000 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ (n=3). (B) NSG-S mice were injected with 1, 500-400 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ (n=8) or 10, 000 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ (n=8). (C) C-Kit^{Mutant} mice were injected with 1, 000-400 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ (n=9) or 4, 000 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ (n=2).

D-F. Normalized lineage distribution for NSG (D), NSG-S (E), and C-Kit^{Mutant} (F) mice engrafted with CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells.
G-I. Representative flow cytometry plots of CD19$^+$ B cell (red gates) and CD33$^+$ myeloid cell (blue gates) distribution in NSG (G), NSG-S (H), and C-Kit^{Mutant} (I) mice engrafted with CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ 18 weeks post-injection.

In summary, we observed sustained human engraftment of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ in three different immunodeficient mouse strains with multilineage output. We also observed that cells in NSG-S yielded engraftment as early as 6 weeks, which could be attributed to the human cytokines overexpressed in this strain of mice. Additionally, we observed very high sustained engraftment in the C-Kit^{Mutant} strain from 12 weeks post injection. Doses of as many as 10, 000 cells of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ failed to engraft, indicating that the population likely contains very limited numbers of SRCs and by applying positive CD117 selection to our gating strategy we significantly enrich for CD34$^-$ SRCs, detecting engraftment with 600 cells, when without CD117 a minimum of 6, 000 cells is required to obtain engraftment.

3.4 CD34$^-$/CD38$^-$/CD93$^-$ subpopulation does not yield human engraftment irrespective of CD117 expression

Prompted by the recent study identifying a population of potent CD34$^-$ SRCs that had almost no CD93 expression, but expressed GPI-80, we proceeded to investigate human engraftment of CD34$^-$/CD38$^-$/CD93$^-$ cells (Sumide et al., 2018). In previous studies of this cell population has not yielded robust *in vivo* engraftment even by injection of cell doses as big as 100, 000 cells (Danet et al., 2002). However, we observed that a subset of the CD34$^-$/CD38$^-$/CD93$^-$ expressed CD117. In addition, we detected a subpopulation of CD34$^-$/CD38$^-$/CD93$^-$ cell expressing distinctly higher levels of CD117 than in the CD34$^-$/CD38$^-$/CD93$^+$ subset (Fig. 3.7A-C). To test self-renewal and differentiation capacity transplantations of CD34$^-$/CD38$^-$/CD93$^-$/CD117$^{+/-}$ in immunodeficient mice were undertaken. For this purpose, we used the NSG-S relying on our findings that in this mouse strain, CD34$^-$ SRCs can establish

engraftment quicker that in the NSG and C-Kit^{Mutant}. We transplanted mice with CD34$^-$/CD38$^-$/CD93$^-$/CD117$^+$ cell doses ranging between 200 and 500 cells, and approximately 3, 000 CD34$^-$/CD38$^-$/CD93$^-$/CD117$^-$ cells. We tracked human chimerism via bone marrow aspirations at six, twelve, and eighteen weeks post-injection, and analysed total bone marrow at 24 weeks. We didn't detect the presence of any human cells in mice injected with CD34$^-$/CD38$^-$/CD93$^-$/CD117$^+$ except for one mouse 24 weeks post-injection that received a cell dose of 200 cells (Fig. 3.7D). The engraftment was only 0.2% and no differentiated lineages were detected. This result could also not be reproduced in subsequent experiments where higher doses of CD34$^-$/CD38$^-$/CD93$^-$/CD117$^+$ were injected. We detected no engraftment at any time point in mice injected with 3, 000 CD34$^-$/CD38$^-$/CD93$^-$/CD117$^-$ cells.

In summary, the CD34$^-$/CD38$^-$/CD93$^-$/CD117$^{+/-}$ fractions do not appear enriched for cells with SRC capacity (Table 3.1). Our gating strategy, which excludes cells that may be expressing CD93 at intermediate or low levels could be also excluding SRCs described by Sumide and colleagues (Fig. 3.7A) (Sumide et al., 2018). It is possible that CD34$^-$/CD38$^-$/CD93$^-$/CD117$^{+/-}$ contain SRCs that could not be captured in our assay due to the low cell doses or because the NSG-S mouse strain does not provide suitable microenvironment and cytokine profile to drive the self-renewal of these particular populations. Our study of the CD34$^-$/CD38$^-$/CD93$^-$/CD117$^{+/-}$ populations as potential sources of SRCs are highly limited by the low cell numbers injected and more experiments are required to confirm whether these compartments are entirely devoid of SRC activity. However, having compelling evidence of the success of CD93 as a positive selection marker for CD34$^-$ SRC isolation, we decided to focus our interest on the CD93$^+$ fraction of the CD34$^-$/CD38$^-$

population.

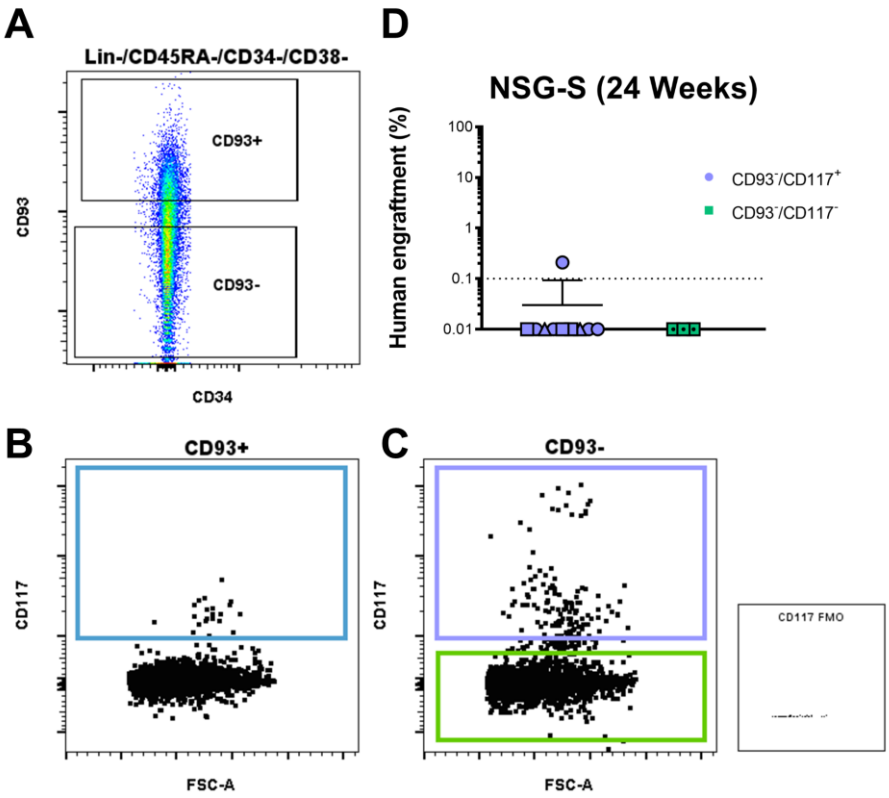

Figure 3.7. Transplantation of CD34-/CD38-/CD93-/CD117+/- does not yield stable human engraftment and multilineage output

A-C. Representative flow cytometry plots and gating for (A) $CD34^{-}/CD38^{-}/CD93^{+}$ and $CD34^{-}/CD38^{-}/CD93^{-}$ populations and (B) $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+}$ expression phenotype and (C) $CD34^{-}/CD38^{-}/CD93^{-}/CD117^{+/-}$ populations.
D. Quantification of human engraftment NSG-S mice at 24 weeks post-injection with $CD34^{-}/CD38^{-}/CD93^{-}/CD117^{+/-}$ populations. Point shapes represent independent experiments (circles – 200 cells, n=5; triangles – 400 cells, n=2; squares - 500 cells, n=3; Green squares – 3, 000 cells).

Population phenotype	**Injected cell dose**	**Engrafted/Tested**
$CD34^{-}/CD38^{-}/CD93^{-}/CD117^{+}$	200	1/5
	400	0/2
	500	0/3
$CD34^{-}/CD38^{-}/CD93^{-}/CD117^{-}$	3000	0/3

Table 3.1. Breakdown of injected cell doses of CD34-/CD38-/CD93-/CD117+/- in NSG-S mice and corresponding engraftment rates.

3.5 CD117 significantly purifies CD34$^-$/CD38$^-$/CD93$^+$ SRCs

After establishing that transplantation of doses between 600 and 1, 500 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells yielded stable multi-lineage engraftment in NSG, NSG-S and C-Kit^{Mutant} mice, we speculated that the SRC frequency of the population is potentially much higher than 1/1000 cells. Therefore, we undertook a limiting dilution analysis (LDA) in NSG-S mice, injecting doses of as little as 100 cells. In addition, we injected CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ cells to confirm our earlier results that the population contained very limited, if any at all, SRCs that can repopulate NSG-S mice.

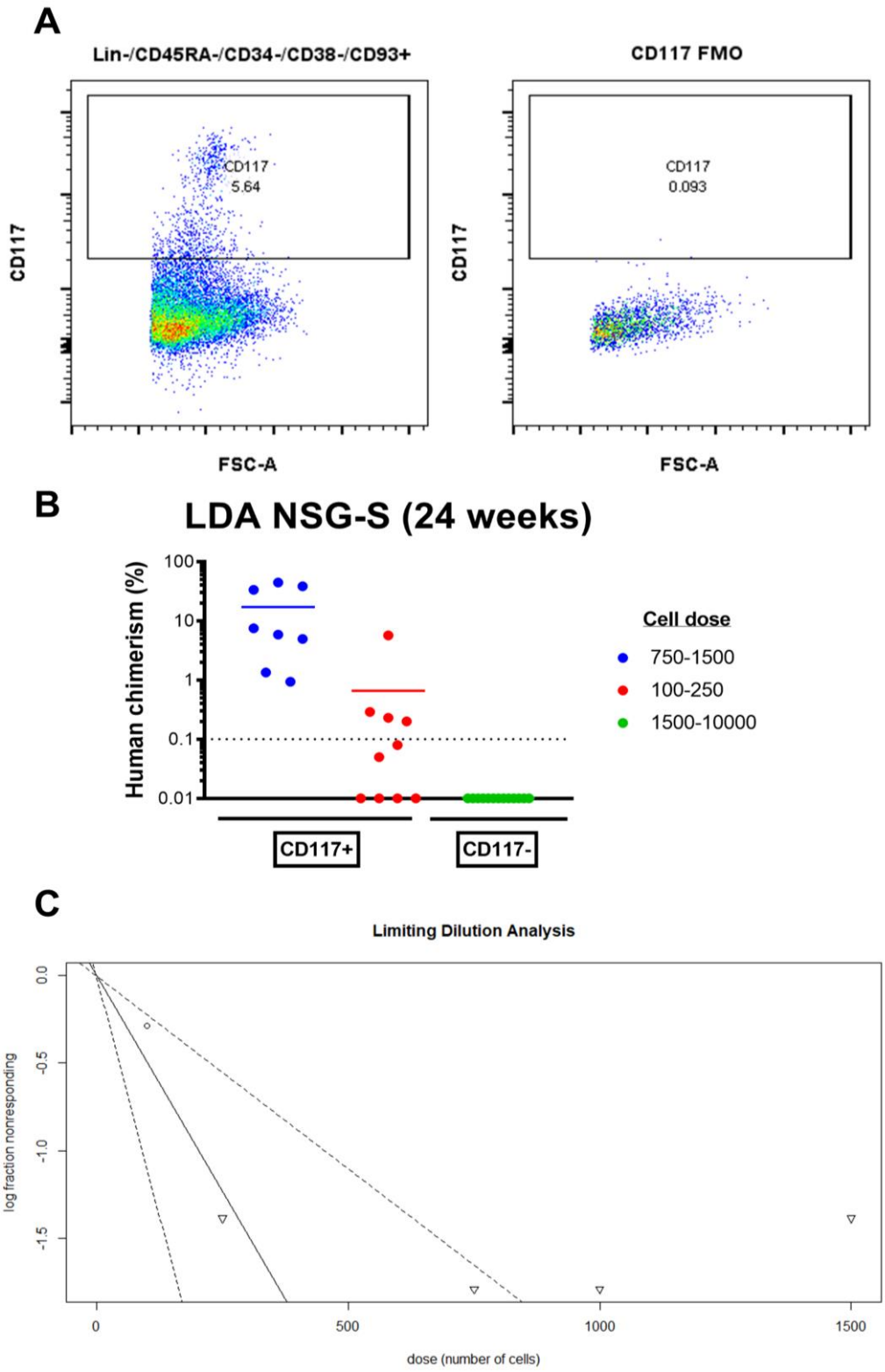

Figure 3.8. Limiting dilution analysis of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells

A. Representative flow cytometry plots for gating strategy for CD117$^+$ cells in the CD34$^-$/CD38$^-$/CD93$^+$ population and CD117 FMO control.
B. Quantification of human chimerism at 24 weeks post-injection in NSG-S mice injected with limiting doses of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ and CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$ cells. Figure represents collation of data from four independent experiments, including data represented in figures 4 and 3 (cell doses 1,500 and 1, 000 for CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$).
C. Limiting dilution analysis plot visualising injected CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cell doses and log of proportion of non-engrafted mice.

Cell dose	**Number of mice injected**	**Number of mice engrafted**
1500	2	2
1000	3	3
750	3	3
250	2	2
100	8	2

Table 3.2. Break-down of injected CD34-/CD38-/CD93+/CD117+ cell doses in NSG-S mice and engraftment outcome at 24 weeks post-injection.

In addition to our data showing positive human engraftment for cell doses of 1, 500 and 1, 000 (Fig. 3.6), we injected mice with cell doses of 750, 250 and 100 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells in two independent experiments (Fig 3.8A). We observed human engraftment in all mice injected with 750 and 250 cells (Fig. 3.8B Table 3.2.). Furthermore, we detected human repopulation in two out of eight mice injected with as little as 100 cells (Fig. 3.8B, Table 3.3.). Using these data, we calculated the estimated stem cell frequencies, using the R package *statmod*, which uses maximal likelihood estimation (Hu & Smyth, 2009). The stem cell frequency was estimated at 1/203 cells, with the upper and lower 95% confidence intervals estimated at 1/90.7 and 1/453 cells, respectively. Compared to the previous stem cell frequency for the CD34$^-$/CD38$^-$/CD93$^+$ population, estimated at about 1/6000 cells, positive selection for CD117 improves stem cell frequency by 30-fold. However, due to the fact only one of the tested cell doses resulted into a mix of non-responding vs responding mice, effectively making a single cell dose determining in the HSCs frequency estimation, it is possible this new HSC frequency is inaccurate. Further experiments with additional cell doses and more replicates would be required to confirm the validity of this LDA.

The previously estimated 1/6000 SRC frequency for the CD34$^-$/CD38$^-$/CD93$^+$ population was calculated based on repopulation kinetics in NSG mice. We have calculated a new SRC frequency in NSG-S and found that the addition of CD117 to the CD34$^-$ SRC gating strategy significantly purifies the population for repopulating cells to an average of one SRC per 203 CD34$^-$/CD38$^-$/CD93$^+$ cells. While these experiments suggest positive selection for CD117 is an effective purification strategy for CD34$^-$ SRCs, the 30-fold CD34$^-$ SRC improvement may be also attributed to the fact the NSG-S was used in our limiting dilution experiments; we have shown NSG are not as supportive of CD34$^-$ SRC engraftment as NSG-S and C-Kit^{Mutant} mouse strains (Fig. 3.6). Nonetheless, these LDAs in NSG-S assert the power of CD117 as a marker for CD34$^-$ SRCs and the utility of this mouse strain as a model system for confirming HSC functionality.

3.6 High expression of CD117 within the CD34$^-$/CD38$^-$/CD93$^+$ defines the populations with the most potent *in vivo* repopulation potential

In all initial experiments validating CD117 as a positive selection marker for CD34$^-$ SRCs, we sorted CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells by drawing the gate for all positive events based on the CD117 FMO control. We observed a distinct population expressing the marker as well as a small number of cells expressing lower levels of CD117 but still falling in the positive CD117 gate (Fig. 3.9A). We also observed the CD34$^-$/CD38$^-$/CD93$^-$ population contained a subset of cells expressing very high levels of CD117, which was absent in the CD34$^-$/CD38$^-$/CD93$^+$ population. These cells, however, yielded no human engraftment.

On the other hand, CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ reliably repopulate immunodeficient mouse strains and positive CD117 selection significantly improved the SRC frequency of the CD34$^-$/CD38$^-$/CD93$^+$ population. Reports on murine haematopoiesis have indicated that low expression of CD117 delineates quiescent long-term HSCs (Shin et al., 2014). Therefore, we undertook a series of experiments to establish whether CD34$^-$ SRC can be further defined by the level of CD117 expression and aid us in improving the CD34$^-$ SRC frequency by refinement of the gating strategy.

Relying on FMO controls we sorted $CD34^-/CD38^-/CD93^+$ into $CD117^{Hi}$ and $CD117^{Mid}$ subsets, and injected the cells intravenously into NSG, NSG-S and C-Kit^{Mutant} mice (Fig. 3.9A). We tracked the engraftment over a period of 24 weeks. At week 24 we sacrificed the mice, harvesting bone marrow and spleen to check for final human chimerism.

In mice injected with the $CD117^{Hi}$ fraction we observed a reproducible engraftment pattern that mimicked observations in transplantation experiments using $CD34^-/CD38^-/CD93^+/CD117^+$ cells (Fig. 3.9B, E, H, Fig. 3.6). In NSG, transplantation of 500 $CD117^{Hi}$ yielded low engraftment, while doses of 200 cells failed to engraft (Fig. 3.9B). At 24 weeks, mice transplanted with 500 $CD117^{Hi}$ cells were repopulated with $CD19^+$ B cells and $CD33^+$ myeloid cell in both the bone marrow and the spleen (Fig. 3.9C & D). Almost no $CD3^+$ T cells could be detected in NSG as we previously observed in transplantations of $CD117^+$ (Fig. 3.9C-D, Fig. 3.6). In C-Kit^{Mutant} mice as little as 200 $CD117^{Hi}$ cells generated multi-lineage human engraftment that steadily increased from 12 weeks post-transplantation until the end of the experiment at 24 weeks (Fig. 3.9H). Spleens of engrafted C-Kit^{Mutant} mice were also repopulated with $CD19^+$ B cells and $CD3^+$ T cells (Fig. 3.9D, Fig. 3.9F). In NSG-S we observed reproducible engraftment of 200 $CD117^{Hi}$ cells, detected as early as 6 weeks post-transplantation (Fig. 3.9C). The human engraftment in these mice was overall lower than in C-Kit^{Mutant} mice (Fig. 3.9E). As previously observed in transplantations of $CD117^+$ cells in NSG-S, $CD117^{Hi}$ also generated a lineage distribution biased toward $CD33^+$ myeloid cells (Fig. 3.9F & G).

In NSG mice injected with as many as 500 $CD117^{Mid}$ cells, no human engraftment was detected (Fig. 3.9B). In C-Kit^{Mutant} mice we observed very low engraftment in a single mouse that persisted throughout the experiment, while in two additional mice we detected very low engraftment at 24 weeks (Fig. 3.9H). This human engraftment constituted entirely of $CD19^+$ B cells (Fig. 3.9I). One of these mice had 4.1% human engraftment in the spleen, comprised mostly of $CD19^+$ B cells (Fig 3.9H & J). In NSG-S mice, we detected engraftment in a single mouse transplanted with 500 $CD117^{Mid}$ cells. It persisted throughout the experiment, and at 24 weeks we also detected splenic engraftment (0.13%) (Fig. 3.9E). The bone marrow engraftment in these two mice was similar to that in mice transplanted with

$CD117^{Hi}$ cells with the most abundant lineage being $CD33^{+}$ myeloid cells and a smaller fraction of $CD19^{+}$ B cells (Fig. 3.9F). One of the mice had a low splenic engraftment that consisted mostly of $CD33^{+}$ myeloid cells (Fig. 3.9G)

From these series of experiments, we observed robust and steady engraftment of the $CD117^{Hi}$ fraction of the $CD34^{-}/CD38^{-}/CD93^{+}$ population. Mice repopulated with $CD117^{Hi}$ cells also had detectable human engraftment in the spleen. In addition, multilineage engraftment comprised of $CD19^{+}$ B cells and $CD33^{+}$ myeloid cells was detected, in all mouse strains injected with the $CD117^{Hi}$ fraction. $CD3^{+}$ T cells were detected in both the bone marrow and spleen in NSG-S and C-Kit^{Mutant} mice transplanted with $CD117^{Hi}$ cells. We observed limited human engraftment in some mice injected with $CD117^{Mid}$ cells, but predominantly at 24 weeks, with engraftment below 1% in most cases and no $CD3^{+}$ cells were detected in any of the mice injected $CD117^{Mid}$ cells. In C-Kit^{Mutant} mice reconstituted with $CD117^{Mid}$ cells, the lineage distribution was almost entirely comprised of $CD19^{+}$ B cells in both the bone marrow and the spleen. On the other hand, in NSG-S mice reconstituted with $CD117^{Mid}$ cells, we observed both myeloid and B cell engraftment. However, engraftment was detected only in 2 out of 9 mice (Table 3.4). The low engraftment and limited lineage reconstitution suggest that the $CD117^{Mid}$ fraction of the $CD34^{-}/CD38^{-}/CD93^{+}$ population holds poorer reconstitution potential than the $CD117^{Hi}$ fraction. We cannot discount the possibility that this low engraftment may suggest the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Mid}$ is highly quiescent and requires longer period of time to establish substantial engraftment. However, due to time constraints we could not extend *in vivo* experiments longer than 24 weeks. Therefore, we proceeded to test how much we could improve the stem cell frequency of the population by refining our gating strategy to $CD117^{Hi}$ cells.

Strain	**Cell Dose**	**Engrafted/Tested**	
		$CD117^{Hi}$	$CD117^{Mid}$
NSG	200	3/3	*
	500	2/2	0/4
NSG-S	200	1/1	*
	250	*	1/2
	300	*	0/3
	500	4/4	1/4

C-Kit^{Mutant}	200	3/3	*
	500	3/3	2/6

Table 3.3. Breakdown of cell doses and engrafted mice in repopulation assays for CD34^{-}/CD38^{-}/CD93^{+}/CD117$^{Hi/Mid}$.

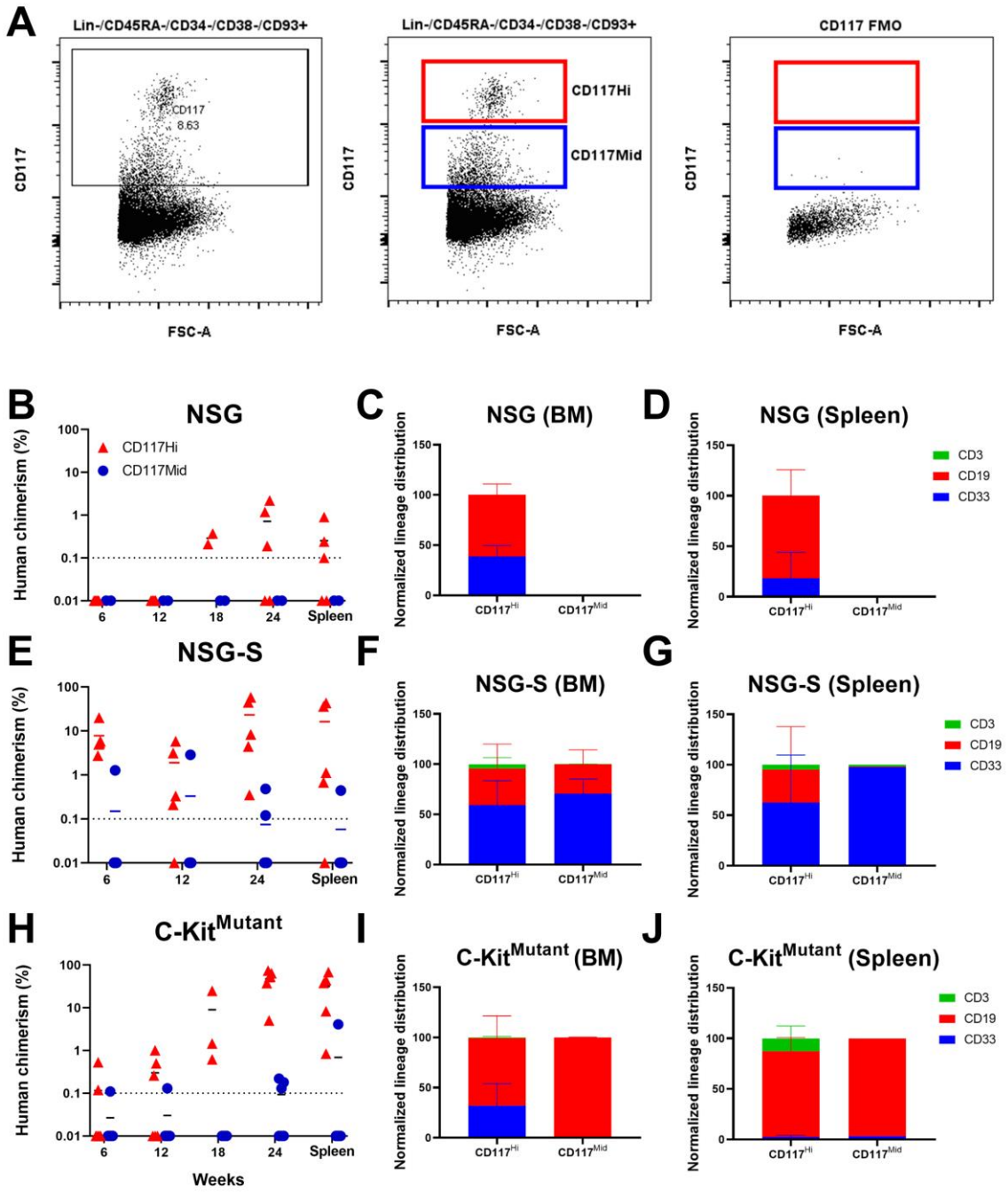

Figure 3.9. High expression of CD117 delineates the CD34^{-}/CD38^{-}/CD93^{+} cells with most robust *in vivo* repopulation capacity

A. Representative flow cytometry plot of the gating strategy for $CD117^{Hi}$ (red gates) and $CD117^{Mid}$ (blue gates) $CD34^{-}/CD38^{-}/CD93^{+}$ cells, and corresponding CD117 FMO control.
B. Quantification of human chimerism in radiation-conditioned NSG mice. Doses of injected $CD117^{Hi}$ cells were 200 (n=3) and 500 cells (n=2). Doses of injected $CD117^{Mid}$ cells were 500 (n=4).
C-D. Normalised lineage distribution in (C) bone marrow and (D) spleens of NSG mice 24 weeks after transplantation of $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi/Mid}$ cells. (C) $CD117^{Hi}$ – n=3; (D) $CD117^{Hi}$ – n=3;
E. Quantification of human chimerism in radiation-conditioned NSG-S mice. Doses of injected $CD117^{Hi}$ cells were 200 (n=1), 500 (n=4). Doses of injected $CD117^{Mid}$ cells were 500 (n=4) and 250 (n=1).
F-G. Normalised lineage distribution in (F) bone marrow and (G) spleens of NSG-S mice 24 weeks after transplantation of $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi/Mid}$ cells. (F) $CD117^{Hi}$ – n=5; $CD117^{Mid}$ – n=2; (G) $CD117^{Hi}$ – n=5; $CD117^{Mid}$ – n=2;
H. Quantification of human chimerism in $C\text{-}Kit^{Mutant}$ mice. Doses of injected $CD117^{Hi}$ cells were 200 (n=3) and 500 (n=3). Doses of injected $CD117^{Mid}$ were 500 (n=6).
I-J. Normalised lineage distribution in bone marrow (I) and spleens (J) of $C\text{-}Kit^{Mutant}$ mice 24 weeks after transplantation of $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi/Mid}$ cells. (I) $CD117^{Hi}$ – n=6; $CD117^{Mid}$ – n=3; (J) $CD117^{Hi}$ – n=6; $CD117^{Mid}$ – n=1;

3.7 Selection for high-expressing CD117 cells within the $CD34^{-}/CD38^{-}/CD93^{+}$ improves SRC frequency

To investigate whether selection for $CD117^{Hi}$ cells could further improve the SRC frequency within the $CD34^{-}/CD38^{-}/CD93^{+}$ population, we undertook LDA analysis by injecting limiting doses of $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ cells in NSG-S and $C\text{-}Kit^{Mutant}$ mice. As we had previously observed human engraftment in mice injected with as few as 200 hundred cells, we injected cell doses of approximately 50, 100, and 200 in $C\text{-}Kit^{Mutant}$ mice. We sacrificed the mice 24 weeks post-injection and detected human engraftment in two out of three mice injected with 200 cells, six out of six mice injected with 100 cells, and six out of eight mice injected with just 50 cells (Table 3.5, Fig. 3.10A). In NSG-S we detected less robust engraftment at 24 weeks. Two out of four mice were engrafted after transplantation with 200 and 400 cells, one out four mice was engrafted with a dose of 100 cells, and three out of four mice with a dose of 50 cells (Table 3.5, Fig. 3.10B). While engraftment levels were low at 24 weeks, multilineage output could still be observed in both $C\text{-}Kit^{Mutant}$ and NSG-S (Fig. 3.10C-D). Furthermore, the normalized lineage distribution correlated with the typical lineage output generated in $C\text{-}Kit^{Mutant}$ and NSG-S mice transplanted

with $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs. $CD19^+$ B cells were the predominant lineage in the $C\text{-}Kit^{Mutant}$ strain, while $CD33^+$ myeloid cells were most abundant in NSG-S (Fig. 3.10C & D).

Using these new data, we recalculated the estimated stem cell frequency and obtained values of 1 per 52.3 $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ cells in $C\text{-}Kit^{Mutant}$ mice with the 95% lower and upper confidence intervals estimated at 99.2 and 27.5 cells, respectively. In NSG-S, the new estimated stem cell frequency was significantly poorer, estimated at 1 per 242 cells, with the 95% lower and higher confidence intervals estimated at 530 and 111, respectively.

Cell Dose	***Engrafted/Tested***	
	$C\text{-}Kit^{Mutant}$	NSG-S
50	6/8	3/4
100	6/6	1/4
200	2/3	2/3
400	*	2/4
Estimated Frequency	1/52.3	1/242
Upper Limit	1/99.2	1/530
Lower Limit	1/27.5	1/111

Table 3.4. Breakdown of injected cell doses of $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ injected in C-KitMutant and NSG-S mice, and corresponding numbers of engrafted and tested mice.

Our LDA analysis demonstrated that by selecting the high-expressing CD117 cells within the $CD34^-/CD38^-/CD93^+$ population we achieved positive multilineage human engraftment in the $C\text{-}Kit^{Mutant}$ and NSG-S mouse stains with cell doses as small as 50. This reiterated the potency of CD117 as a positive selection marker for $CD34^-$ SRCs. By applying positive selection for this new marker and refining our gating strategy we have achieved over 100-fold improvement on the purity of the population.

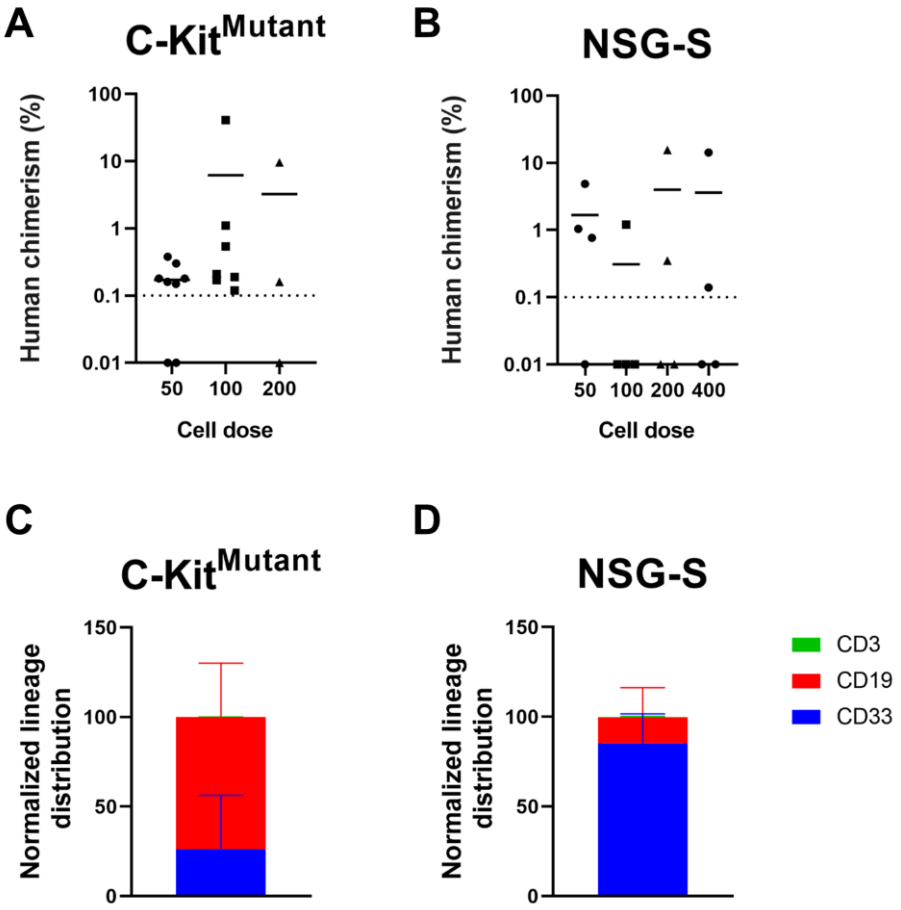

Figure 3.10. Selection for high-expressing CD117 cells within the CD34$^-$/CD38$^-$/CD93$^+$ population purifies further for SRCs

A-B. Quantification of human chimerism as percentage values of all single and live cells in total bone marrow harvested 24 weeks post transplantation of limiting doses of CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi} cells in C-Kit^{Mutant} (A) and irradiation-conditioned NSG-S (B) mice.
C-D. Normalized lineage distribution of all engrafted mice from A-B (C-Kit^{Mutant}, n=15; NSG-S, n=8).

3.8 CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells possess long-term repopulation capacity

To test the long-term repopulation capacity of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^{+/Hi}$ cells we undertook serial transplantations in C-Kit^{Mutant} mice. We started with this strain as we had observed robust engraftment 24 weeks post-transplantation of both

$CD34^-/CD38^-/CD93^+/CD117^+$ and $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ cells. 24 weeks following transplantation of 1, 000 $CD34^-/CD38^-/CD93^+/CD117^+$ or 200-500 $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ cells, we sacrificed engrafted mice and evaluated human engraftment, differentiation outputs and haematopoietic stem cell profile. Interestingly, after gating for all Live, single, $hCD45^+$, Lin^- cells, we observed that the CD34, CD38, CD93 cell surface marker phenotype of the human grafts at 24 weeks resembled the primary $CD34^-/CD38^-/CD93^+/CD117^+$ population (Fig. 3.11B). The grafts were depleted from mouse cells, $hCD45^+$ content was quantified, and the cells were transplanted into secondary $C\text{-}Kit^{Mutant}$ recipients. We checked engraftment via BM aspiration at 6 weeks and sacrificed animals 12 weeks post-transplantation. Engraftment and lineage distribution were quantified using flow cytometry.

Twelve weeks post-transplantation of $hCD45^+$ cells, we observed serial repopulation in all secondary $C\text{-}Kit^{Mutant}$ recipients that received a $CD34^-/CD38^-/CD93^+/CD117^+$ transplant (Fig. 3.11C). In addition, we detected multi-lineage engraftment at both 6 and 12 weeks (Fig. 3.11D). Notably, at 6 weeks, the lineage distribution profile resembled our observations in primary recipients with $CD19^+$ B cells being the most abundant mature cell type (Fig. 3.11D, Fig. 3.6). At 12 weeks, however, there was a significant lineage shift toward $CD33^+$ myeloid cells (Fig. 3.11D). While lineage bias toward the production of myeloid cells could be suggestive of exhaustion of the repopulating HSC pool, the analysis of the CD34 and CD38 cell surface marker phenotypes of the secondary grafts at the end of the secondary transplant still highly resembled the CD34 versus CD38 profiles observed in the primary transplants (Fig. 3.11B & E). In contrast, tangible differences in the CD93 and CD117 cell surface marker profiles could be seen between the primary and secondary transplants, with a visible depletion of the $CD93^+$ and $CD117^+$ fractions in the latter, supporting the possibility of stem cell pool exhaustion (Fig. 3.11B & E).

We conducted the same experiments to test long-term repopulation capacity of $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ cells. Taking into consideration estimated SRC frequencies of about 1 per 52 cells in $C\text{-}Kit^{Mutant}$ mice, we speculated that injecting cell doses of 200 (approximately 4 SRCs) ought to generate human engraft as this has been previously achieved with injecting equivalent or smaller SRC doses of both $CD34^-/CD38^-/CD93^+/CD117^+$ and just $CD34^-/CD38^-/CD93^+$ (Anjos-Afonso et al., 2013). While all primary recipients injected with, 200, 400 or 500 $CD34^-/CD38^-$

/CD93^{+}/CD117^{Hi} were repopulated with multilineage engraftment at 24 weeks, we observed engraftment in 1/3, 2/3, and 1/3 secondary recipients that received grafts generated from 500, 400, and 200 CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} cells, respectively (Table 3.5, Fig. 3.11C). Due to the lower primary cell doses, final engraftment in primary C-Kit^{Mutant} recipients ranged between 4.8 and 10.6%, which was significantly lower than the engraftment levels of achieved by transplantation of 1, 000 CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} cells (Fig. 3.11A). However, at 24 weeks the HSC profile of mice repopulated with CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} resembled that of C-Kit^{Mutat} mice repopulated with CD117^{+}, with similar distribution of CD34 versus CD38 expression, and presence of CD93^{+}/CD117^{+} cells within the CD34^{-}/CD38^{-} fraction (Fig. 3.11B). Despite the low level of engraftment at 12 weeks in CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} repopulated secondary C-Kit^{Mutant} recipients, we could still observe a CD34 versus CD38 expression profile, resembling that in secondary recipients transplanted with CD117^{+} cells (Fig. 3.11E). In contrast to the CD117^{+} transplanted secondary recipients, secondary C-Kit^{Mutant} recipients of CD117^{Hi} cells seemed to lose CD93 and CD117 expression completely (Fig. 3.11E). However, the secondary engraftment of CD117^{Hi} was low in the majority of the repopulated mice (between 0.4-1.6%), making a comprehensive analysis of the HSC phenotype challenging in these secondary recipients (Fig. 3.11C). More experiments and replicates with higher cell doses will be required to properly evaluate the cell surface marker phenotype changes over the course of CD34^{-}/CD38^{-}/CD93^{+}/CD117$^{+/Hi}$ repopulation in C-Kit^{Mutant} mice.

Population	**1° Cell Dose**	**2° Cell Dose (hCD45^{+})**	**Donor to Recipient Graft Ratio**	**Engrafted/ Tested**
CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+}	1,000	14x10^{6}	1:1	3/3
CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}	500	200x10^{3}	1:3	1/3
	400	200x10^{3}	1:1	1/1
	400	200x10^{3}	1:2	1/2
	200	125x10^{3}	1:3	1/3

Table 3.5. Overview of tested cell doses in primary and secondary C-Kit^{Mutant} recipients in five independent experiments.

Primary cell dose represents the number of $CD34^-/CD38^-/CD93^+/CD117^+$ cells injected intravenously per each primary recipient. Secondary cell dose represents the number of human $CD45^+$ cells injected per secondary recipient. Donor to recipient ratio represents the number of primary recipients to the number of secondary recipients

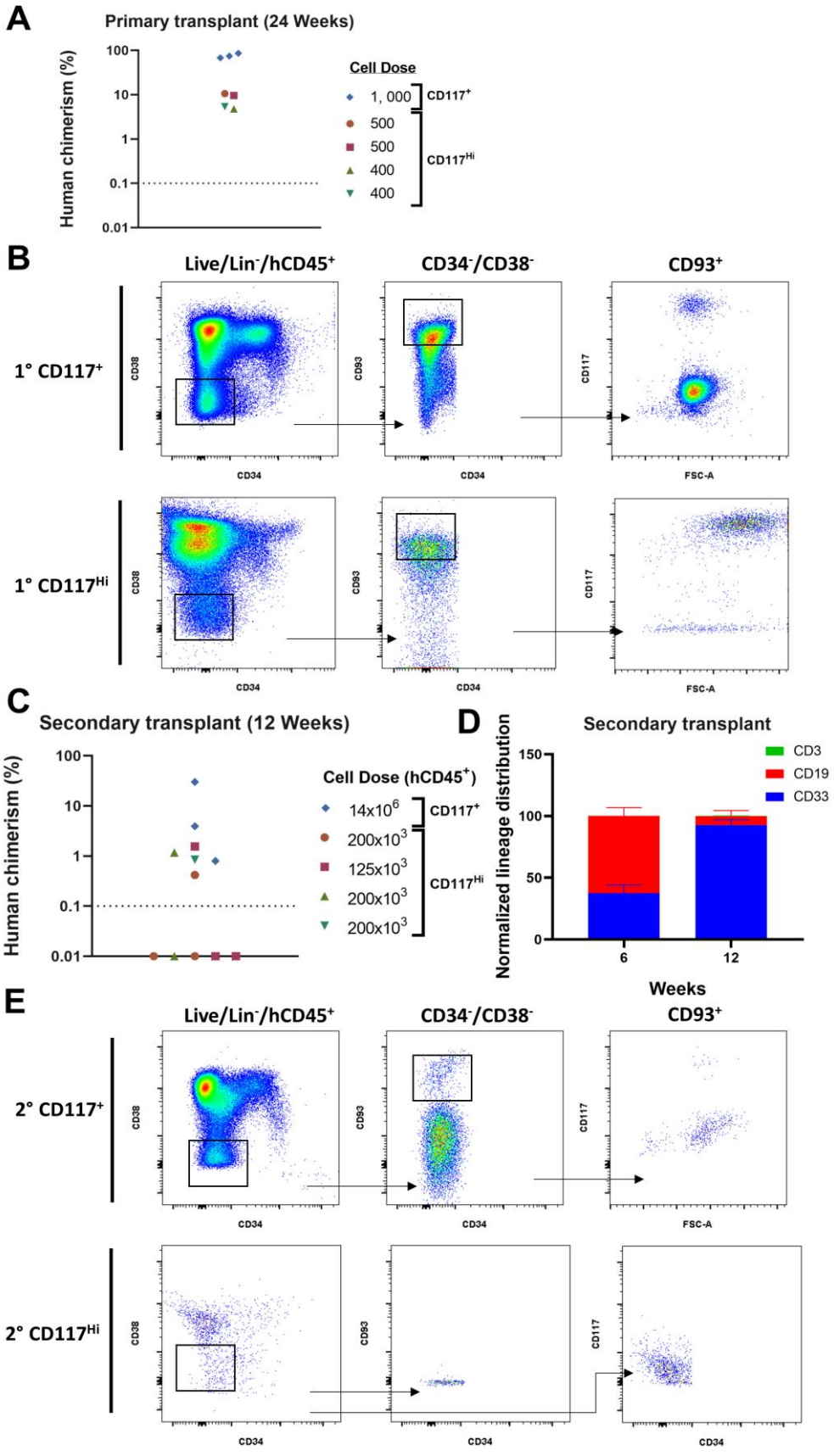
A
Primary transplant (24 Weeks)
Human chimerism (%)
100
10
1
0.1
0.01
Cell Dose
1, 000
500
500
400
400
CD117+
CD117Hi
B
Live/Lin-/hCD45+
CD34-/CD38-
CD93+
1° CD117+
1° CD117Hi
CD38
CD34
CD93
CD117
FSC-A
C
Secondary transplant (12 Weeks)
Human chimerism (%)
100
10
1
0.1
0.01
Cell Dose (hCD45+)
14x10^6
200x10^3
125x10^3
200x10^3
200x10^3
CD117+
CD117Hi
D
Secondary transplant
Normalized lineage distribution
150
100
50
0
6
12
Weeks
CD3
CD19
CD33
E
Live/Lin-/hCD45+
CD34-/CD38-
CD93+
2° CD117+
2° CD117Hi
CD38
CD34
CD93
CD117
FSC-A

Figure 3.11. CD34^{-}/CD38^{-}/CD93^{+}/CD117$^{+/Hi}$ cells have long-term *in vivo* repopulation potential

A. Quantification of total human engraftment as a percentage value of all single live cells in harvested bone marrow of primary C-Kit^{Mutant} recipients. Each data point represents a mouse. Each set of data points represents and independent experiment from Table 4 (diamonds-Row 1; circles-Row 2; inverted triangle-Row 3, triangles-Row 4; squares-Row 5).
B. Representative flow cytometry plots of primary human grafts in C-Kit^{Mutant} 24 weeks post transplantation of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+}.
C. Quantification of total human engraftment as a percentage value of all single live cells in harvested bone marrow of secondary C-Kit^{Mutant} recipients. Each data point represents a mouse. Each set of data points represents and independent experiment from Table 2 (diamonds-Row 1; circles-Row 2; inverted triangle-Row 3, triangles-Row 4; squares-Row 5).
D. Normalized lineage distribution at 6 and 12 weeks post-injection in secondary C-Kit^{Mutant} recipients (Data represents 3 mice; Table 2 – Row 1)
E. Representative flow cytometry plots of secondary human grafts in C-Kit^{Mutant} 12 weeks post transplantation of hCD45^{+} cells from CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} primary grafts.

Next, we undertook serial transplantations in NSG-S mice. We injected primary recipients with 1, 000 and 500 cells in two independent experiments. The engraftment of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} is lower in NSG-S compared to C-Kit^{Mutant} mice at 24 weeks post injection (Fig. 3.11A, Fig. 3.12B). Due to the low engraftment the CD45^{+} cell yield was much lower than what we obtained from C-Kit^{Mutant} mice, but we proceeded with secondary transplantations in a 1:1 primary to secondary recipient ratio (Table 3.6). No human engraftment was detected in secondary NSG-S recipients in serial transplantation of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} cells (Fig. 3.12C, Table 3.6). This could be attributed to the lower doses of hCD45^{+} cells that secondary NSG-S recipients were transplanted with compared to secondary C-Kit^{Mutant} recipients. Another explanation for the failure of the secondary transplantation of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} in NSG-S could be that the overexpression of key human cytokines in this mouse strain pushes the self-renewal of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} and depletes the population of long-term repopulating cells quicker than in C-Kit^{Mutant} mice. Analysis of the HSC profile of primary CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+} grafts at 24 weeks post transplantation showed that the CD34 versus CD38 expression phenotype had shifted predominantly toward CD34^{-} populations, which contain mostly mature cells (Fig. 3.12A). In addition, a tangible shift of CD117 expression could be observed (Fig. 3.12A). Almost no

$CD117^-$ cells were present, while all positive CD117 cells were distributed between $CD117^{Very\ High}$ and $CD117^{High}$ populations (Fig 3.12A). Prior to transplantation, NSG-S mice require irradiation conditioning, which can result in damage to the bone marrow niche, adding an additional challenge for secondary $CD34^-/CD38^-/CD93^+/CD117^+$ transplants to engraft. Due to the low engraftment of $CD34^-/CD38^-/CD93^+/CD117^+$ cells in NSG mice, we did not attempt serial transplantations in this strain.

In summary, we observed that serial transplantation of approximately 1, 000 $CD34^-/CD38^-/CD93^+/CD117^+$ yields robust multi-lineage secondary engraftment in C-Kit^{Mutant} mice. Lower primary doses between 200-500 $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ cells yielded secondary engraftment in 4 out of 9 secondary C-Kit^{Mutant} recipients. In this strain, we successfully achieved secondary engraftment with a primary dose of as little as 200 $CD34^-/CD38^-/CD93^+/CD117^{Hi}$. In NSG-S mice, we detected no secondary engraftment. Possible reasons for this could be that the strain is exhausting the SRC pool in the primary recipients, the low primary engraftment at 24 weeks, and unfavourable niche conditions for $CD34^-$ SRCs in secondary recipients. Our findings strongly suggest that the C-Kit^{Mutant} strain maintains $CD34^-$ SRC long-term and could be superior to NSG-S in studying the functionality of this population.

Population	**1° Cell Dose**	**2° Cell Dose ($hCD45^+$)**	**Donor to Recipient Graft Ratio**	**Engrafted/ Tested**
$CD34^-/CD38^-$ $/CD93^+/CD117^+$	1,000	$150x10^3$	1:1	0/3
	500	$90x10^3$	1:1	0/3

Table 3.6. Overview of tested cell doses in primary and secondary NSG-S recipients in two independent experiments.

Primary cell dose represents the number of $CD34^-/CD38^-/CD93^+/CD117^+$ cells injected intravenously per each primary recipient. Secondary cell dose represents the number of human $CD45^+$ cells injected per secondary recipient. Donor to recipient ratio represents the number of primary recipients to the number of secondary recipients

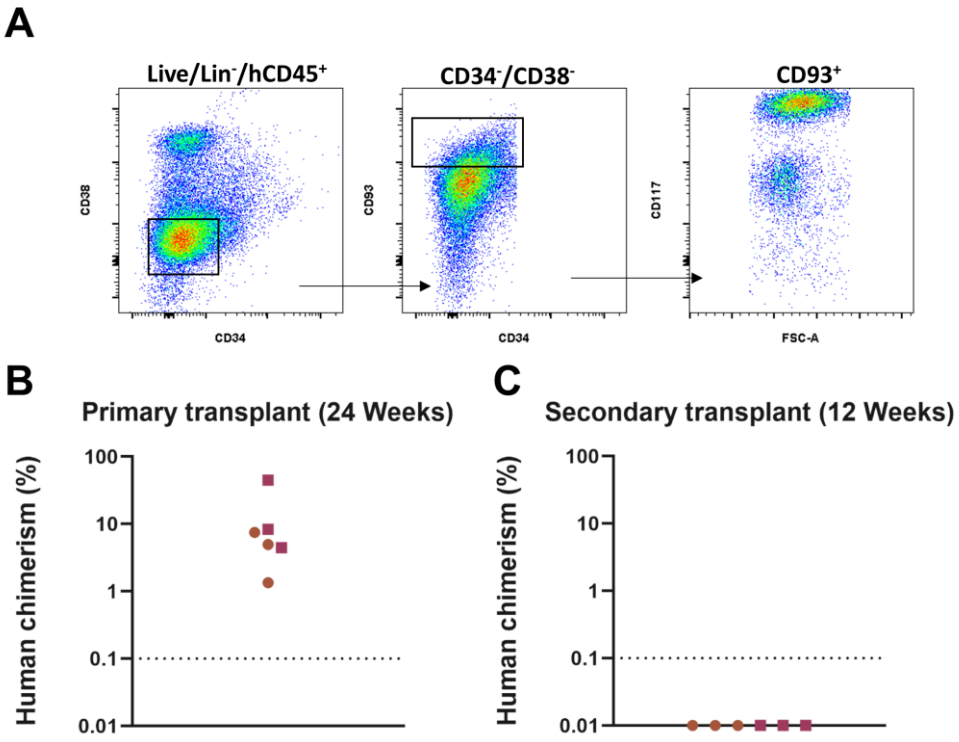

Figure 3.12. The CD34-/CD38-/CD93+/CD117+ do not repopulate secondary NSG-S recipients.

A. Representative flow cytometry plot of primary grafts 24 weeks post transplantation of $CD34^-/CD38^-/CD93^+/CD117^+$ cells in NSG-S mice.
B. Quantification of total human engraftment as a percentage value of all single live cells in harvested bone marrow of primary NSG-S recipients. Each data point represents a mouse. Each set of data points represents and independent experiment from Table 3 (squares-Row 1; circles-Row 2).
C. Quantification of total human engraftment as a percentage value of all single live cells in harvested bone marrow of secondary NSG-S recipients. Each data point represents a mouse. Each set of data points represents and independent experiment from Table 3 (squares-Row 1; circles-Row 2).

3.9 $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs produce functional $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+$ with *in vivo* repopulation potential in C-Kit^{Mutant} mice

After establishing that the C-Kit^{Mutant} model is more suited than NSG-S for the maintenance of the long-term repopulation capacity of $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs, we questioned whether primary transplants of this population can produce

functional $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs, capable of generating secondary multi-lineage engraftment. To investigate this, we injected two primary recipients with approximately 500 $CD34^-/CD38^-/CD93^+/CD117^+$ cells and sacrificed at 24 weeks. Following depletion of murine cells, we stained the human primary grafts and sorted for $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+/CD117^+$ cells (Fig. 3.13A & B). We injected three secondary mouse recipients with 2, 000 $CD34^+/CD38^-$ and another three with 1, 200 $CD34^-/CD38^-/CD93^+/CD117^+$ cells and monitored engraftment over a period of 24 weeks. In our first experiment we observed consistent engraftment in a single mouse injected with $CD34^+/CD38^-$ (Fig. 3.13C). At 24 weeks about 80% of these cells were comprised of $CD33^+$ myeloid cells and approximately 20% of $CD19^+$ B cells (data not shown). Marginal human engraftment was detected in a single mouse injected with $CD34^-/CD38^-/CD93^+/CD117^+$ at 18 weeks post injection, but engraftment was too low to assess multi-lineage output (Fig. 3.13C). No splenic engraftment was detected in any of the mice at 24 weeks.

Following these results, we speculated that our secondary transplant gating strategy is eliminating SRCs from the $CD34^-$ fraction by placing the gate for $CD117^+$ cells too high (Fig. 3.13B). We observed that CD117 expression by $CD34^-/CD38^-/CD93^+$ cells 24 weeks post transplantation in C-Kit^{Mutant} mice is higher compared to that observed in $CD34^-/CD38^-/CD93^+/CD117^+$ that were initially transplanted (Fig. 3.13B). In section 4 of this chapter, we demonstrated that the $CD34^-/CD38^-/CD93^-$ fraction contained a population expressing high levels of CD117 that held no *in vivo* repopulation potential. This could suggest that the $CD34^-/CD38^-/CD93^+$ population with elevated CD117 expression does not represent SRCs. It is possible that the repopulating $CD34^-/CD38^-/CD93^+/CD117^+$ downregulate CD117 expression after extended occupancy in the murine bone marrow niche and by selecting for high-expressing CD117 cells we could have been potentially excluding the $CD34^-$ SRCs. Therefore, we repeated the experiment by sorting from the primary $CD34^-/CD38^-/CD93^+/CD117^+$ grafts for $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+$ cells without selecting for $CD117^+$ in the latter (Fig. 3.13D).

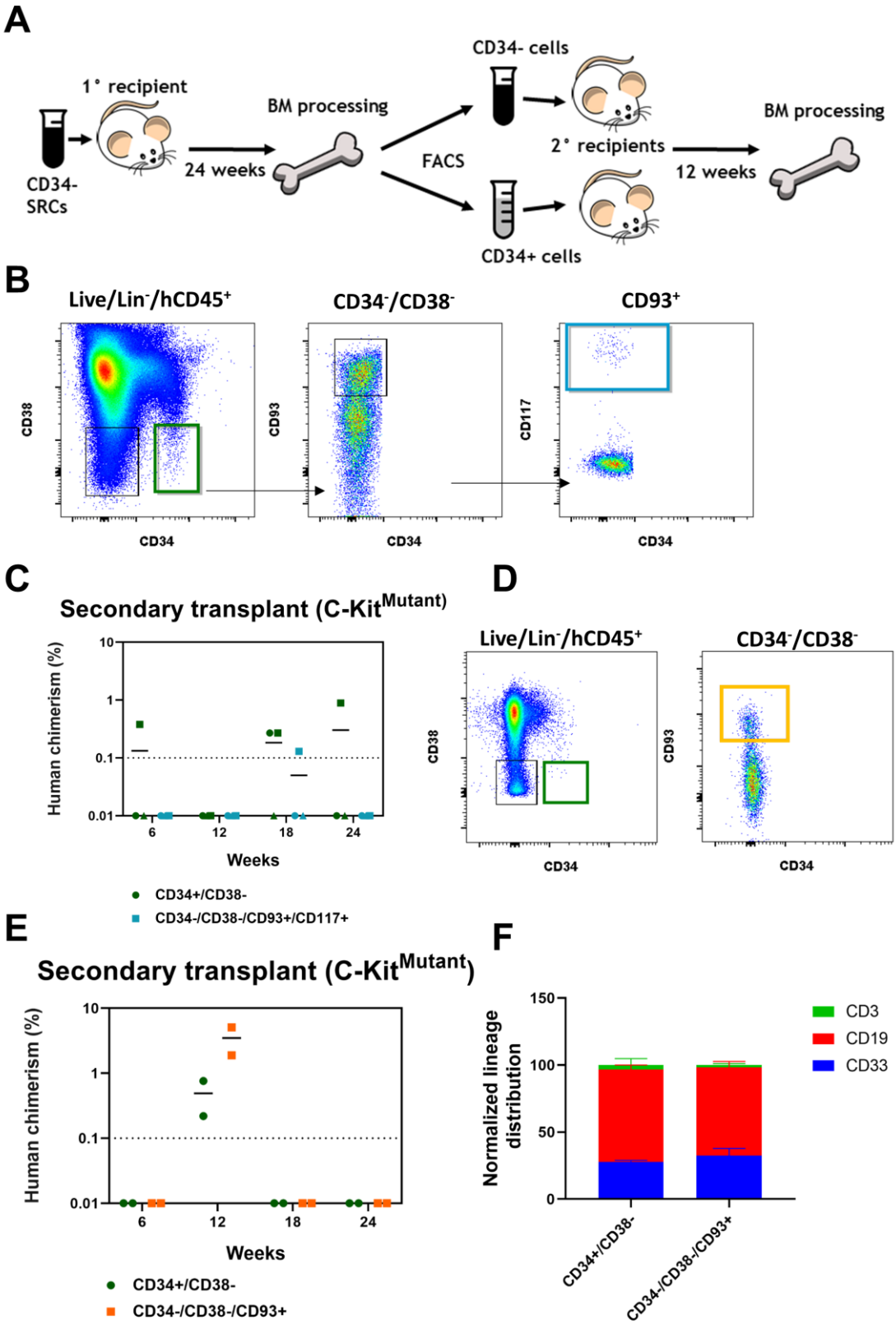

Figure 3.13. Primary CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ SRCs produce repopulating CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ and CD34$^+$/CD38$^-$

A. Schematic of serial transplantation of $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+/CD117^+$ or $CD34^-/CD38^-/CD93^+$ cells obtained from primary human grafts generated by $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs.
B. Representative flow cytometry plots of gating strategy for sorting $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+/CD117^+$ cells from primary human grafts obtained from C-Kit^{Mutant} mice 24 weeks post transplantation with primary $CD34^-/CD38^-/CD93^+/CD117^+$
C. Quantification of human engraftment of $CD34^+/CD38^-$ (n=3) and $CD34^-/CD38^-/CD93^+/CD117^+$ (n=3) as a percentage value of all live single cells from bone aspirates (weeks 6-18) and processed bone marrow from legs and hips (24 weeks). Point shapes distinguish between individual mice.
D. Representative flow cytometry plots of gating strategy for sorting $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+/CD117^+$ cells from primary human grafts
E. Quantification of human engraftment of $CD34^+/CD38^-$ (n=2) and $CD34^-/CD38^-/CD93^+$ (n=2) as a percentage value of all live single cells from bone aspirates (weeks 6-18) and processed bone marrow from legs and hips (24 weeks).

In this experiment, we injected mice with 2, 000 $CD34^+/CD38^-$ and 2, 200 $CD34^-/CD38^-/CD93^+$ cells obtained from two primary mouse recipients injected with 500 $CD34^-/CD38^-/CD93^+/CD117^+$ cells and tracked human engraftment over 24 weeks. At 12 weeks post-injection, we detected human engraftment of over 1% in two out of two mice injected with $CD34^-/CD38^-/CD93^+$ (Fig. 3.13E). We also observed engraftment in the two mice injected with $CD34^+/CD38^-$ (Fig 3.13E). Most notably, at 12 weeks we also detected multilineage output which was similar between the two cohorts of mice and mimicked the lineage output typically observed in primary C-Kit^{Mutant} recipients transplanted with $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs. Unfortunately, the engraftment didn't persist until 24 weeks. We also didn't detect any splenic engraftment at the end of the experiment.

In these experiments we attempted to investigate the capacity of $CD34^-$ SRCs to produce functional $CD34^-$ and $CD34^+$ repopulating cells. In two independent experiments, we detected secondary engraftment of $CD34^+/CD38^-$ obtained from $CD34^-/CD38^-/CD93^+/CD117^+$ primary transplants. In the first experiment, one mouse injected with $CD34^+/CD38^-$ sorted from $CD34^-/CD38^-/CD93^+/CD17^+$ primary xenografts had long-term engraftment that persisted until 24 weeks in the secondary recipient. In the second experiment, we detected multi-lineage engraftment at 12 weeks in two different mice. In this experiment, we also detected secondary engraftment of $CD34^-/CD38^-/CD93^+$ obtained from $CD34^-/CD38^-/CD93^+/CD117^+$ primary transplants with multi-lineage engraftment in two out of two tested mice.

It has been previously demonstrated in NSG that the $CD34^-/CD38^-/CD93^+$ population can produce repopulating $CD34^+$ and $CD34^-$ SRCs, however these experiments were only 12 weeks long in the primary transplant and cells were sorted only based on CD34 expression (Anjos-Afonso et al., 2013). Here we have pushed the experiment by extending both the primary and secondary repopulation time to 24 weeks. While we detected multi-lineage secondary engraftment by both $CD34^+$ and $CD34^-$ grafts, the repopulating capacity of the injected cells appeared to diminish by the 48-week time point. This could be due to the fact that the SRC pool is exhausted due to the limited number of $CD34^-/CD38^-/CD93^+/CD117^+$ injected in the primary recipients or the possibility that the mouse model is pushing the population to exhaust. To answer these questions, the experiments would have to be repeated with a larger number of primary cells transplanted and different mouse models. In addition, the mouse bone marrow microenvironment could affect the expression of cell surface markers of human SRCs, rendering their isolation based on highly specific cell surface phenotypes difficult. While these results indicate that $CD34^-/CD38^-/CD93^+/CD117^+$ SRCs can produce repopulating $CD34^+/CD38^-$ and $CD34^-/CD38^-/CD93^+$ SRCs, further experiments are required with more replicates and larger primary cell doses to firmly establish the long-term repopulation capacity and *in vivo* population dynamics.

3.10. Chapter summary and discussion

In the beginning of this chapter, we investigated the expression of multiple cell surface markers by $CD34^-/CD38^-/CD93^+$ cells in attempt to identify new purification marker for $CD34^-$ SRCs. We showed that we could not reproduce the antibody staining for two recently reported markers (CD133 and GPI-80) that have been shown to improve the $CD34^-$ SRC frequency to approximately 1/8.1 cells. Their report indicated their $CD34^-$ SRC population differed from the one investigated in this thesis; they did not observe any CD93 expression in their population. That prompted us to investigate the repopulation capacity of $CD34^-/CD38^-/CD93^-/CD117^{+/-}$ cells. We did not observe human engraftment in mice injected with these population, suggesting

that the CD93$^-$ populations are a potentially poor source of SRCs. More experiments with larger cell doses need to be undertaken to confirm this.

We identified CD117 as good CD34$^-$ SRC positive selection marker and demonstrated that by adding it to our gating strategy, we could achieve human engraftment in three different immunodeficient mouse strains. Furthermore, we observed a distinct repopulation kinetics and multilineage output between NSG, NSG-S and C-Kit^{Mutant} mice transplanted with the same numbers CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells. In NSG we observed very low engraftment, while in NSG-S and C-Kit^{Mutant} mice, it was significantly higher. This was a major step forward for this project as the low engraftment of CD34$^-$ SRC in the traditionally-used NSG mouse strain has made the study of this population challenging. Interestingly, in NSG-S, we observed engraftment as early as 6 weeks post-transplantation, when positive engraftment of CD34$^-$ SRCs usually is detected from 12 weeks post-transplantation. It is easy to speculate that the overexpression of three human cytokines in the NSG-S mouse strain could be pushing the earlier engraftment of CD34$^-$ SRCs. We did not investigate the reason behind this, but cell cycle status analysis could reveal whether human grafts generated from CD34$^-$ SRCs contain more proliferative cells at 6 weeks in NSG-S mice compared to 12 weeks in NSG mice. NSG-S mice express human SCF, IL-3 and GM-CSF. CD117 is the receptor for SCF, so we can expect that once homed to the bone marrow of NSG-S mice CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells will be subjected to constitutive KIT signalling activation. Gain of function mutations associated with this pathway have been reported to result in expansion of the myeloid compartment, which is consistent with our observation that CD33$^+$ myeloid cells are more abundant in NSG-S compared to NSG and C-Kit^{Mutant} mice (Billerbeck et al., 2011; Willinger et al., 2011). The deregulation of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ HSCs by the overexpression SCF in NSG-S mice could also potentially explain why we could not achieve successful secondary engraftment in this strain. It is possible that the overexpression of IL-3 and GM-CSF also play a role in stimulating the early engraftment of CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$, but the expression of the corresponding receptors on the surface of these cells is unknown.

From our repopulation kinetics analysis for CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$ cells, we established that these SRCs expand the most within the C-Kit^{Mutant} strain, while retaining their long-term self-renewal capacity as demonstrated by their ability to repopulate secondary mouse recipients. In NSG-S we could achieve quicker

repopulation readouts, as early as 6 weeks post-injection, but failed to achieve engraftment in secondary NSG-S recipients. This suggests that the NSG-S mouse strain could be useful in identifying potential SRC candidate populations quickly, but may exhaust them faster as well. We proceeded by refining our gating strategy for this population and demonstrating that the majority of SRCs are located within the $CD117^{High}$ fraction of the $CD34^{-}/CD38^{-}/CD93^{+}$ population. We confirmed this though *in vivo* repopulation assay that revealed the multilineage engraftment pattern of the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ population that was consistent with our earlier data on the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+}$ population, suggesting that it is the high-expressing CD117 cells that drive the human engraftment in the NSG, NSG-S and C-Kit^{Mutant} strains. To finalize this work, we performed limiting dilution analysis for the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ population in NSG-S and C-Kit^{Mutant} strains, which resulted in an estimated stem cell frequency of 1/242 and 1/52.3, respectively. It is highly likely that technical issues may have affected the outcomes of the LDA analysis, but strain-specific differences between NSG-S and C-Kit^{Mutant} mice may also contribute to the different estimated stem cell frequencies. Nonetheless, by identifying CD117 as an effective positive selection marker for $CD34^{-}$ SRCs, we have finally brought the purity of this population at a level that allows us to investigate its transcriptome at a single-cell level.

Chapter 4. Single-cell transcriptomics reveals the heterogeneity of the CD34^{-} SRCs in post-natal haematopoiesis

4.1 Introduction

Since the early 2000s, the haematopoietic stem cell field has been driven forward by rapid advancements in the understanding of the structure of the haematopoietic hierarchy. The discovery and introduction of new surface markers helped redraw the haematopoietic tree, showing that the megakaryocyte lineage branches off early in the hierarchy, while the myeloid and lymphoid lineages do not dissociate from each other until later in the maturation process (Adolfsson et al., 2005; Doulatov et al., 2010; Sanjuan-Pla et al., 2013; Yamamoto et al., 2013). Studies in the mouse were also strongly suggesting that both the MPP and HSC pools are highly heterogenous (Cabezas-Wallscheid et al., 2014; Dykstra et al., 2007; Pietras et al., 2015; Sanjuan-Pla et al., 2013). These reports were indicating that the compartmentalized view of the haematopoietic hierarchy is too rigid to accurately represent the HSPC population and the dynamic processes of differentiation. Functionality assays based on *in vitro* and *in vivo* methods were also suggesting that HSC primed for a particular lineage *in vitro* could change their lineage output *in vivo* and vice versa (Carrelha et al., 2018; Rodriguez-Fraticelli et al., 2018). The implications of these findings were multi-faceted, but one strong inference was that lineage priming occurs at the level of single HSCs. But xenotrasnplantations and *in vitro* assays are likely to provide suboptimal conditions to test the differentiation and self-renewal of every single HSC. These assays are also unable to provide comprehensive differentiation readouts. Furthermore, in order to gain a better understanding of how HSC pools are maintained and cell fate determined, we need not only information about biological functionality but the molecular profiles of single cells.

Initially, single-cell gene expression assays were limited to a small number of cells and a predetermined panel of genes, restricting the scope of new findings (Vogelstein et al., 2006). However, advancements in single-cell RNA-sequencing

(scRNAseq) technologies resulted in rapid reporting of the transcriptomes of murine HSPCs. Human studies followed soon after, mirroring what had been previously reported in the mouse system; the human HSC pool is highly heterogenous and individual HSCs co-expressed transcriptional programmes associated with multiple committed lineages, (Pellin et al., 2019; Velten et al., 2017; Zheng et al., 2018). Interestingly, despite evidence that oligopotent progenitors appear to be more enriched in umbilical cord blood compared to adult bone marrow, comparison of transcriptomic data between adult and post-natal $CD34^+$ HSCs suggested that early lineage priming processes are conserved between bone marrow and cord blood (Zheng et al., 2018).

Despite the growing number of reports on scRNAseq in human haematopoietic cells, very little has been done to investigate the molecular profile of the $CD34^-$ stem cells. In most studies, $CD34^-$ cells are either depleted immuno-magnetically or discarded from the scRNAseq analysis. To our knowledge, there is only one recent report on $CD34^-$ SRCs (defined as $CD34^-/CD133^+/GPI\text{-}80^+$) (Sumide et al., 2018), using microarrays to assay the expression of 79 target genes in 23 $CD34^-/CD133^+/GPI\text{-}80^+$ cells ($CD34^-$ SRCs) and 33 $CD34^+/CD38^-/CD45RA^-/CD90^+$ cells ($CD34^+$ SRCs) (Sumide et al., 2018). Comparison between the two populations showed similar patterns of expression for genes like KIT, RUNX1, TAL1, BMI1, DNMT3A, TGFBR1, TGFBR2 (Sumide et al., 2018). While the expression of most genes seemed to correlate between the two populations, IFTM1, MPL, IKZF1 appeared to be expressed at lower levels by $CD34^-$ SRCs (Sumide et al., 2018). Only two genes were reported to be more highly expressed in $CD34^-$ SRCs, MYB and EZH2 (Sumide et al., 2018). Their 1/8.1 SRC frequency implies that approximately 2.8 of their 23 single $CD34^-/CD133^+/GPI\text{-}80^+$ cells are actual SRCs. Interestingly, around 8 of their $CD34^-$ single cells clustered closely to the $CD34^+$ SRCs in their PCA plot. Unfortunately, the authors did not make the single-cell RT-PCR data available, which would have allowed us to investigate which genes from their panel, these eight cells express.

To build on this, the authors undertook bulk gene expression analysis of sorted $CD34^-/CD133^+/GPI\text{-}80^+$ and $CD34^+/CD38^-/CD45RA^-/CD90^+$ cells, showing that the biological processes and pathways associated with upregulated genes in the former population were related to hypoxia, megakaryocyte/erythroid differentiation, Wnt signalling, and angiogenesis (Sumide et al., 2018). While this data is intriguing and

the CD34$^-$ SRC frequency is remarkably good, it has implications for the validity of their global gene expression analysis; only about 1/8 of the data represents the transcriptome of SRCs. Being technically challenging assays, *in vivo* LDAs tend to underestimate actual SRC numbers in tested populations, but still a significant proportion of the cells analysed are non-SRCs. Drawing meaningful conclusions about the gene expression profile of the bulk CD34$^-$/CD133$^+$/GPI-80$^+$ population is thus challenging.

In order to gain better insight into the molecular profile of impure populations, scRNAseq is a powerful method. Provided enough cells are sequenced at sufficient read depth, HSC signatures can be captured. In Chapter 3, we demonstrated that the CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi} population contains functional SRCs with long-term repopulation capacity and multi-lineage engraftment. We also showed that the SRC frequency can reach approximately 1/100 cells. Following these functional studies, we undertook a scRNAseq of the CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi} population in order to gain further insight into the molecular profile of the elusive CD34$^-$ SRC.

4.2 Characterisation of cell types within the CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi} population

To characterise the transcriptional state of CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi}, we sorted approximately 2000 cells to be processed using the 10X Genomics Single Cell 3' v.3.1 Reagent Kit. 972 cells were sequenced with an average depth of 193,887 reads per cell, reaching approximately 80% sequencing saturation. After processing and mapping of reads, we used Seurat to further analyse the data. As part of quality control checks, we visualized the number of genes per cell (nFeature), number of unique molecular identifiers mapped to the human genome (nCount), and percentage of mitochondrial gene counts per cell (percent.mt) (Fig. 4.1A-C). We filtered out cells with less than 200 genes and over 5000 genes to eliminate low quality cells and doublets. We also eliminated cells with over 10% mitochondrial gene counts as high expression of mitochondrial genes indicates dying cells (Ilicic et al., 2016). This left us with 695 cells for further analysis. Visualization of percent.mt versus nCount indicated that there are still some cells in the dataset with low number of RNA counts and high mitochondrial content relative to the rest of the cells (Fig.

4.1D). Also, when we plotted nCount versus nFeature, we could see that there are a number of cells with both high number of genes and counts per cell, which could be indicative of doublets (Fig. 4.1E). We decided not filter the dataset further, but instead investigate if these potential doublets cluster together or separately as the latter would suggest doublets.

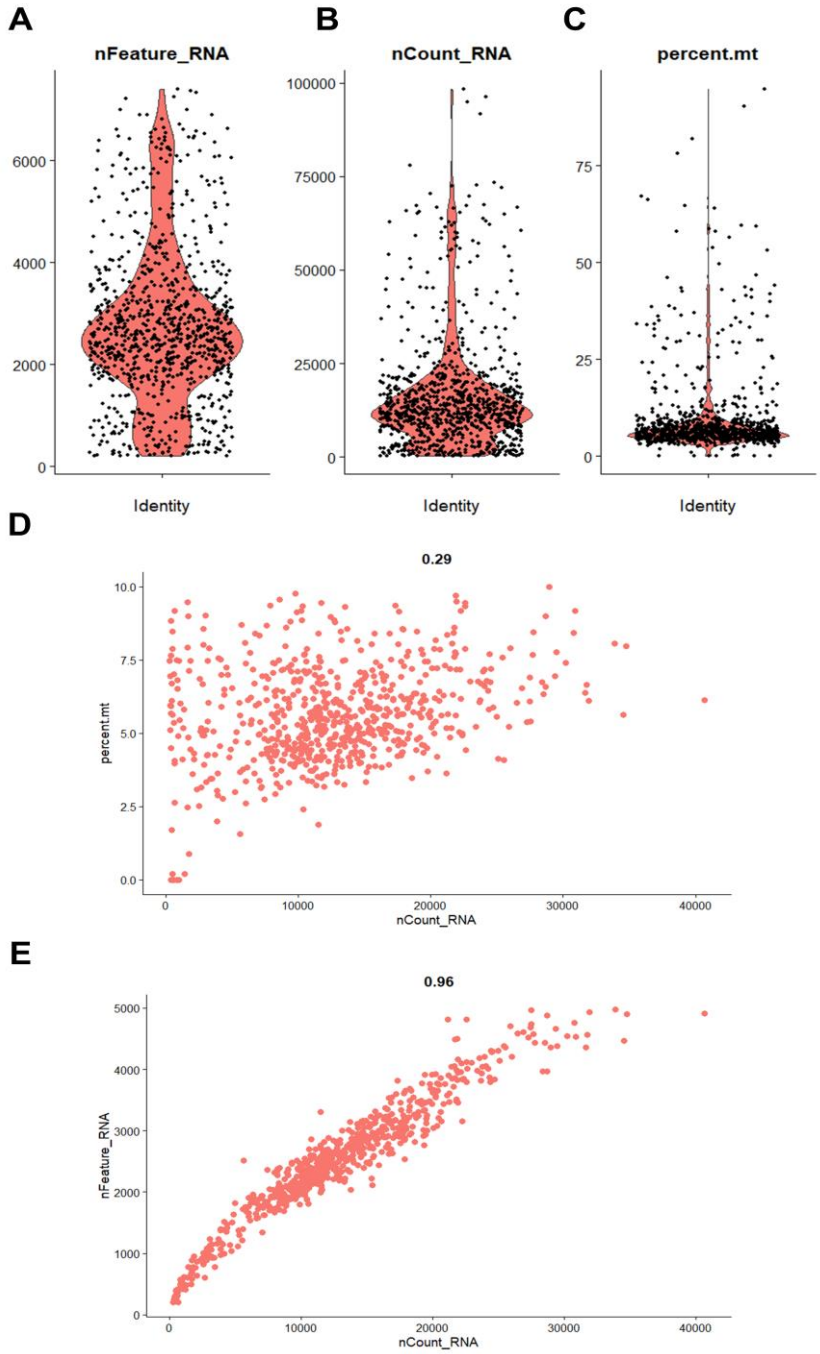

Figure 4.1. Quality control of scRNAseq data for CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}

A-C. Plot of detected genes per cell (A), RNA counts per cell (B), and mitochondrial gene content per cell (C).

D. Scatter plot of RNA counts per cell versus percentage of mitochondrial genes per cell.
E. Scatter plot of count of RNA counts per cell versus gene counts per cell.

We next performed linear dimensional reduction of the data. The principal component analysis (PCA) clustered a large portion of the cells together, suggesting that the bulk of the sample is similar in gene expression compared to a small portion of outlier cells (Fig. 4.2A). We were interested in visualizing the gene sets that primarily contribute to the variance and plotted heatmaps of the first 12 principal components (PCs) that explain the variance of the data (Fig. 4.2B). We observed that clear structure in the PC plots up to the PC_9, suggesting that most of the variance in the data can be explained by the first 9 PCs (Fig. 4.2B). We also used a statistical test (JackStraw) to determine the significance of the different PCs which generates p-values that show how strong the association of each gene is with its principle component. The first five principle components contained genes that are significantly associated with their respective PC. The distribution of p-values from PC_10 grew steadily closer to the p-value uniform distribution (dashed line), which indicates the enrichment for genes with low p-values begins to decrease (Fig. 4.2C). Another approach we took to determine how many PCs to use in the downstream analysis was to perform an Elbow Plot, which plots the standard deviations of each PC (Fig. 4.2D). The “knee” of the curve begins levelling off around cluster 12, which suggests that a high amount of information of the variance of the data is captured in these components (Fig. 4.2D). This is in agreement with the data structure for each PC according to the heatmaps (Fig. 4.2D). As the JackStraw analysis suggested that the first 20 PCs are significant, we proceed with 15 PCs in the downstream analysis.

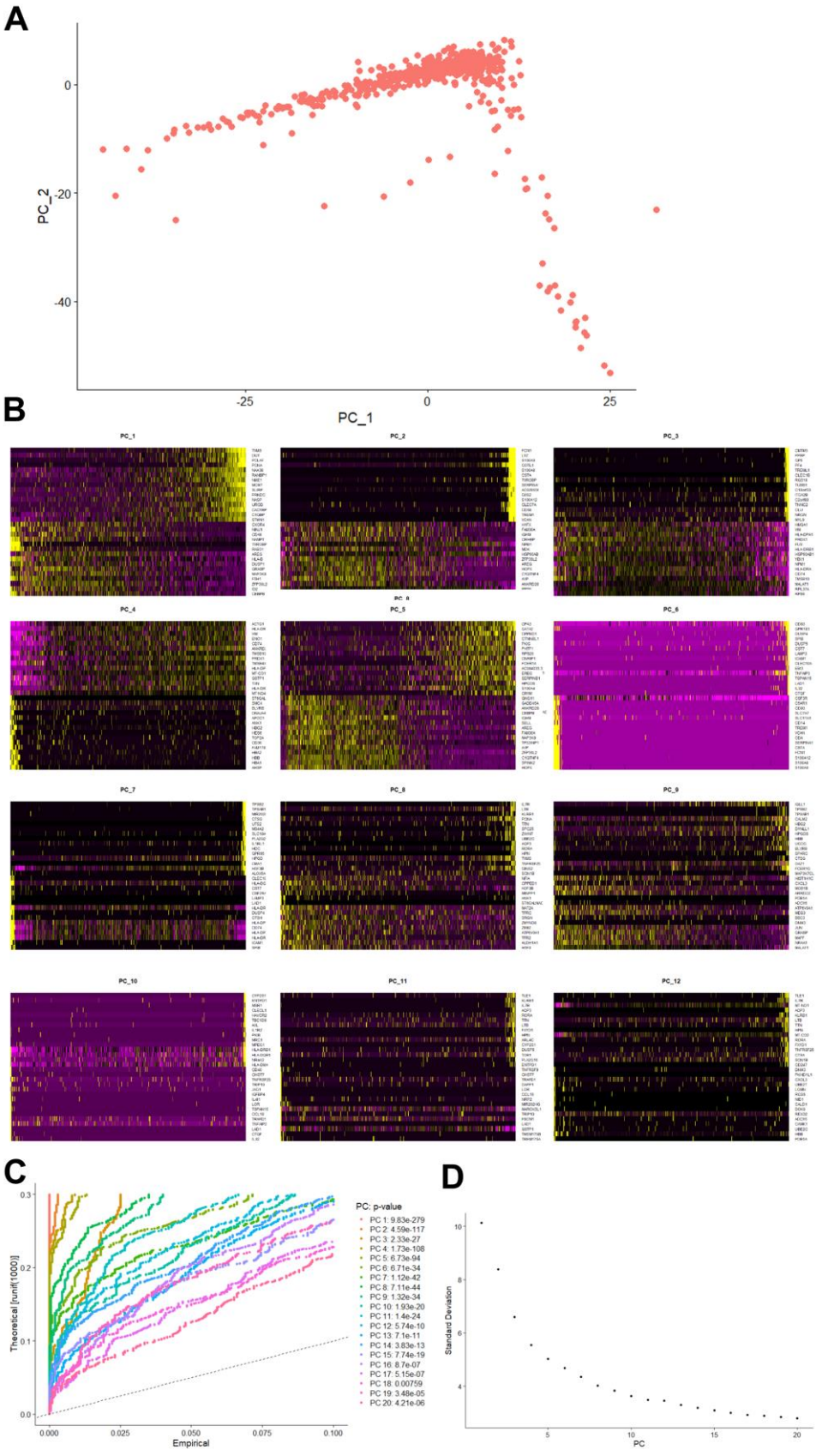
A
PC_2
PC_1
B
C
Theoretical [runif(1000)]
Empirical
PC: p-value
PC 1: 9.83e-279
PC 2: 4.59e-117
PC 3: 2.33e-27
PC 4: 1.73e-108
PC 5: 6.73e-94
PC 6: 6.71e-34
PC 7: 1.12e-42
PC 8: 7.11e-44
PC 9: 1.32e-34
PC 10: 1.93e-20
PC 11: 1.4e-24
PC 12: 5.74e-10
PC 13: 7.1e-11
PC 14: 3.83e-13
PC 15: 7.74e-19
PC 16: 8.7e-07
PC 17: 5.15e-07
PC 18: 0.00759
PC 19: 3.48e-05
PC 20: 4.21e-06
D
Standard Deviation
PC

Figure 4.2. Principal component analysis and determinations of significant principle components

A. Principle component analysis plot.
B. Heatmaps of the first 12 principal components levels of expression on a subset of cells. Yellow and purple colours represent upregulation and downregulation of genes, respectively.
C. Quantile-quantile plot where each curve represents a PC. Every point represents the p-value of a variable feature (gene) plotted against a universal p-value. The dashed line represents the threshold.
D. Elbow plot of the standard deviation of the first 10 PCs.

We grouped cells based on their expression profiles. To do this we used graph-based clustering followed by nonlinear dimensionality reduction (UMAP). Eight clusters were identified with the first five containing the majority of the cells (Fig. 4.3A). We observed that the graph-based clusters were well-separated. By mapping the number of genes per cell (n_Features) versus the number of counts per cell (n_Counts) and projecting the clusters onto the plot, we observed that cluster 5 is predominantly comprised of cells with both high number of genes and counts, which can be indicative of doublets (Fig. 4.3B). We also observed that the percentage of mitochondrial genes seemed evenly distributed between 2.5 and 10% between all clusters, with the exception of clusters 4 and 5, where mitochondrial gene expression was over 6% in a proportion of the cells (Fig. 4.3C). While high expression of mitochondrial genes can suggest that the cells in these clusters are in the process of dying, elevated mitochondrial function and biogenesis can also be associated with the natural processes in some subsets of haematopoietic cells.

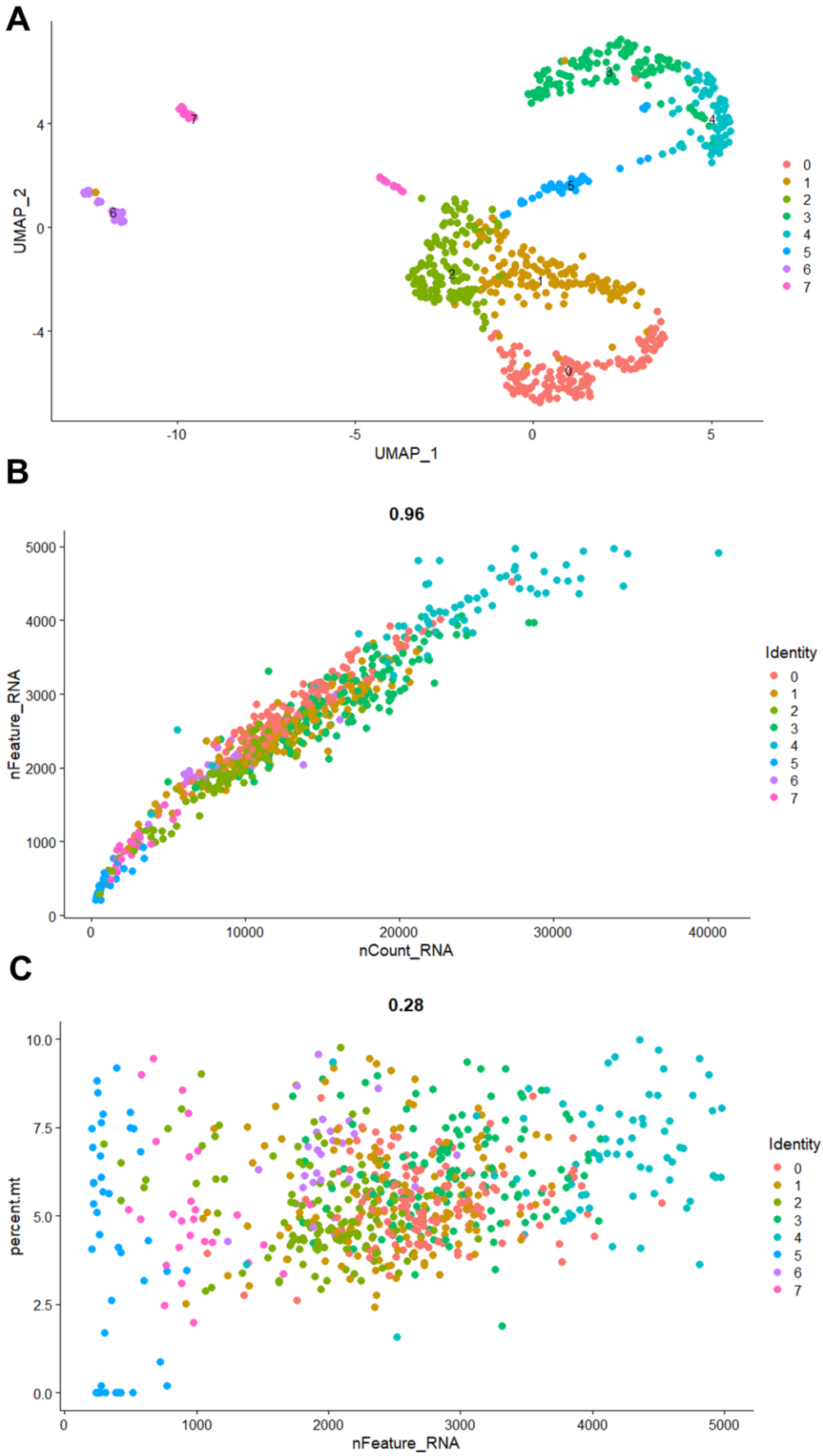
A
UMAP_2
UMAP_1
4
0
-4
-10
-5
0
5
7
6
3
4
5
2
1
0
Identity
0
1
2
3
4
5
6
7
B
0.96
nFeature_RNA
nCount_RNA
5000
4000
3000
2000
1000
0
0
10000
20000
30000
40000
Identity
0
1
2
3
4
5
6
7
C
0.28
percent.mt
nFeature_RNA
10.0
7.5
5.0
2.5
0.0
0
1000
2000
3000
4000
5000
Identity
0
1
2
3
4
5
6
7

Figure 4.3. Non-linear dimensionality reduction and variable feature mapping across clusters

A. Visualisation of graph-based cluster onto dimension reduction plot (UMAP).
B. Scatter plot of RNA counts per cell versus percentage of mitochondrial genes per cell and projection of graph-based clusters.
C. Scatter plot of count of RNA counts per cell versus gene counts per cell and projection of graph-based clusters.

We next scored cells according to their cell cycle phase status using a previously described method that has been incorporated as a function into Seurat (Tirosh et al., 2016). We mapped scores onto the UMAP plot and observed that most clusters were comprised of predominantly G0/G1 cells (Fig. 4.4A). Clusters 5 and 4 consisted of large portion of cells in the S and G2M phases (Fig. 4.4A). Stacked bar plots of the proportion of cells according to their cell cycle identity revealed that cells in clusters 1 through 3 were pretty similar in terms distribution of cell cycle phases (Fig. 4.4B). Clusters 6 and 7 contained the smallest proportions of cells in the S or G2M phases (Fig. 4.4B). This heterogeneity of cell cycle state between clusters can be contributing to some of the principle components that explain the variance of the data. However, rescaling the data to regress cell cycle genes out is not a procedure advised for datasets where mixtures of differentiation processes may entail different cell cycle states.

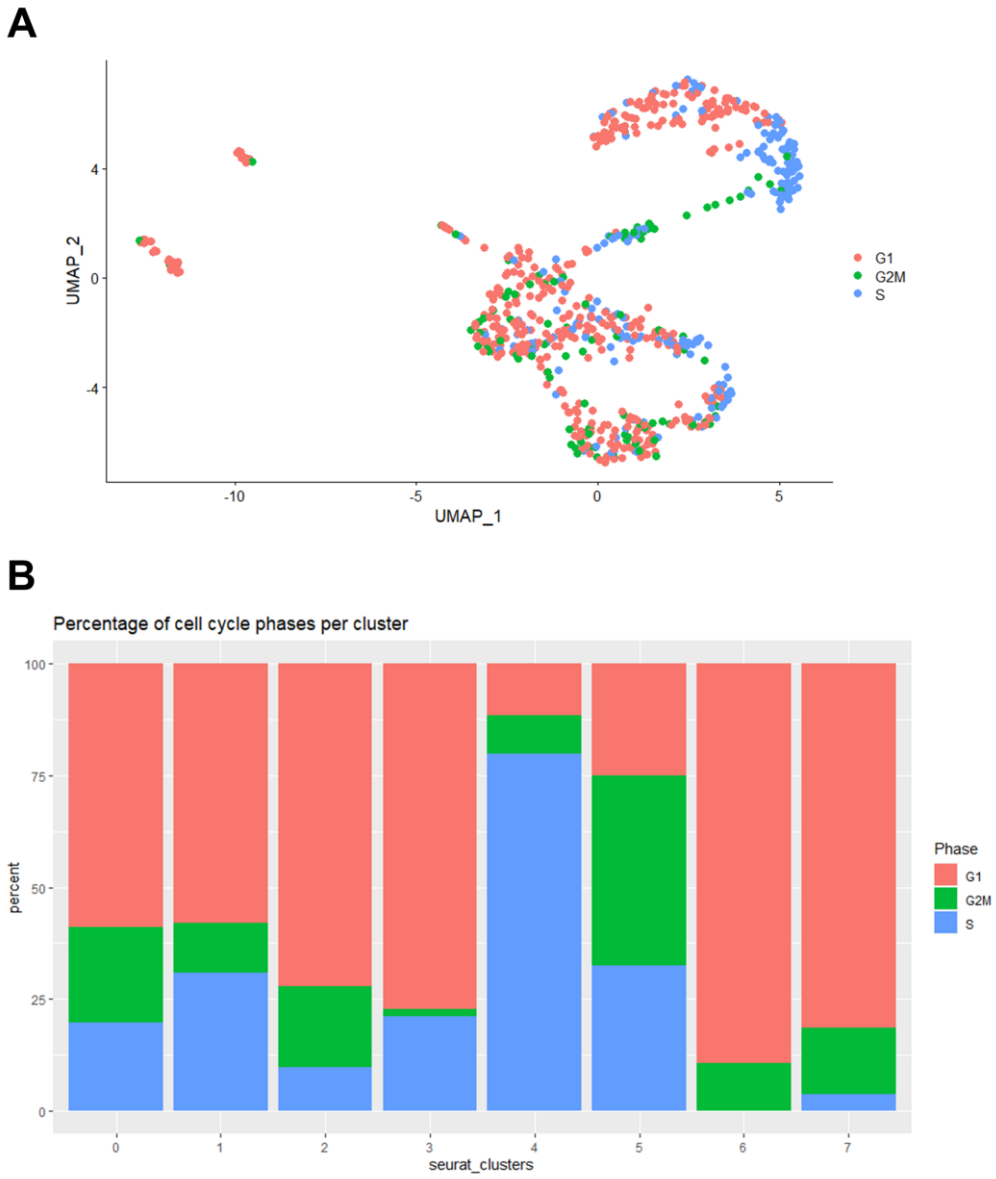

Figure 4.4. Cell cycle phase distribution between clusters

A. Mapping of cell cycle phases of individual cells onto UMAP plot.
B. Stacked bar graphs of the distributions of cells in different cell cycle phases in each cluster.

After quality control, clustering and assessment of cell cycle phases, we looked into the expression of genes known to be expressed by HSCs. We observed that the majority of the genes we investigated, were highly expressed by cells in clusters 0, 1, 2 (henceforth C0, C1, and C2) (Fig. 4.5A). MEIS1, MLLT3, SOCS2, ETV6, KIT were also expressed in clusters 3 and 4 (henceforth C3 and C4). MLLT3 was recently reported to play an important regulatory role in the self-renewal and maintenance of human HSCs (Calvanese et al., 2019). SOCS2 is a suppressor of

JAK/STAT-induced signalling and has been shown to be the most highly expressed member of the protein family in HSCs compared to differentiated cells (Georgantas et al., 2004; Ivanova et al., 2002). In mouse SOCS2 has been shown to be important for regulating HSC expansion in response to cytokine stimulation (Vitali et al., 2015).

The expression of the genes HOPX, CD34, AVP, and SPINK2 was mostly restricted to clusters C0, C1, and C2. CXCR4 was also highly expressed in these clusters, but also in clusters 6 and 7 (henceforth C6 and C7). In mouse, CXCR4 has been shown to be important for maintaining the quiescence of HSCs, but can also be expressed by a B, T, NK cells and granulocytes (Nie et al., 2008; Sugiyama et al., 2006). The genes with most restricted expression were HLF and CLEC9A in C0. Another interesting observation was that PROM1 was also mostly expressed by cells in C0 and C1. Seeing that a significant number of cells expressed CD34 and PROM1 (CD133) was surprising as the sequenced cells are defined as $CD34^-$ and we observed no CD133 cell surface marker expression by $CD34^-/CD38^-/CD93^+$ cells. It is possible that during our sorting strategy we capture cells with low expression of CD34, which would explain why we can detect its expression. It is also possible both CD34 and PROM1 are expressed but degraded as result of post-transcriptional regulation.

It was particularly exciting that the expression HOXA9 and HOXB4 could be detected in the dataset. HOXA9, HOXB3, and HOXB4 have been shown to be expressed by HSCs and downregulated during lineage commitment (Lawrence et al., 2005; Sauvageau et al., 1994; Velten et al., 2017). A small number of cells in C0, C1, and C2 expressed HOXB3, but HOXA9 and HOXB4 were expressed by a larger number of cells (data for HOXB3 not shown). HOXA9 and HOXB4 expression was restricted to clusters 0 through 4 and strongly emphasised in cluster 0. Furthermore, HOX genes require protein cofactors to efficiently bind DNA, and we observed that two well-known HOX cofactors, PBX1 and MEIS1, were also expressed in C0 (W. F. Shen et al., 1996, 1997). HOXA9 and HOXB4 and their effectors have been repeatedly featured in gene sets associated with human HSCs (Doulatov et al., 2010; Guo et al., 2013; Velten et al., 2017)

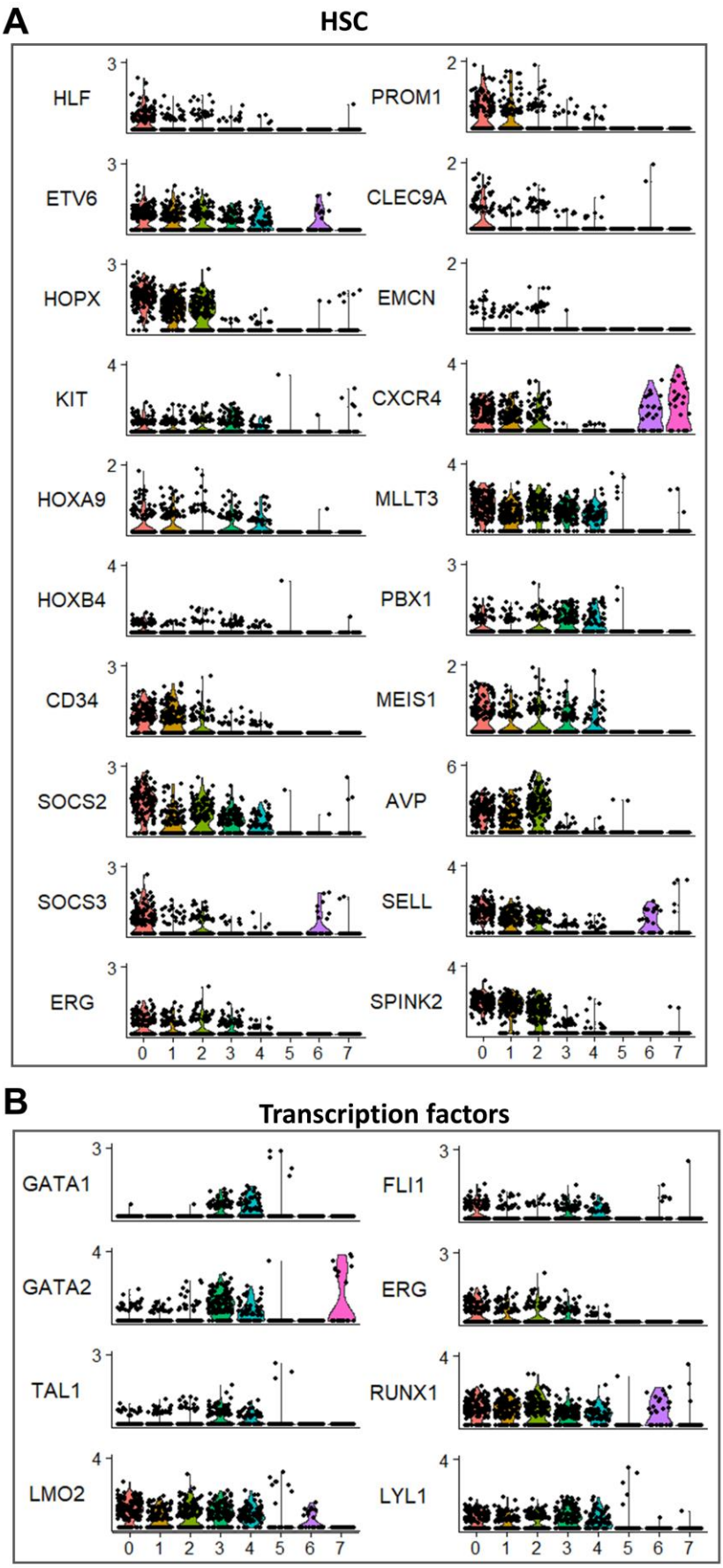
A
HSC
HLF
PROM1
ETV6
CLEC9A
HOPX
EMCN
KIT
CXCR4
HOXA9
MLLT3
HOXB4
PBX1
CD34
MEIS1
SOCS2
AVP
SOCS3
SELL
ERG
SPINK2
0 1 2 3 4 5 6 7
B
Transcription factors
GATA1
FLI1
GATA2
ERG
TAL1
RUNX1
LMO2
LYL1
0 1 2 3 4 5 6 7

Figure 4.5. HSC gene expression signatures in CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} cells

A. Expression of genes associated with HSC function and phenotype across clusters.
B. Expression of heptad transcription factor genes and GATA1.

Transcription factors often must act in concert to activate stem cell and differentiation programmes. RUNX1, GATA2, TAL1, LMO2, FLI1, ERG, and LYL1, collectively known as the heptad transcription factors have been shown to have overlapping binding in human CD34^{+} HSPCs and a murine progenitor cell line (Beck et al., 2013; Wilson et al., 2015). All seven factors were expressed by cells in clusters 0 through 4 (Fig. 4.5B).

GATA2 and GATA1, transcription factors, heavily implicated in megakaryocytic/erythroid differentiation, were both highly expressed in clusters 3 and 4 (Fig. 4.5B). GATA1 has been well characterised as a lineage-specific transcription factor of crucial importance during erythroid and megakaryocyte development (Vodyanik et al., n.d.; Weiss & Orkin, 1995; Welch et al., 2004). Together with ZFPM1 (FOG-1) GATA1 supplants GATA2 at GATA DNA binding sites to orchestrate the erythroid differentiation programme (Anguita et al., 2004; Hamlett et al., 2008; Hoffmann & Spengler, 2019; Welch et al., 2004). ZFPM1 expression correlated with GATA1 in C3 and C4 (Fig. 4.6B). Transferrin receptor genes, TRFC and TFR2 which are important for intracellular iron trafficking, were also most highly expressed in these clusters along with expression of the core erythropoietic LDB1-complex, TAL1, LMO2, LDB1, and TCF3 (E2A) (Love et al., 2014). EPOR was detected only at very low levels in both clusters, which may be due to the technical limitations. A number of other erythrocyte-associated genes such as CNRIP1, ANK1, and CD36 were detected in both clusters, but were higher in C4 (Psaila et al., 2016).

On the other hand, a number of megakaryocyte-specific genes were also expressed in clusters 3 and 4. RUNX1, FLI1, and NFE2, known transcriptional regulators of megakaryocyte maturation, were expressed in these populations, though at low levels (Fig. 4.6B). Genes CLU (Clusterin) and ITG2B (CD41), both associated with megakaryocyte maturation, were also expressed at low levels in clusters 3 and 4 (Fig. 4.6B) (Tschopp et al., 1993). It is possible that the presence of GATA2 in these cells is responsible for inducing the expression of these genes (Huang et al., 2009).

During erythroid maturation heme synthesis precedes globin production, and we observed that a number of genes coding for enzymes in the heme synthesis pathway were expressed in C3 and C4 (Fig. 4.6A). Notably, HMBS, UROD, UROS, PPOX, FECH appeared more highly expressed in C4 compared to C3 (Fig. 4.6A). In contrast a large number of globin genes as well as haemoglobin chaperon gene, AHSP, were highly expressed by cells in cluster 5 (henceforth C5), which identifies this population of cells as more mature erythroid cells (Fig. 4.6A). Heme synthesis genes were not expressed by cells in C5, and globin genes were only poorly expressed in C3 and C4 (Fig. 4.6A).

The process of erythroid differentiation is intimately coupled with the cell cycle. Cell cycle scoring demonstrated that C3 is predominantly composed of cells in G1, whereas C4 is highly enriched for cells in the S phase (Fig. 4.4). This was of particular interest, as it has been shown in mouse foetal liver, that erythroid terminal differentiation occurs in the S-phase of a single cell cycle (Pop et al., 2010). Studies in cell lines have also suggested a role for GATA1 in modulating G1-phase arrest during erythroid maturation (Rylski et al., 2003). It is possible that C3, C4 and C5 represent three distinct stages of erythroid maturation. While both C3 and C4 expressed genes indicating active heme synthesis, cells in C3 were predominantly arrested in G0/1, while C4 cells were in the S-phase. This could indicate distinct erythroid stages, but differential expression of cell cycle genes could also explain the split of the two largely similar clusters by cell cycle phase. With the high expression of globin genes, C5 would certainly comprise of more mature erythroid cells.

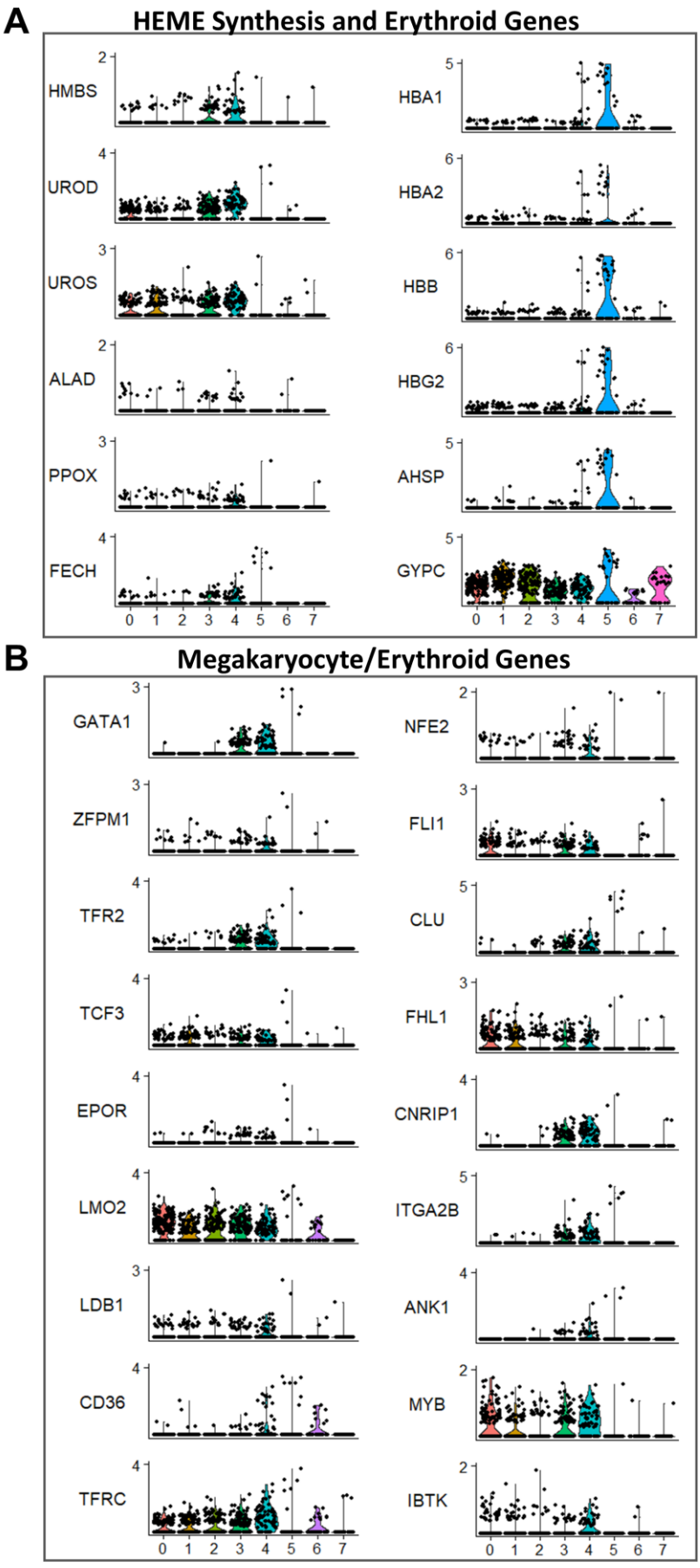
A
HEME Synthesis and Erythroid Genes
HMBS
UROD
UROS
ALAD
PPOX
FECH
HBA1
HBA2
HBB
HBG2
AHSP
GYPC
0 1 2 3 4 5 6 7
B
Megakaryocyte/Erythroid Genes
GATA1
ZFPM1
TFR2
TCF3
EPOR
LMO2
LDB1
CD36
TFRC
NFE2
FLI1
CLU
FHL1
CNRIP1
ITGA2B
ANK1
MYB
IBTK
0 1 2 3 4 5 6 7

Figure 4.6. Megakaryocyte and erythroid gene expression

A. Violin plots of expression of genes coding for key heme synthesis enzymes (column 1) and globin proteins (column 2).
B. Violin plot of expression of genes involved in erythroid and megakaryocyte maturation.

In C6, we found that the top most highly expressed genes, were LYZ, SPI1, and CEBPB, which was strongly suggestive of myeloid cells (Fig. 4.7A). In addition, we found that the genes CD14, ANPEP (CD13), IRF8, and CD68 were active (Fig. 4.7A). These code for cell surface markers strongly associated with monocytes at various stages of maturation (Ingersoll et al., 2010; Ziegler-Heitbrock & Hofer, 2013). However, the mature monocyte marker FCGR3A (CD16) was not expressed (data not shown). The two subunits of the IFNγ (Interferon-γ) receptor were also both expressed in C6. IFNγ signalling has been shown to be important in inducing HLA-DR expression in monocytes, which we also could detect at high levels in cluster 6 (Orchansky et al., 1986). TLR2, a monocyte-associated toll-like receptor, was also present in C6 cells (Fig. 4.7A) (Ropert, 2019).

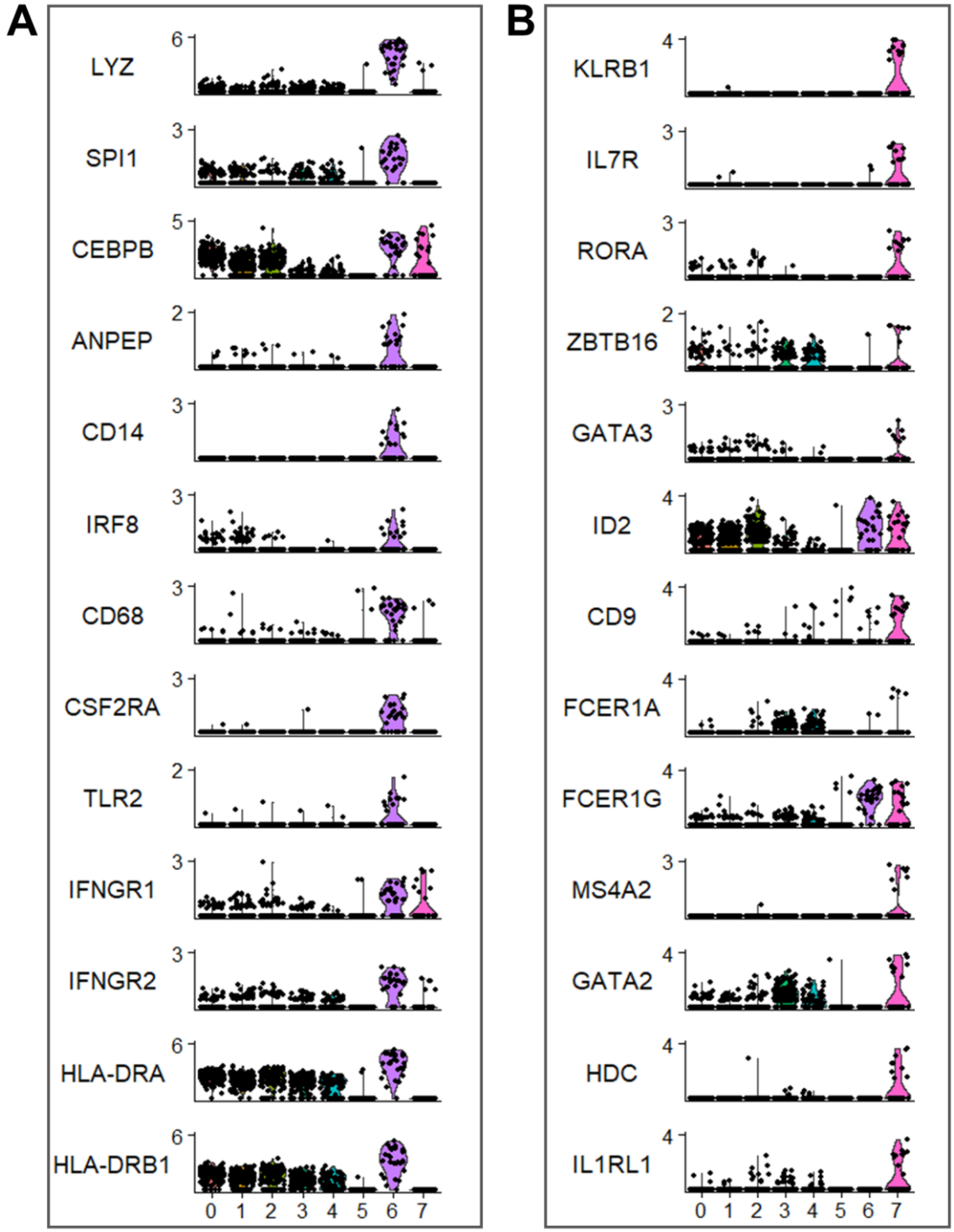

Figure 4.7. Highly expressed genes in clusters 6 and 7

A. Violin plots of the expression of myeloid genes across clusters.
B. Violin plot of the expression of NK and mast cell/basophil markers across clusters.

Cluster 7 (C7) was split in two distinct UMAP populations, one closer to cluster 6 and one closer to cluster 2. One of the top expressed genes in the cluster was KLRB1 (CD161), which is a notable NK cell marker (Fig. 4.7B) (Baev et al., 2004; Berzins et al., 2005). When looking into other NK cell genes, we found that neither of the classical CD16 and CD56 cell surface markers were expressed in the whole dataset. While, NK cells can be double negative for both of these, genes such as IFNγ, CD90, PRF1, GZMA and GZMB also weren't expressed in the dataset. In addition, key transcription factors for NK cell development, EOMES and TBX2 were missing (Luetke-Eversloh et al., 2013).

KLRB1 is also strongly associated with lymphoid or lymphoid like cell populations. EOMES and TBX21 are also crucial in the differentiation of innate lymphoid cells (ILCs) of type 1, while RORC and GATA3 have been shown to be upregulated in the differentiation of ILCs type 3 and ILCs type 2, respectively (Spits et al., 2013). While RORC was not expressed, GATA3 was present in C7. RORA, a transcription factor implicated in ILC type 2 cells, could also be detected. Additionally, IL7Rα, an ILC 2 cell surface marker, was also expressed at high levels by cells in C7 (Fig. 4.7B).

Type 2 ILCs secrete IL-5 and IL-13, but we could observe no transcripts for either of the interleukins. As IL-7 signalling is important in the maturation of innate lymphocytes and the IL-7 receptor was present in these cells, it is possible that this population consists of immature type 2 ILCs. This is further supported by the presence of ID2 in these cells, which has been shown to be expressed in NK and ILC precursors (Fig. 4.7B) (Boos et al., 2007).

Compelled by the partitioned structure of the cluster, we investigated the distribution of some of the markers. We found that all markers involved in the ILC2 gene signature localized to just one of the subclusters (termed C7.1). Cluster 7.2, localizing closer to C2 expressed high levels of CD9, CD63, CD151, HDC and a number of FCεRs (FC-epsilon receptors) (Fig. 4.8B). FCεRs are IgE receptors that can be expressed on B cells, eosinophils, basophils and mast cells. In C7.2, we observed no expression of the low-affinity FCER2, characteristic of B cells. However, all three high affinity FCεRs, FCER1A, FCER1G, and MS4A2 (FCERI) were expressed only in C7.2 (Fig. 4.8B) (Prussin & Metcalfe, 2006). Additionally, CD9, CD63 and CD151 tetraspanins, have been shown to be important in degranulation of mast cells and to interact with FCεRs (Köberle et al., 2012). Another distinctive

feature of this subcluster was the expression of GATA2, which was absent from C7.1 (Fig. 4.8B). GATA2 has also been shown to be important in mast cells by regulating the expression of HDC, a key enzyme in the biosynthesis of histamine in mast cells and basophils (Y. Li et al., 2016).

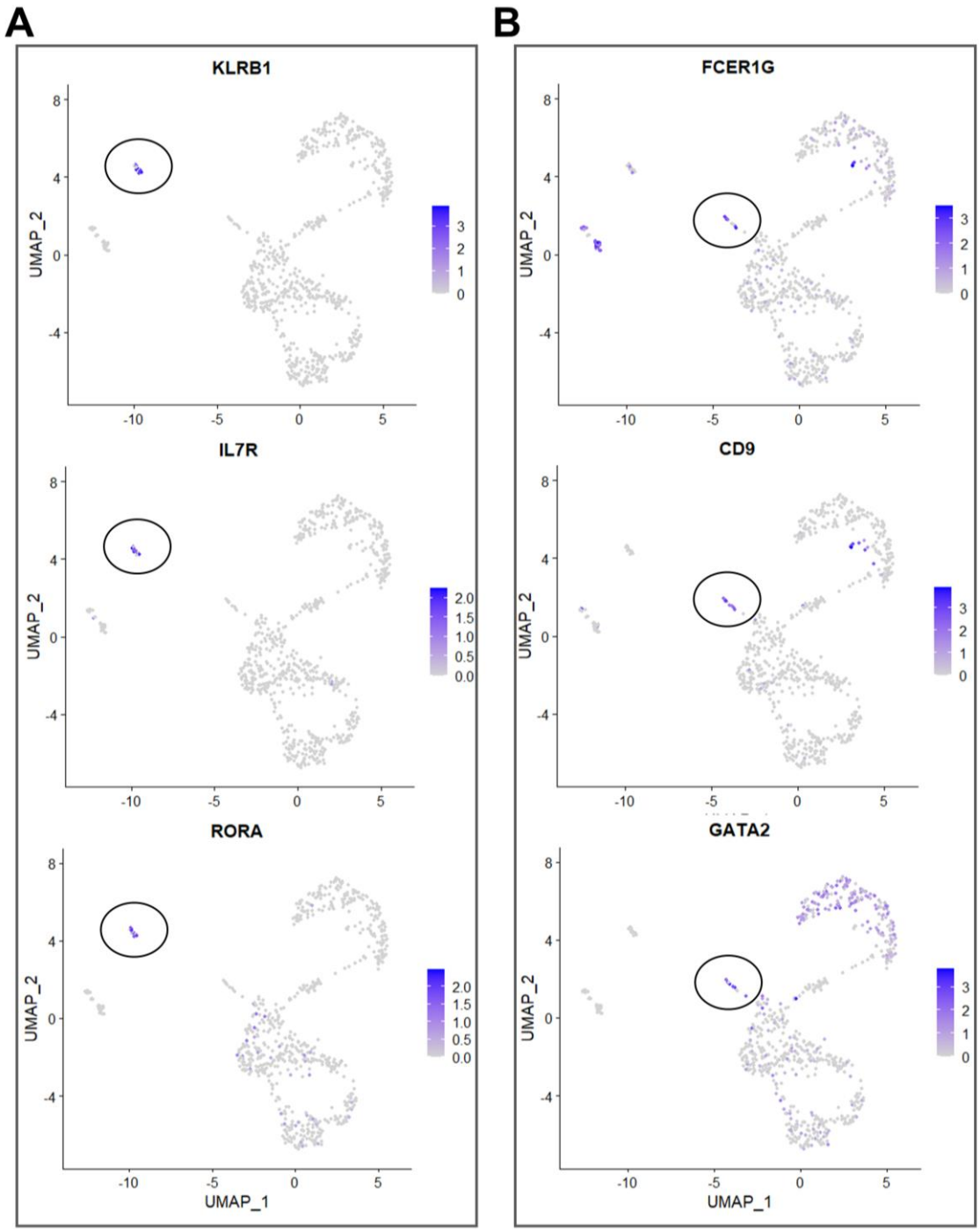

Figure 4.8. Distinct expression patterns between subclusters in cluster 7

A. Three highly expressed genes (KLRB1, IL7R, RORA) in subcluster 7.1 (black circle).
B. Three highly expressed genes (FCER1G, CD9, GATA2) in subcluster 7.2. (black circle)

Following the examination of the highest expressed genes in the eight clusters, in our dataset generated from the sequencing of FACS-sorted, $CD34^{-}$/$CD38^{-}$/$CD93^{+}$/$CD117^{Hi}$ cells, we established that C0, C1, and C2 were highly enriched in classical HSC genes. Some of these genes were also expressed in C3, but in the UMAP projection of the data, C3 was localised further from the C0, C1, and C2 group, and placed closer with C4. C3 and C4 were expressing genes associated with the erythroid and megakaryocyte lineages as well as a number of erythroid differentiation genes such as GATA1, TFR2, TFRC and genes coding for heme synthesis enzymes. Further down the erythroid lineage, we placed C5 with the high expression of globin genes and chaperone, AHSP, suggesting haemoglobin assembly in these cells.

With the expression of CEBPB, SPI1, CSF2RA, we identified C6 cells as myeloid. The expression of CD14, ANPEP (CD13) and IRF8 further suggested that C6 is likely monocytes.

Finally, we managed to divide C7 into two subclusters, C7.1 and C7.2. We found that C7.1 was expressing genes KLRB1 and IL7R, and transcription factors associated with innate lymphoid cells of type 2 such as GATA3 and RORA. C7.2, conversely expressed a few distinct cell surface markers and tetraspanins, CD9 and CD63. In addition, FCεRs, important mediators of the allergen-induced immune response, localised specifically to C7.2.

4.3 Characterisation of HSC molecular signatures

Due to contaminating differentiating cells that we characterised in the dataset, the next step in our analysis was to subset C0, C1, and C2, which we found were enriched for HSC signature genes. Despite high expression of erythroid and heme synthesis genes, we also included C3 in this part of the analysis as this cluster expressed all of the heptad transcription factors and consisted largely of cells in the G1 phase. After subsetting, the data were rescaled to ensure subtle differences in

gene expression between the clusters will be captured. Despite increasing the resolution of the analysis, the clustering algorithm yielded 4 clusters again (Fig. 4.9). This indicates 4 clusters sufficiently group the data according to patterns of gene expression.

As the data was clustered again, the new cluster identities do not correspond to the numbers that may have applied in the initial analysis. For clarity, the cluster identity in this follow-up analysis will be referred to HSC0, HSC1, HSC2 and HSC3. Relying on UMAP for dimensionality reduction, we found that HSC0 and HSC3 clustered closely together (Fig. 4.9A). HSC1 clustered close to HSC3, but was further away from HSC2. The gene expression heatmap of the top 25 expressed genes per cluster showed that HSC2 had high expression of a number of genes associated with the erythroid lineage, GATA2, CNRIP1, and ZEB2 (Fig. 4.9B). This cluster also appeared to share few of its top 25 most highly expressed genes with other clusters (Fig. 4.9B). Conversely, HSC0, HSC1, and HSC3 appeared to share the expression of some genes (Fig. 4.9B).

In HSC0, ID1 and ID2 stood out as the top expressed genes. In mouse, the former has been shown to be suppressor of lymphoid fate, while in humans they have been reported to be more enriched in HSCs compared to progenitor populations (Cochrane et al., 2009; Doulatov et al., 2010; Jankovic et al., 2007; Notta et al., 2011). ID2 has further been confirmed to be highly expressed in human HSCs and shown to drive cells toward the myeloid lineage and maintain HSC self-renewal (van Galen et al., 2014). MEG3, coding for a long-noncoding RNA, specifically expressed in murine LT-HSCs was expressed in HSC0 (Sommerkamp et al., 2019). HSC1 was characterised by high expression of HOPX, SOCS2, and SOCS3 which have all been associated with HSCs (Fig. 4.9B). Another top expressed gene in this cluster was CEBPB, which was also present in C0-2 in the initial analysis.

Finally, HSC3 stood out with a number of genes associated with the lymphoid lineage such as IGHM, MZB1, and LTB1. Two NFκβ inhibitors, NFKBIA and NFKBIE were also in the HSC3 top 25 (Fig. 4.9B).

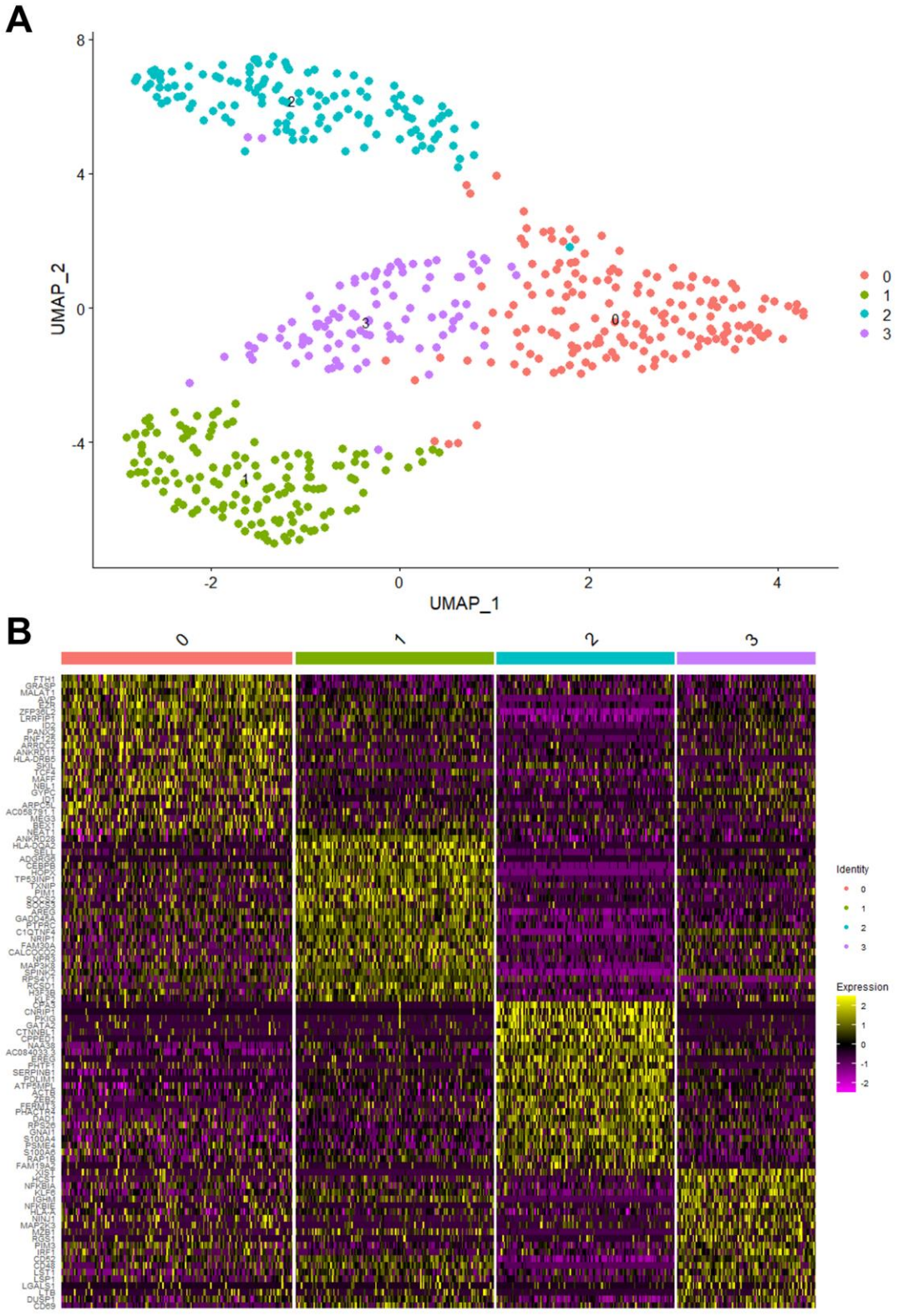

Figure 4.9. Re-clustering of HSC-like subsets from primary analysis

A. Newly calculated clusters of the data subsets mapped onto a UMAP projection of the data.
B. Heatmap of top 25 highly expressed genes across the four clusters.

As the top expressed genes in HSC2 were associated with the erythroid and megakaryocyte lineages, it is likely this cluster represented C3 in the initial analysis, which also had an MEP signature. We checked for the expression of similar genes in HSC2, and found genes such as GATA1, TFR2, HBD, and CNRIP1, distinctly erythroid genes to be expressed (Fig. 4.10). Other detectable genes in this cluster included GATA1 binding partners, TCF3 (E2A), and LMO2 (Fig. 4.10). Two other important GATA1-interacting proteins ZFPM1 (FOG-1) and LDB1 were also expressed, but by a small number of cells. Similar to what we observed in C3 in the initial analysis, a small number of cells expressed low levels of EPOR, which could indicate early stage of erythroid development. What we found compelling was that ITGA2B (CD41) and PLEK transcripts were also present at high levels in significant number of cells. However, we could not detect characteristic megakaryocyte genes such as THPO, MPL, VWF, and NFE2.

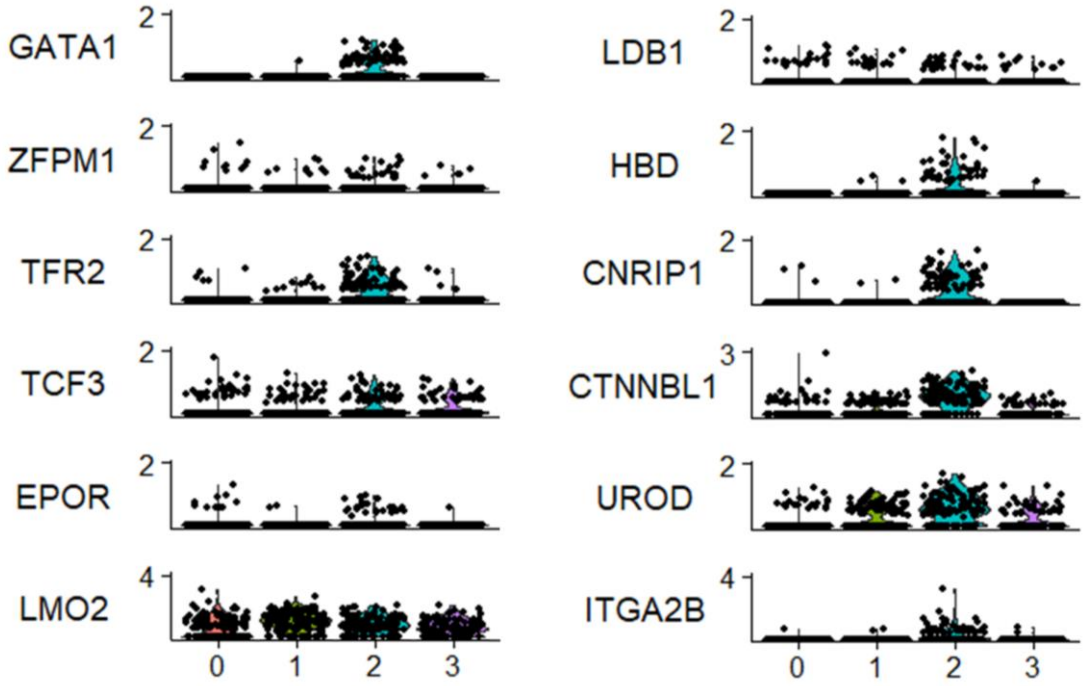

Figure 4.10. Erythroid gene signature in HSC2

Violin plots of expression of erythroid genes and ITGA2B across HSC0-3.

A few haematopoietic transcription factors were present in HSC2, most notably GATA1, GATA2 and TAL1 (Fig. 4.11). KLF1, however, a potent transcriptional regulator of the erythroid fate was absent. GATA2 was more highly expressed than GATA1, which taken together with the low expression of GATA1 partners ZFPM1, TCF3 and LDB1 implies the cells in this cluster may still retain

capacity to differentiate toward the megakaryocyte fate (Fig. 4.10 & 4.11). It is also possible with the presence of both FLI1 and GATA1 that this is an MEP-like population with the potential to generate both lineages (Fig. 4.11).

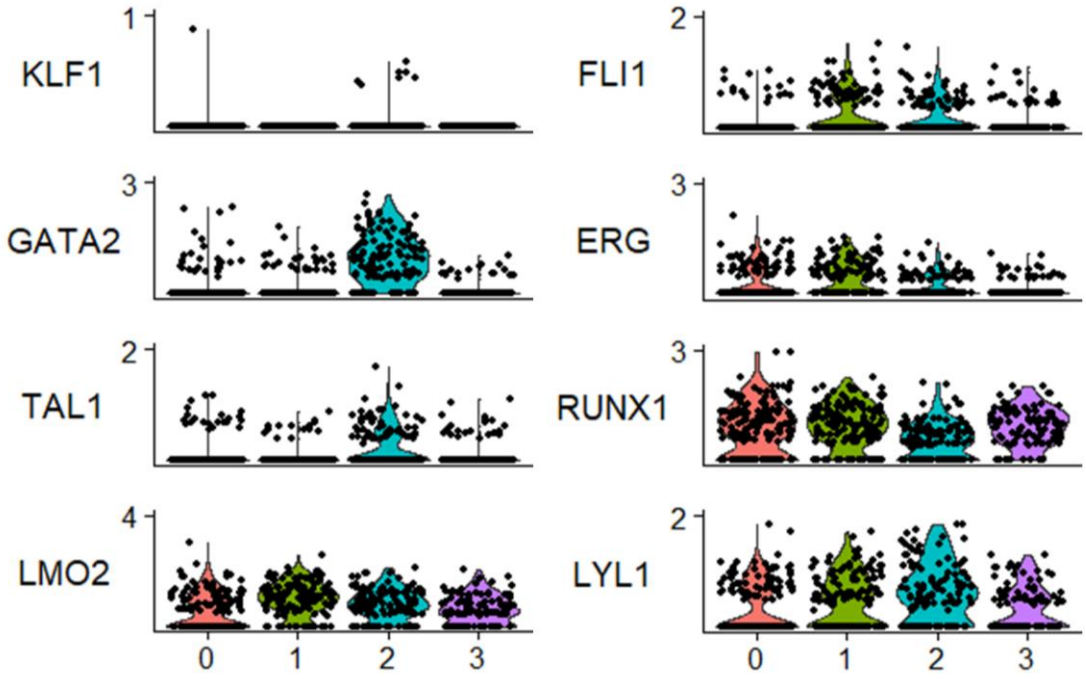

Figure 4.11. Haematopoietic transcription factor expression across clusters
Violin plots of haematopoietic transcription factor genes (points represent individual cells).

GATA2 was expressed by few cells in clusters HSC0, 1 and 3, while LMO2, RUNX1, and LYL1 featured in all four (Fig. 4.11). In addition, HSC1 expressed FLI1 and ERG.

Given the expression of lymphoid genes in HSC3 and suppressors of the lymphoid fate in HSC0, we investigated the status of other key genes involved in the regulation of lymphoid fate commitment (Fig. 4.9B). We didn't observe any expression for EBF1, PAX5 or CD19, and found only small number of cells expressing GFI1 (Fig. 4.12). EBF1 target genes CD79A and CD79B also weren't expressed. On the other hand, ETS1 was present in HSC0 and HSC1, and IKZF1 highly expressed in all clusters but at lower levels in HSC2. Looking more closely at other genes associated with lymphoid cells, HSC3 expressed LTB, BCR, ADA, and SYK (Fig. 4.12). These genes were also present in HSC1, but were absent from HSC0. Given the expression of ID1 and ID2, lymphoid fate may be supressed in this

cluster. TCF7 regulator of T-cell and innate lymphoid cells was also upregulated in HSC3 (Harly et al., 2020). Overall, HSC3 and HSC1 showed a robust expression of lymphoid genes, though key genes in lymphoid commitment were not detected, suggesting lymphoid priming rather than engagement of lymphoid commitment.

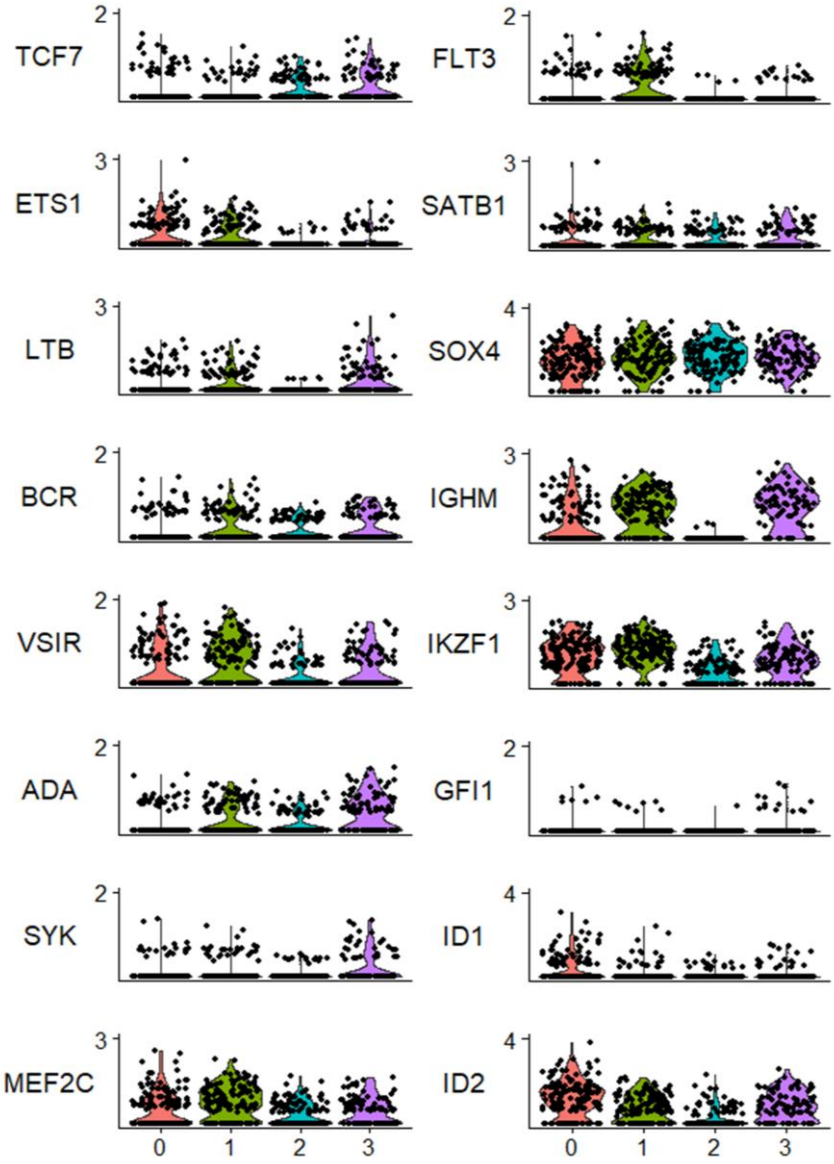

Figure 4.12. Lymphoid gene expression in HSC0-3

Violin plots of the expression of lymohoid genes across clusters (points represent expression in individual cells).

Suppression of lymphoid fates in HSCs has been tied with HSC maintenance as well as myeloid priming, so we looked into some myeloid genes. CEBPB was highly expressed in HSC1, but it was expressed also by a significant number of cells

in HSC0 and to lesser degree HSC3 (Fig. 4.13). Other members of the CEBP family were not expressed. SPI1 was only present in few cells in HSC1, HSC2 and HSC3. We also observed some LYZ expression in both HSC1 and HSC3. Interestingly, CSF3R was expressed in HSC0, 1 and 3 (Fig. 4.13). In addition, a few genes previously described as markers of immature myeloid progenitors, CD52, SPINK2 and SELL, were all expressed in these clusters as well (Velten et al., 2017). EVI2B, a target gene of CEBPA and a known regulator of granulopoiesis, was also expressed in HSC0, HSC1, and HSC3 (Zjablovskaja et al., 2017). In terms of other transcription factors, we looked into ERG, EGR1, GATA2, MYB and IRF8. All except IRF8 were present in HSC2, while ERG, EGR1 and MYB were also detectable in HSC1 (Fig. 4.13). While MLPs tend to express predominantly lymphoid genes and GMPs predominantly myeloid, LMPPs tend to share gene signatures with both populations (Karamitros et al., 2018). We also showed that HSC0-2 express PBX1 which has been implicated in supressing myeloid differentiation, while maintaining the lymphoid potential of LMPPs (Ficara et al., 2013).

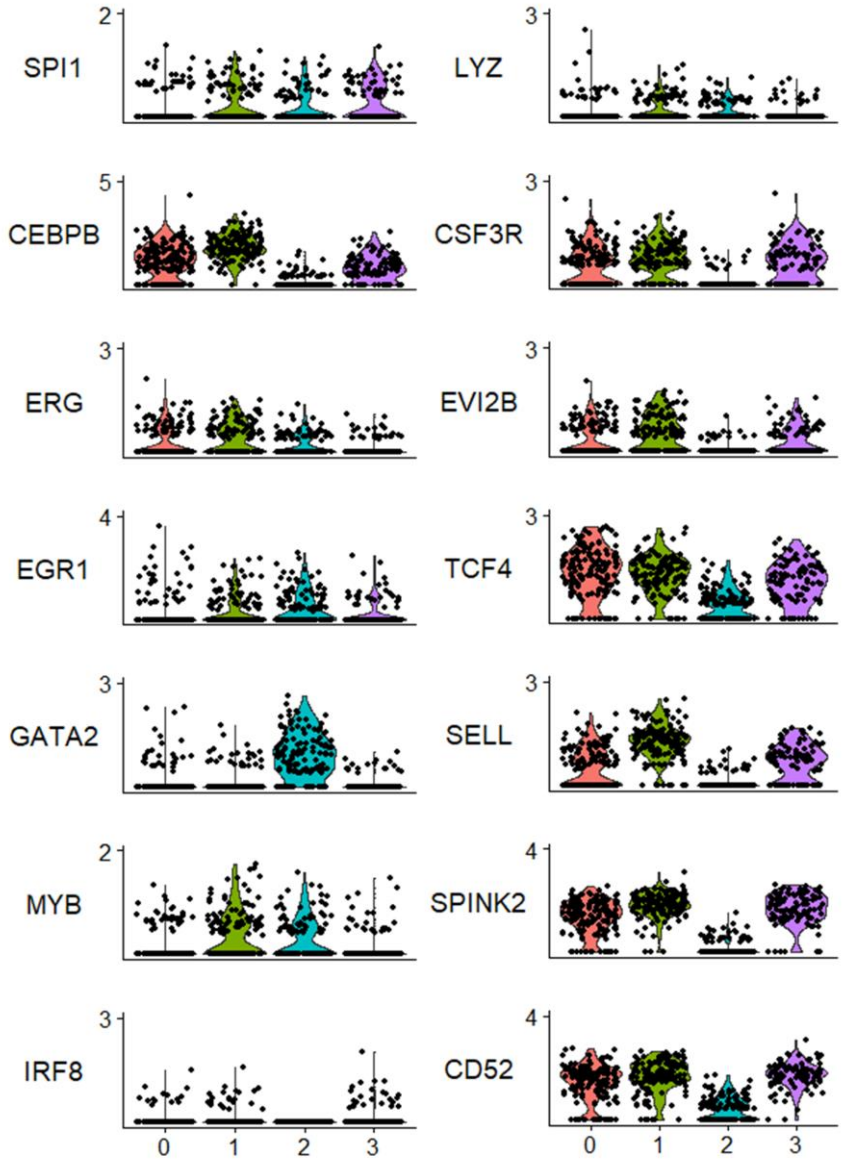

Figure 4.13. Myeloid gene signatures

Violin plot of expression transcription factors and genes associated with myeloid cell fates in HSC0, 1, 2, and 3.

Finally, we looked into the expression of HSC-like genes. By looking at about 20 genes that have well-described associations with HSC either functionally or as cell surface markers, we found that the majority of them were most highly expressed in HSC1 (Fig. 4.14). This included HOPX, HLF, MLLT3 and SPINK2. HOX genes HOXA9 and HOXB4 were also present at higher levels in this cluster. Suppressors of cytokine signalling SOCS2 and SOCS3 were also highly expressed, and taken into account that CSF3R is also present in HSC1, suggests that cells in this cluster

may be refractory to G-CSF (CSF3) stimulation. In terms of cell surface marker expression, this cluster was also well endowed with expression of CD34, PROM1 and CXCR4. While flow cytometry analysis has been unable to confirm the expression of PROM1 on the surface of $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ cells, we cannot discount the possibility that it could still be transcribed. On the other hand, CXCR4 has been suggested previously as a potential purification marker for $CD34^{-}$ HSCs (Bhatia et al., 1999; Danet et al., 2002). A compelling marker we found to be expressed by HSC1 was CLEC9A, which has been recently reported to mark a $CD34^{Low}$ population enriched for HSC with long-term multipotent repopulation potential (Belluschi et al., 2018).

In summary, our downstream analysis identified the expression of erythroid and megakaryocyte genes in HSC2. Interestingly, we found a number of lymphoid genes to be expressed in HSC0, 1, and 3, with HSC3 expressing higher levels of more mature lymphoid markers. While we found key lymphoid transcription factors like ETS1, TCF7 and IZKF1 to be expressed, some important regulators of more advanced lymphoid maturation like GFI1, EBF1 and PAX5 were absent which could indicate early lymphoid priming. In a similar pattern, we found myeloid TFs SPI1 and CEBPB to be present in HSC1, but very few cells to express IRF8. On the other hand, CSF3R was present in HSC0, 1, and 3, and low levels of LYZ in HSC1 and HSC2. Finally, we showed that a number of HSC regulators and markers were present in HSC0, 1, and 3, with high levels in HSC1. Some of these regulators included key transcription factors and transcription factor binding partners like HLF, HOPX, HOXA9, HOXB4, MEIS1, and PBX1. This promiscuous expression of transcription factors associated with different lineages was highly compelling as it has been previously shown to be characteristic of early HSCs (Miyamoto & Akashi, 2005). What we also found interesting was the expression of CLEC9A, which has been recently shown to mark a population of immature populations of HSCs in cord blood (Belluschi et al., 2018).

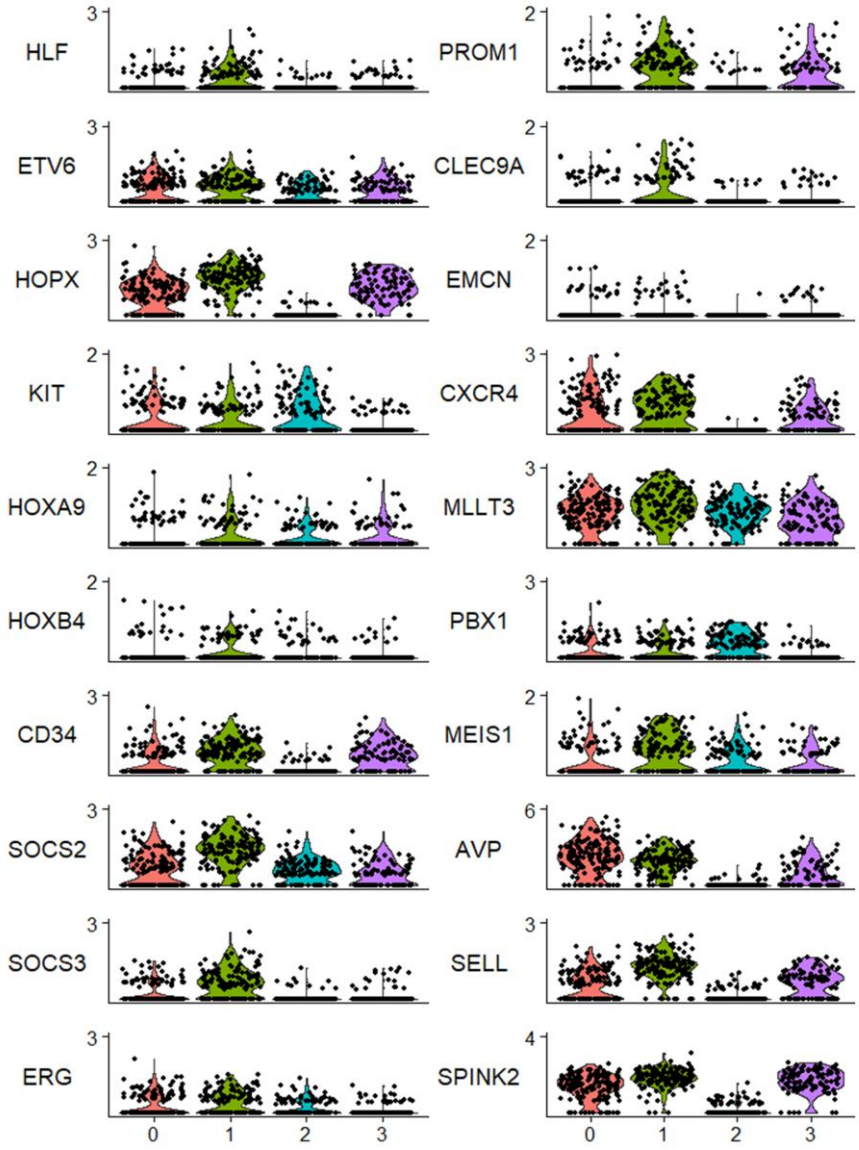

Figure 4.14. HSC transcription factors, regulators and markers

Violin plots of expression of functional and cell surface markers in clusters HSC0, 1, 2, and 3

4.4 Molecular pathway enrichment in HSC-like subsets of the CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} population

Our attempt to characterise the identities of the subset HSC-like clusters, while successfully pointing out some differences, was still strongly suggesting that HSC0, HSC1, and HSC3 are highly similar. To further elucidated the differences between these populations we undertook GO processes enrichment analysis using the R package ClusterProfiler, which relies on the Bioconductor GO.db package as a gene

ontology repository. We calculated differentially expressed genes for each HSC cluster and excluded insignificant genes (Adjusted p-value>0.05). For the GO analysis, we calculated enriched GO processes separately for upregulated and downregulated processes for each cluster.

HSC0 stood out with having far more downregulated than upregulated genes, 132 and 56, respectively. Consequently, we found far more downregulated processes compared to upregulated ones. One of the top upregulated process that immediately stood out was related to regulation of mRNA catabolism, suggesting degradation of mRNA being an active process in these cells (Fig. 4.15). Another upregulated process that was indicative of the potential quiescence of this cluster compared to the rest was downregulation of protein secretion. Ossification is process that has been shown to be regulated by BMP signalling, and we observed that both negative regulation of ossification and BMP signalling are downregulated in this cluster. We also saw positive regulation of T cell-mediated immunity to be upregulated in HSC0, which was driven solely by the expression of HLA genes. On the other hand, the processes that were downregulated in this cluster included oxidative phosphorylation and ATP synthesis (Fig. 4.15). Quiescent HSCs are thought to be metabolically inactive and reliant on glycolysis rather than oxidative phosphorylation (Simsek et al., 2010; Takubo et al., 2013; Umemoto et al., 2018; Vannini et al., 2016). This could be a strong indicator that this is the most quiescent portion of the HSC-like subset, and its low metabolic state may also explain why catabolic mRNA processes are highly active.

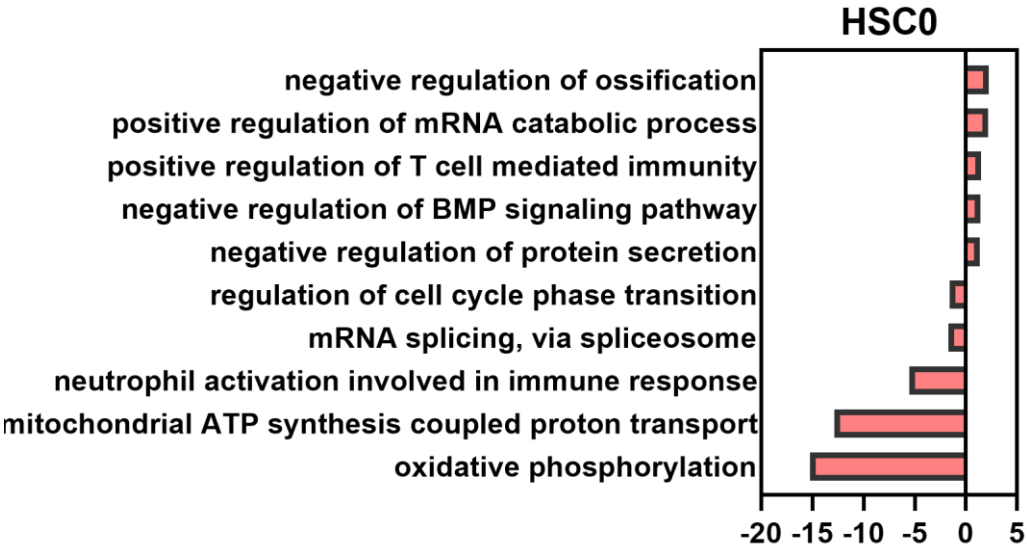

Figure 4.15. Gene ontology processes enrichment in subset HSC0 of the CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} population

Bar plot of significantly enriched GO processes. Values on plotted on the x axis represent –log10 of adjusted p values for upregulated processes and log10 of adjusted p values for downregulated processes. Respective GO processes accession numbers are GO:0030279, GO:0061014, GO:0002711, GO:0030514, GO:0050709, GO:1901987, GO:0000398, GO:0002283, GO:0042776, GO:0006119

The most highly enriched process in HSC1, “response to interferon-gamma”, came as a surprise as we could detect no expression of the IFNγ receptors, IFNGR1 and IRNGR2 (Fig. 4.16A). Similarly, we could only detect very low expression of well-known targets of IFNγ signalling like IRF2, IRF8, and IRF9, while others like IRF1, RELA and JUN were also expressed in other clusters (Fig. 4.16B). Two other processes that were upregulated were related to myeloid cell differentiation and neutrophil degranulation. We already noted that this cluster stood with a robust myeloid gene signature. Interestingly, erythrocyte homeostasis also came up in the upregulated processes due to genes like KLF2 and ZEB1. Downregulated processes involved RNA splicing, which we found compelling as it has been shown that quiescent HSCs are more enriched for unspliced transcripts (Bowman et al., 2006). However, alternative splicing feature in latter commitment, as it has been shown that the rate of intron retention increases in as granulocytes mature from promyelocytes which has been tied with drop of expression of splicing factors (J. Wong et al., 2013). Another report has also shown that knockout of RUNX1 alternative splicing isoforms in mouse resulted in significantly lower HSC numbers (Komeno et al., 2014).

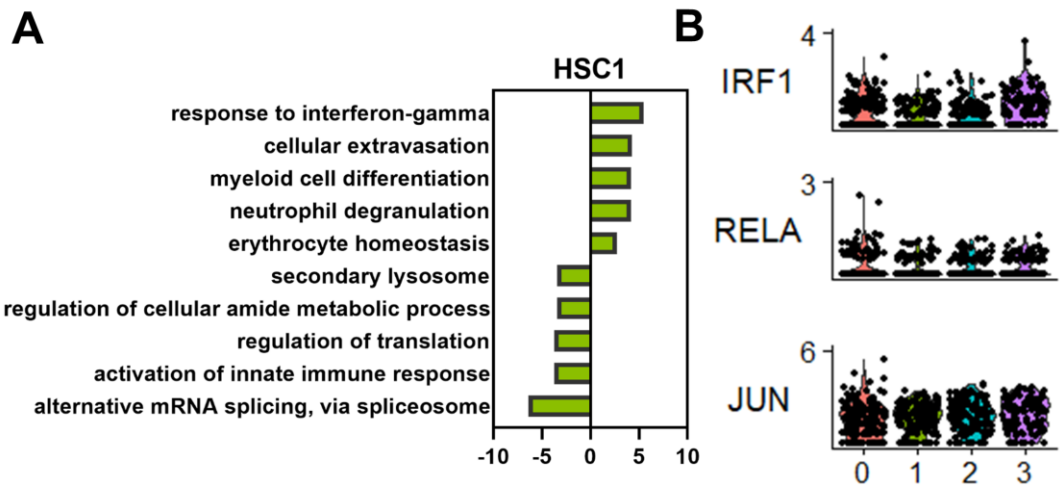

Figure 4.16. Gene ontology processes enrichment in subset HSC1 of the CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} population

Bar plot of significantly enriched GO processes. Values plotted on the x axis represent –log10 of adjusted p values for upregulated processes and log10 of adjusted p values for downregulated genes. Respective GO processes accession numbers are GO:0034341, GO:0045123, GO:0030099, GO:0043312, GO:0034101, GO:0005767, GO:0034248, GO:0006417, GO:0002218, GO:0000380.

In HSC2, we found that the top upregulated processes are neutrophil and platelet degranulation, which involve similar set of genes (Fig. 4.17). This confirms our previous observation that some megakaryocyte genes were observed in this cluster. Another two connected processes were DNA synthesis and telomere maintenance, which could suggest that this cluster is enriched for cells preparing to enter S phase of the cell cycle. Protein folding is also a key process in the gradual maturation of specialised cells such as erythrocytes and megakaryocytes which must assemble large proteins such as haemoglobin or to expand the cell membrane in preparation for endomitosis. Some of the downregulated processes included myeloid and lymphoid differentiation, confirming that HSC2 is enriched for cells committed toward the megakaryocyte-erythroid lineage.

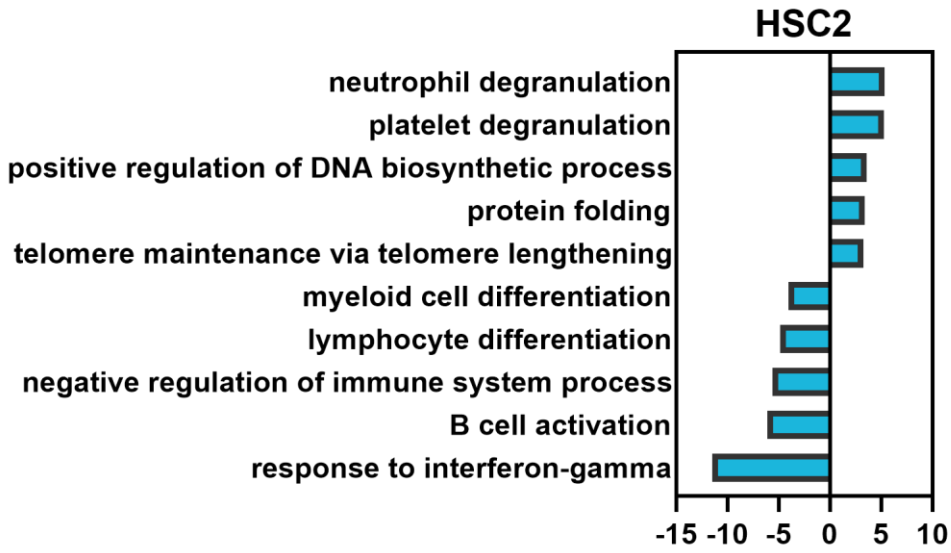

Figure 4.17. Gene ontology processes enrichment in subset HSC2 of the CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} population

Bar plot of significantly enriched GO processes. Values plotted on the x axis represent –log10 of adjusted p values for upregulated processes and log10 of

adjusted p values for downregulated genes. Respective GO processes accession numbers are GO:0043312, GO:0002576, GO:2000573, GO:0006457, GO:0010833, GO:0030099, GO:0030098, GO:0002683, GO:0042113, GO:0034341

In HSC3, we found that the top upregulated processes confirmed our earlier observations that this cluster is highly enriched for more late lymphoid genes. The biological processes upregulated in this cell population were involved with regulation of lymphocyte activation, lymphocyte/leukocyte proliferation and leukocyte migration (Fig. 4.18). In the downregulated processes, what stood out was the downregulation of MHC II class protein complex genes, which suggests that this cluster may represent a population with stronger divergence from the transcriptional programmes of innate immune cells, compared to HSC0 and HSC1.

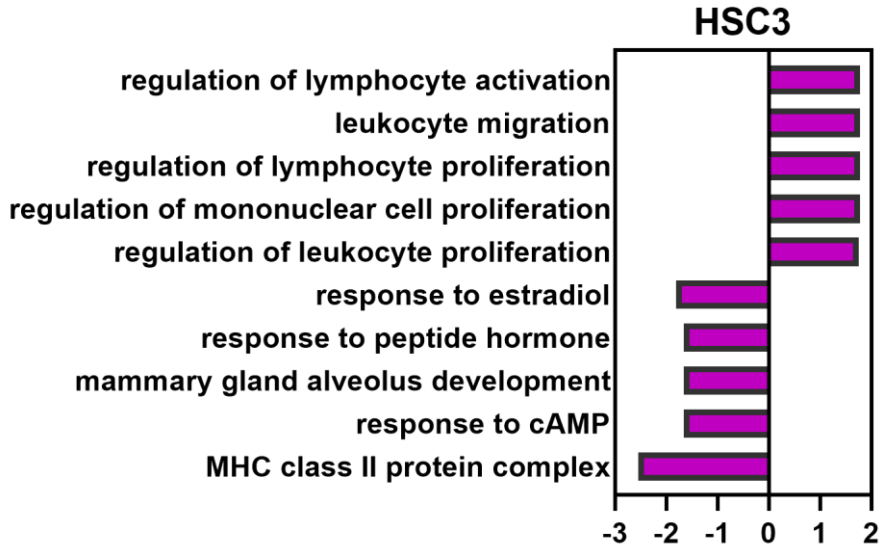

Figure 4.18. Gene ontology processes enrichment in subset HSC2 of the CD34-/CD38-/CD93+/CD117Hi population

Bar plot of significantly enriched GO processes. Values plotted on the x axis represent –log10 of adjusted p values for upregulated processes and log10 of adjusted p values for downregulated genes. Respective GO accession numbers are GO:0051249, GO:0050900, GO:0050670, GO:0032944, GO:0070663, GO:0032355, GO:0043434, GO:0060749, GO:0051591, GO:0042613.

In summary, GO processes upregulated in HSC2 and HSC3 suggested the two populations may be slightly more mature with clearer signatures toward the megakaryocyte-erythroid and lymphoid lineages, respectively. In HSC3 we saw downregulation of MHC II genes with strong upregulated lymphoid signature. In

HSC2, platelet degranulation and protein folding were upregulated, while processes related to the myeloid and lymphoid lineages were downregulated.

On the other hand, HSC0 and HSC1 were harder to type. We found very few upregulated genes in HSC0 and larger number of downregulated genes. One of the most upregulated processes in this cluster was mRNA catabolic processes which suggest active post-transcriptional gene silencing. On the other hand, oxidative phosphorylation and mitochondrial ATP synthesis were downregulated. We found no significant overrepresentation of genes involved in glycolysis in this cluster, but the strong downregulation of genes involved in mitochondrial activity could indicate quiescence. In HSC1, we saw processes like myeloid differentiation, neutrophil degranulation and erythrocyte homeostasis. It has been previously reported that myeloid and erythroid transcriptional programmes overlap (Laurenti et al., 2013). What we also found was significantly downregulated in HSC1, was alternative splicing programmes. Higher frequency of unspliced transcripts in HSCs has been reported before, but not fully elucidated in terms of functionality (Bowman et al., 2006).

4.5 Pre-ranked gene set enrichment analysis reveals different haematopoietic signatures within the HSC-like subsets of the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ population

While HSC2 and HSC3 appeared to contain more committed cells, we were interested to find out if lineage priming programmes may be already occurring in HSC0 and HSC1. Published gene sets generated from highly purified human HSC and progenitor populations and lineage-priming gene modules were used to conduct gene set enrichment analysis on the four subset clusters (Table 4.1). A pre-ranked GSEA was ran for each of the HSC clusters. The average log fold change was used to rank all significant differentially expressed genes for each cluster.

Gene Set Name	Number of Genes	Reference
HAY_BONE_MARROW_CD34_POS_HSC	127	(Hay et al., 2018)
EPPERT_CE_HSC_LSC	41	(Eppert et al., 2011)
EPPERT_HSC_R	126	(Eppert et al., 2011)
ZHENG _PUTATIVE_MEGAKARYOCYTE_PROGENITOR	98	(Zheng et al., 2018)
ZHENG_MEGAKARYOCYTE_ERYTHROID_PROGENITOR	98	
ZHENG _PUTATIVE_EARLY_ERYTHROID_COMMITMENT	100	
ZHENG _HSC_MULTIPOTENT_PROGENITOR	97	
ZHENG_LYMPHOID_PRIMED_MULTIPOTENT_PROGENITOR_1	95	
ZHENG_LYMPHOID_PRIMED_MULTIPOTENT_PROGENITOR_2	97	
ZHENG_GM_PROGENITOR	100	
ZHENG _MULTILYMPHOID_PROGENITOR	97	
ProB_250	250	(Laurenti et al., 2013)
MLP_250	250	
MEP_250	250	
HSC_250	250	
GMP_250	250	
ETP_250	250	
EarlyB_250	250	
CMP_250	250	

Table 4.1. Gene sets used in the gene set enrichment analysis, number of genes in each list and source publication.

We found that HSC0 was significantly enriched for two HSC-specific gene signatures. Both gene sets were generated from population-specific studies that isolated and defined HSCs by positive CD34 cell surface marker expression (Laurenti et al., 2013; Zheng et al., 2018). The normalized enrichment scores (NES) of 3.8 and 2.5 indicated a strong overrepresentation of $CD34^+$ HSC-specific signatures among the top-most upregulated genes in HSC0. This cluster had significant negative enrichment for one GMP gene set, indicating downregulation of myeloid genes.

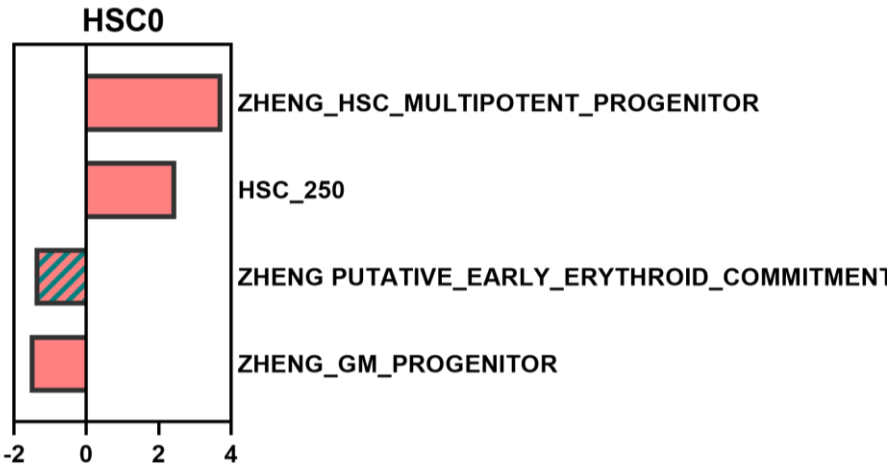

Figure 4.19. HSC0 population-specific enriched gene signatures

Normalised enrichment scores (NES) for significantly (adjusted p-value<0.05) enriched gene signatures in HSC0. Bar with blue fill represents a signature with adjusted p-value=0.085).

Similarly to HSC0, HSC1 was also significantly enriched for two separate HSC signatures, but also for an LMPP signature (Fig. 4.20). More specifically, the authors of the source of this gene set showed that cord blood LMPPs, characterised as $CD34^{+}/CD38^{-}/CD45RA^{+}$ could be subdivided into two populations based on CD52 and CSF3R expression which correspond to a lymphoid-primed and myeloid-primed LMPP, respectively (Zheng et al., 2018). HSC1 is enriched for the LMPP signature associated with stronger lymphoid priming. In accordance with the LMPP-like state of this cluster, one of the negatively enriched signatures is of MEPs, suggesting these cells may be losing their megakaryocyte-erythroid potential. In addition, the neutrophil-associated gene module, TRIM28_YBX1, is also negatively enriched in this cluster, which is supportive of the rest of the analysis, asserting the enrichment in this cluster for lymphoid gene signatures (Velten et al., 2017).

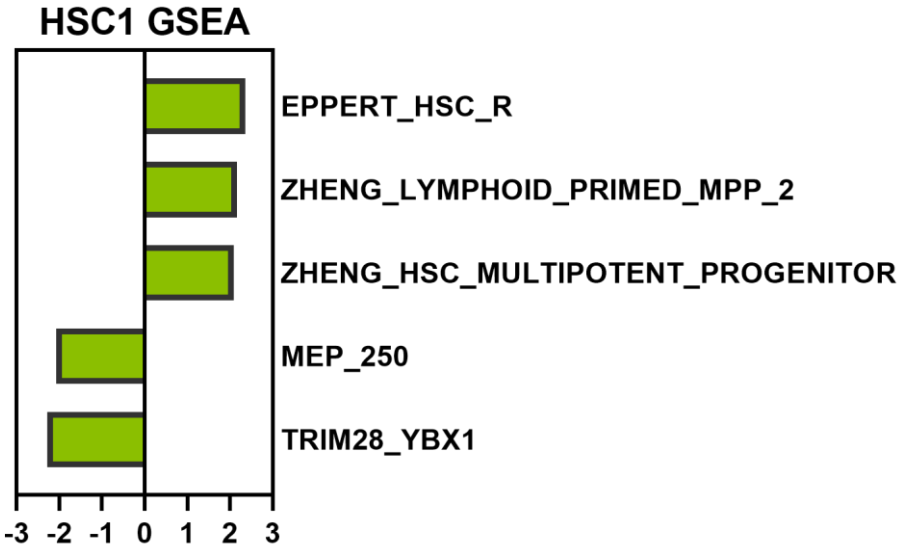

Figure 4.20. HSC1 population-specific enriched gene signatures

Normalised enrichment scores (NES) for significantly (adjusted p-value<0.05) enriched gene signatures in HSC1.

HSC2 was significantly enriched only for MEP, early erythroid and megakaryocyte gene signatures. A number of genes associated with the lymphoid and myeloid lineages were downregulated, including LMPPs, MLPs, early thymic precursors (ETPs) as well as GMPs. Because the differentially expressed genes for each cluster are computed against the rest of the clusters, the presence of so many negatively enriched lymphoid signatures, implies these are positively enriched in the rest of the data. There was also strong negative enrichment for HSC gene signatures, suggesting this cluster is largely comprised of more committed cells.

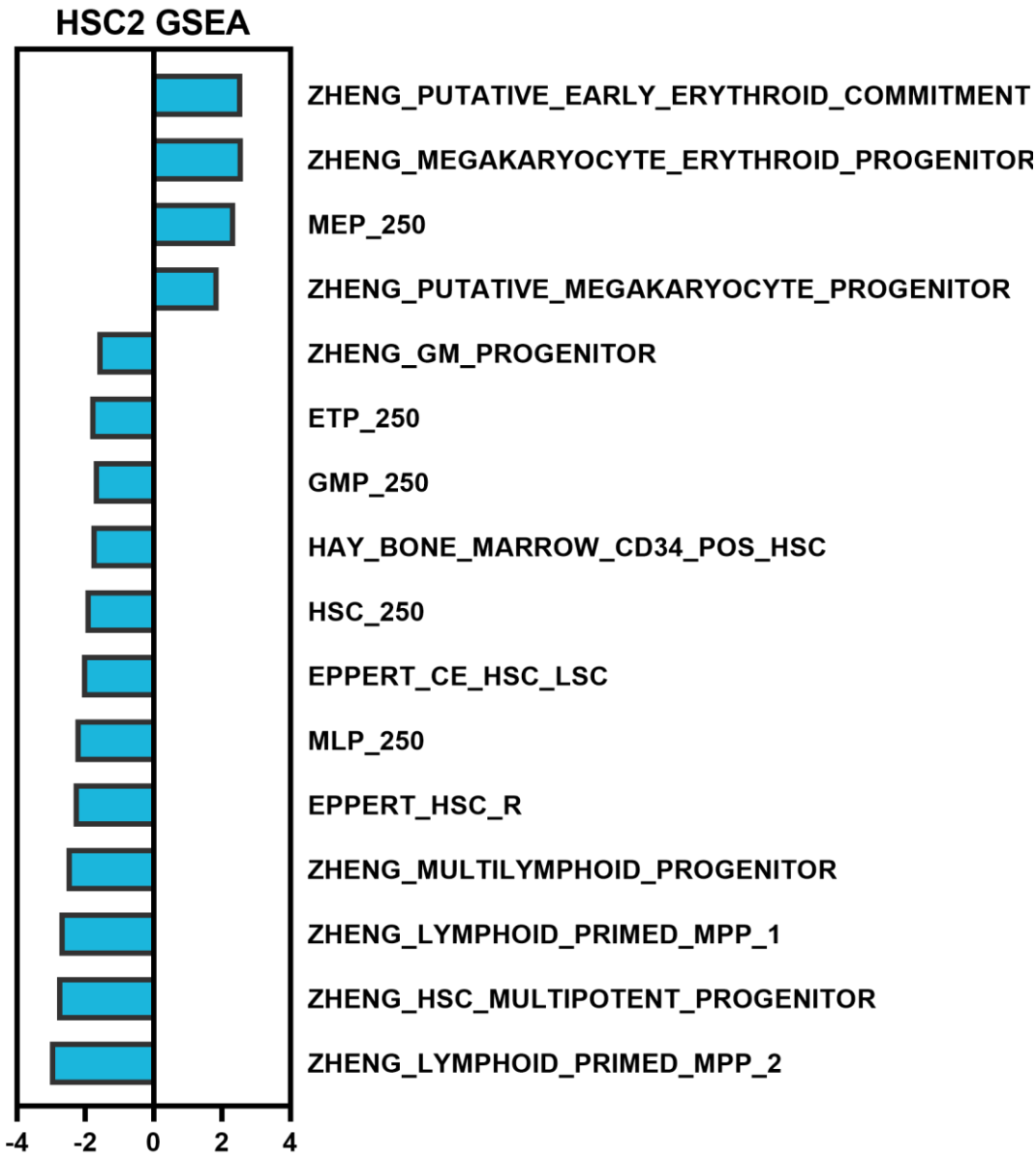

Figure 4.21. HSC2 population-specific enriched gene signatures

Normalised enrichment scores (NES) for significantly (adjusted p-value<0.05) enriched gene signatures in HSC2.

HSC2 was also positively enriched for a number of megakaryocyte/erythroid-associated gene modules (Fig. 4.22) (Velten et al., 2017). Two of these signatures, GMPR_MINPP and NFE2_GATA2 are associated with MEPs, while CBFA2T3_STAT3 and GP1BB_PBX1 are megakaryocyte specific. While we found no enrichment for erythroid-restricted gene modules, we have observed a number of erythroid genes in this cluster as well. The negatively enriched gene signature, mirrored the population specific GSEA in consisting of lymphoid (EBF1_ID3 and

FLT3_SATB1) and HSC (HLF_ZFP36L2) signatures. While the downregulated gene expression profile of this population appeared enriched for some GMP gene signatures, no CMP signatures or myeloid gene modules were significantly downregulated. Therefore, the myeloid potential of this population remains to be addressed.

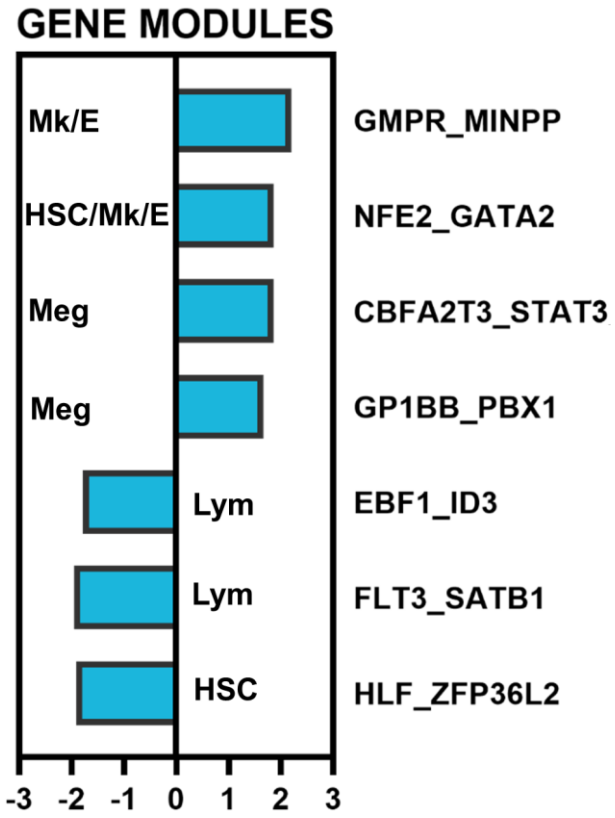

Figure 4.22. HSC2 gene module-specific enriched signatures

Normalised enrichment scores (NES) for significantly (adjusted p-value<0.05) enriched gene signatures in HSC2.

The positively enriched gene signatures in HSC3 were predominantly related to the lymphoid lineage (Fig. 4.23). One lymphoid-restricted gene module, FLT3_SATB1, was also enriched among the upregulated genes of HSC3. Another gene set featuring in the positively enriched list was the ZHENG_GM_PROGENITOR, bringing the HSC3 gene profile closer to LMPPs, which are capable of generating lymphocytes, granulocytes, monocytes and NK cells. The negatively enriched gene sets were related to the erythroid, myeloid

progenitors and HSC populations. An MEP signature was also negatively enriched but not significant.

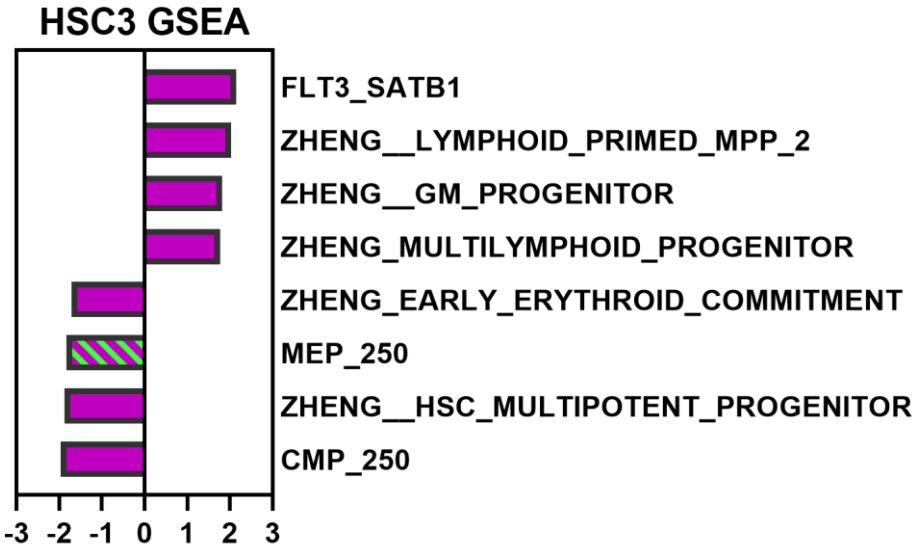

Figure 4.23. HSC3 population-specific enriched signatures

Normalised enrichment scores (NES) for significantly (adjusted p-value<0.05) enriched gene signatures in HSC3. Bar with green fill represents a signature with adjusted p-value=0.08.

Through GSE analysis, we have reasserted the gene expression profiles of HSC2 and HSC3 are respectively enriched for Mk/E and lymphoid transcriptional programmes. HSC2 was positively enriched for multiple population-level and gene-module signatures related to the megakaryocyte and erythroid commitment. HSC3 had positive enrichment in LMPP, MLP and the lymphoid FLT3_SATB1 gene module. Both HSC2 and HSC3 were negatively enriched for HSC gene signatures, suggesting they are more mature than HSC0 and HSC1.

HSC1 was enriched for LMPP as well as HSC gene expression programmes. This suggests a level of lymphoid priming in this subset. On the other hand, HSC0 had no positively enriched programmes apart from ones related to HSCs. With no other lineage-associated genetic programmes prevailing in this subset, it is possible these cells are most immature in the dataset.

To elaborate the differences between HSC0 and HSC1 we conducted pre-ranked GSE analysis on the significant differentially expressed genes between the two clusters. HSC0 was significantly enriched for one HSC signature confirming it as the cluster with the stronger HSC gene expression profile (Fig 4.24). HSC0 was also

enriched for an LMPP signature which also co-expresses neutrophil/monocyte gene programmes (Zheng et al., 2018). HSC1 was enriched for GMP and early erythroid gene signatures, which falls in agreement with earlier observations that this cluster expresses a mixture of genes associated with erythroid and myeloid lineages such as CEBPB, ZEB2 and KLF2. The GSE analysis suggested that the LMPP signature expressing higher lymphoid gene modules was also enriched in HSC1 but this enrichment was not significant (p=0.11).

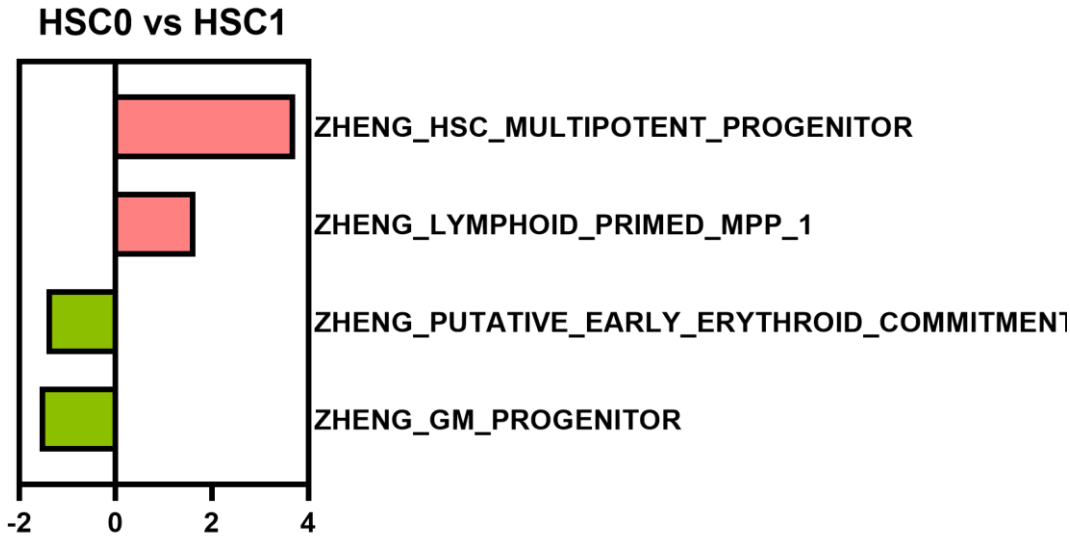

Figure 4.24. HSC0 vs HSC1 population-specific enriched signatures

Normalised enrichment scores (NES) for significantly (adjusted p-value<0.05) enriched gene signatures in differentially expressed genes between HSC0 and HSC1. Positive NES represent enrichment in HSC0 (coral bars) and negative NES represent HSC1 enrichment (green bars).

HSC0 and HSC1 both appeared to be expressing HSC and lymphoid primed gene programmes. While HSC0 has a less dynamic transcriptional profile, HSC1 expresses some myeloid and erythroid gene signatures as well.

4.6 Integration of $CD34^+$ and $CD34^-$ HSC single cell RNAseq datasets

As we have established that the $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ population contains subsets of cells enriched for HSC genetic signatures characterised in the $CD34^+$ population, the next question we wanted to answer is how $CD34^+$ and $CD34^-$

HSCs differ. To achieve this, we integrated a population-specific $CD34^+$ HSC dataset with our own. We chose this dataset because it was conducted on two sorted populations of $CD34^+/CD38^-/CD49f^+$ HSCs, defined as either $CD34^{Low}/CLEC9A^+$ or $CD34^{High}/CLEC9A^-$ (Belluschi et al., 2018). The dataset was used to show myelo-lymphoid lineage restriction occurs in the $CD34^+/CD38\text{-}/CD49f^+/CD34^{High}/CLEC9A^-$ population, while the $CD34^+/CD38\text{-}/CD49f^+/CD34^{Low}/CLEC9A^+$ fraction was more immature (Belluschi et al., 2018).

We started by conducting our own analysis of the published data, using our standard pipeline and obtained two distinct clusters, which is in concordance with the original publication of this data. We then plotted a heatmap of the top 25 expressed genes, finding 8 of them had also appeared among the top expressed genes in our own HSC subset (Fig. 4.25, Table 4.2). Five of these genes overlapped between our HSC1 cluster and the Belluschi 0 (B0) cluster.

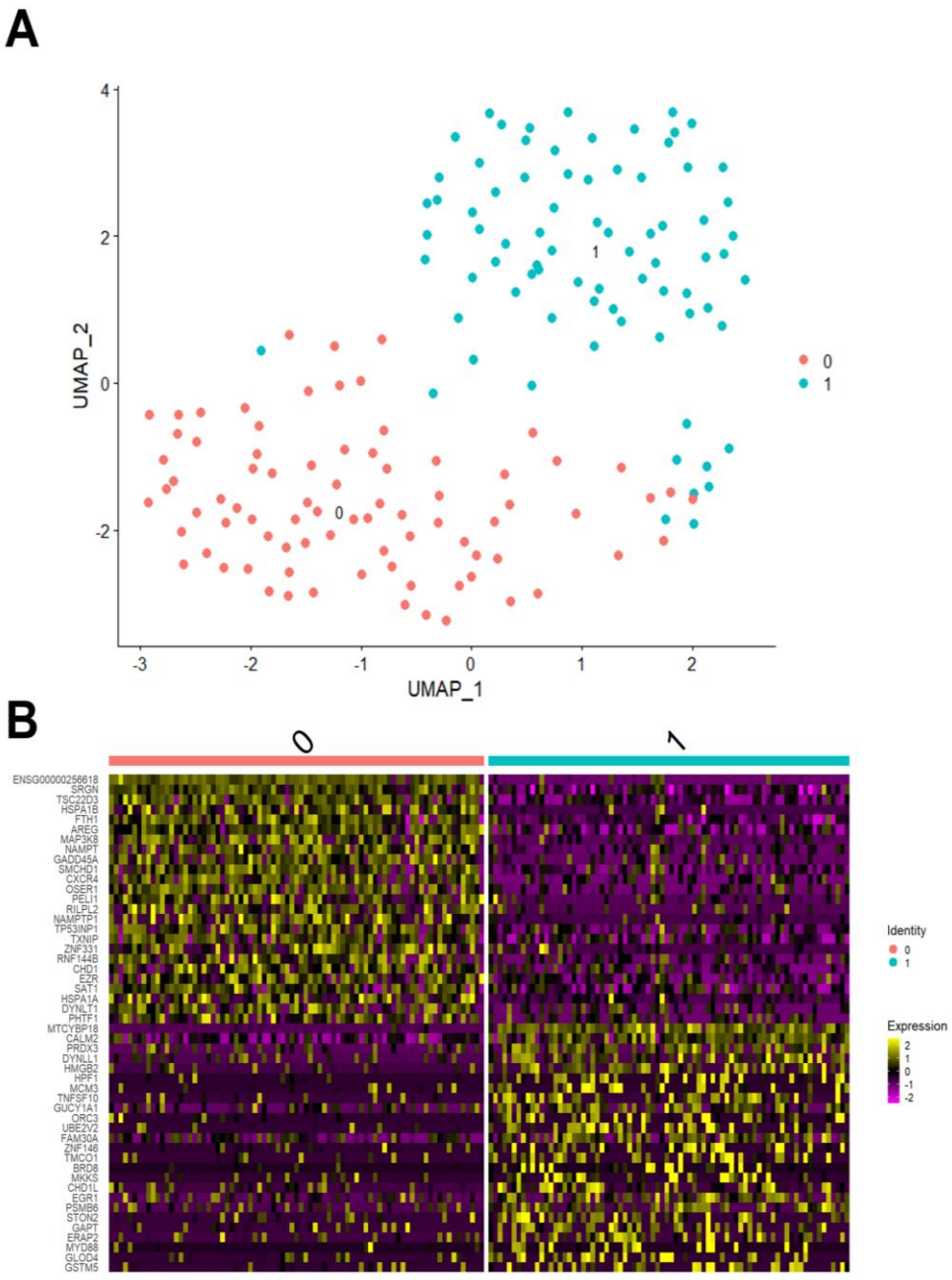

Figure 4.25. Belluschi dataset preliminary analysis

A. Dimensionality reduction (UMAP) plot with projected clusters.
B. Heatmap of the top 25 expressed genes across the two clusters.

GENE	Belluschi Data Cluster	HSC Subset Cluster
FTH1	0	0
EZR	0	0
TP53INP1	0	1
AREG	0	1
GADD45A	0	1
FAM30A	0	1
MAP3K8	0	1
PHTF1	0	2

Table 4.2. Overlapping expression of the top 25 expressed genes between the Belluschi dataset and our HSC subset with corresponding clusters

To confirm the validity of our analysis pipeline, we also plotted a list of genes that had been identified Belluschi and colleagues as differentially expressed between the two clusters (Fig. 4.26). While our normalization methods differ, we observed similar expression pattern as reported in the publication (Belluschi et al., 2018). Taking together, the top 25 expressed genes and the differenatially expressed gene expression paterns (Fig. 4.25B & Fig. 4.26), we can confidently conclude that Belluschi cluster 0 and Belluschi cluster 1 (henceforth referred to as B0 and B1) correspond to the CD49f Subset 1 and CD49f Subset 2, respectively. CD49f subset 1 was defined as the $CD34^{Low}$/$CLEC9A^{+}$ fraction and was shown to have potent long-term multilineage *in vivo* engraftment capacity, while the CD49f Subset 2 ($CD34^{+}$/$CLEC9A^{-}$) had long-term engraftment capacity but was myelolyphoid lineage restricted (Belluschi et al., 2018). The fact that the B0 cluster shared some highly expressed genes with our HSC1 cluster suggested that there may be some homology between these two populations. We proceeded to integrate our two datasets in order to explore both similarities and differences between the two datasets.

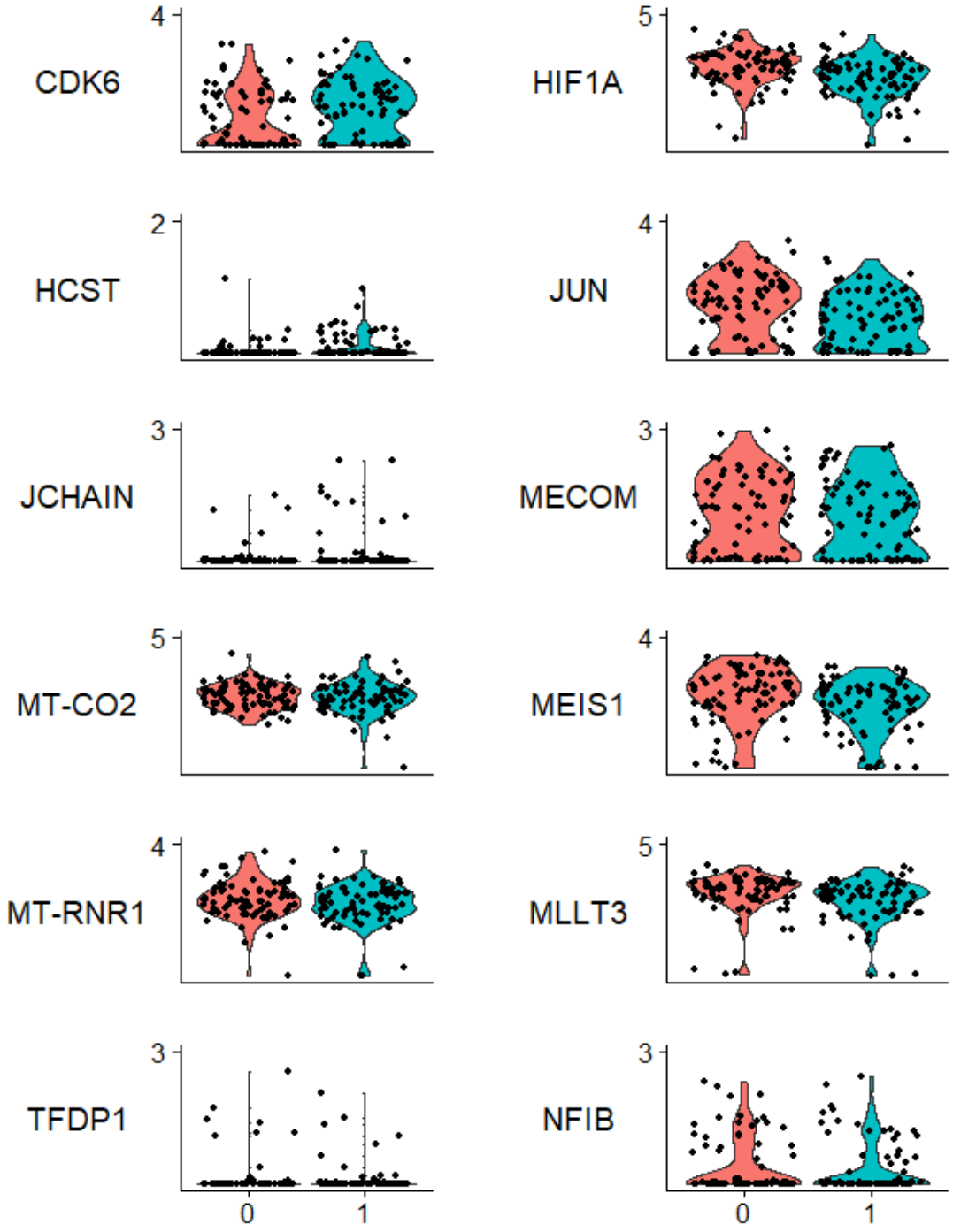

Figure 4.26. Expression of key genes in CD34^{+}/CD38^{-}/CD90^{+}/CD49f^{+} published dataset.

Violin plots for gene expression of genes reported to be expressed at varying levels by the two subpopulations defined as CD34^{Low}/CLEC9A^{+} (cluster 0, coral) and CD34^{High}/CLEC9A^{-} (cluster 1, turquoise) within the CD34^{+}/CD38^{-}/CD90^{+}/CD49f^{+} population.

We used pairwise anchor-based integration provided by the Seurat package to integrate our HSC subset with the Belluschi dataset. After scaling, we undertook dimensionality reduction, but didn't recalculate new clusters. Instead we projected the former HSC subset clusters onto the UMAP projection (Fig. 4.27A). When we plotted UMAP and coloured the cells by their dataset identity, the Belluschi cells were primarily divided between our former HSC0 and HSC1 clusters with few cells also mapping to the former HSC2 cluster (Fig. 4.27B). To determine how the two Belluschi clusters are mapping to our own data, we mapped the cluster number and dataset onto the UMAP projection. The two Belluschi clusters still separated with B0 mapping to HSC0, while B1 mapped to HSC1 (Fig. 4.27C). A small number of B0 and B1 cells were scattered into HSC1 and HSC3, respectively. This visualisation of the integrated data suggested that potential similarities between B1 and HSC1 as well as B0 and HSC0. However, B0 and B1 remained tightly associated together and didn't disperse into HSC0 and HSC1. So, we proceeded to recluster the data.

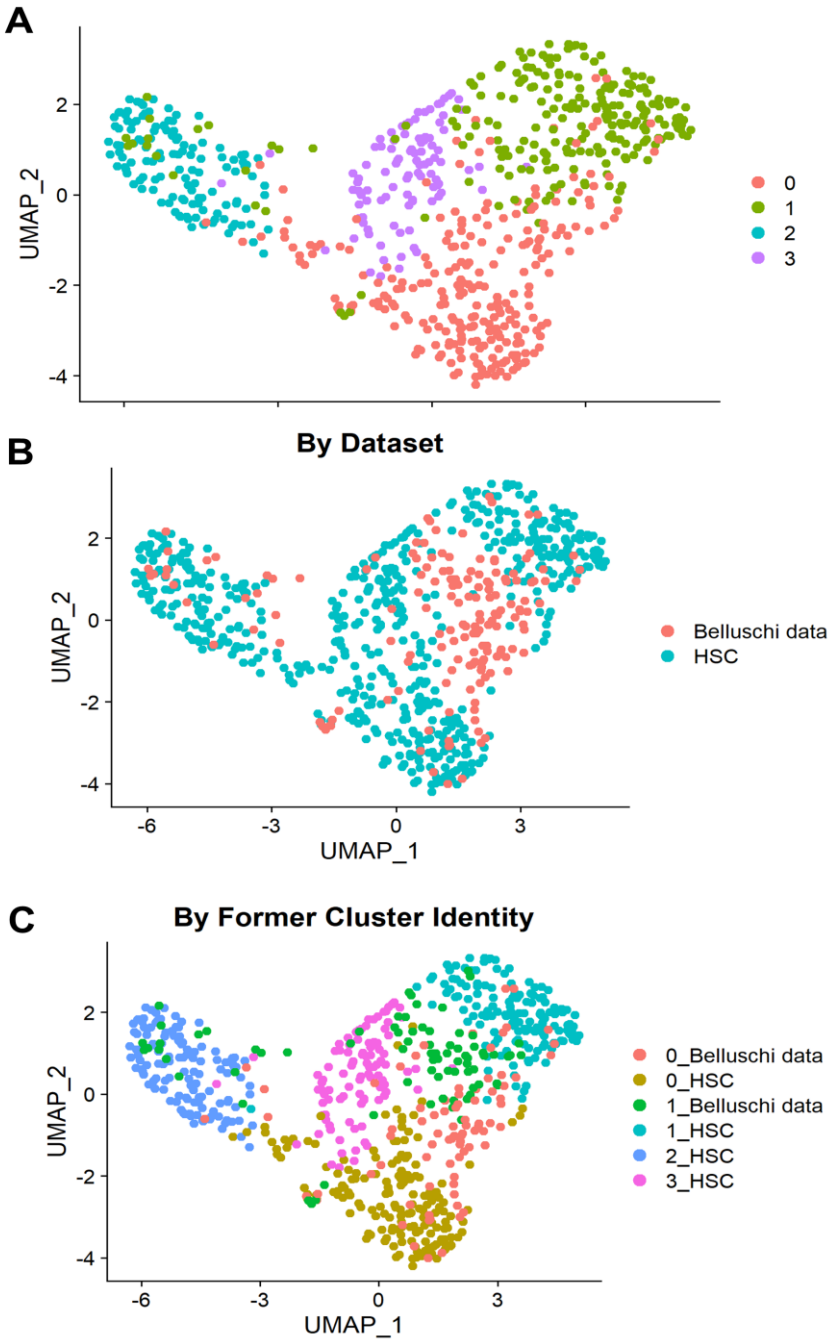

Figure 4.27. Dimensionality reduction of integrated datasets

A. UMAP projection of integrated datasets, where clusters represent former cluster assignements of the HSC subset and the cells from the Belluschi dataset have been assigned to HSC clusters.
B. UMAP projection of integrated datasets, where the two datasets have been mapped in coral (Belluschi data) and turquise (HSC subset).
C. UMAP projections of integrated dataset, where different cluster assignments have been mapped according to their cluster and dataset identity in their own individual analyses.

A total of 5 new clusters were computed for the integrated dataset. The new clusters maintained similar structures to the original HSC subset clustering, but notably some C1 cells that had co-localised with C2 (Fig. 4.27A) were reassigned to C2 in the clustered data (Fig. 4.28A). These cells originally belonged to B0 and B1 (Fig. 4.28C). A few cells that had been C0 but dispersed in the C1 cluster had also been reassigned to C1. An additional cluster, composed of only about 26 cells had also appeared in the reclustering. The original Belluschi clusters, B0 and B1, did not cluster together but remained separated between C0 and C1, asserting similarities in the gene expression profiles between HSC0 and B0 as well as HSC1 and B1.

When we looked at the cell composition of the different clusters, C0 consisted mostly of HSC0 and smaller portion of B0 cells (Fig. 4.28B). C1 was composed mostly of HSC1, but also relatively even proportions of B0 and B1 cells. C2 contained all the HSC2 cells which were identified as MEP-like with some B1 cells also present in this cluster. This was surprising to us as the B1 cluster has been defined in the original publication as having limited erythroid differentiation capacity. This assignment of cells could potentially be also explained by the fact that meyloid and erythroid gene expression signatures overlap. C3 was predominantly composed of HSC3 cells as well as B1 and to lesser degree B0 cells. The small C3 cluster was made up mostly of HSC0 cells.

We hypothesised HSC0 was possibly more immature than HSC1, and this is supported by a significant proportion of B0, an immature HSC population with multilineage capapcity, clustering with HSC0 in our integration analysis. In contrast B1, which is myelo-lymphoid lineage restricted was assigned together with HSC1, HSC2, and HSC3, the latter two being more mature clusters (Fig. 4.28B). In addition, the contribution B0 gradually diminished from C1 to C3, which can suggest that C1 and C3 are less enriched for immature HSC gene expression signatures.

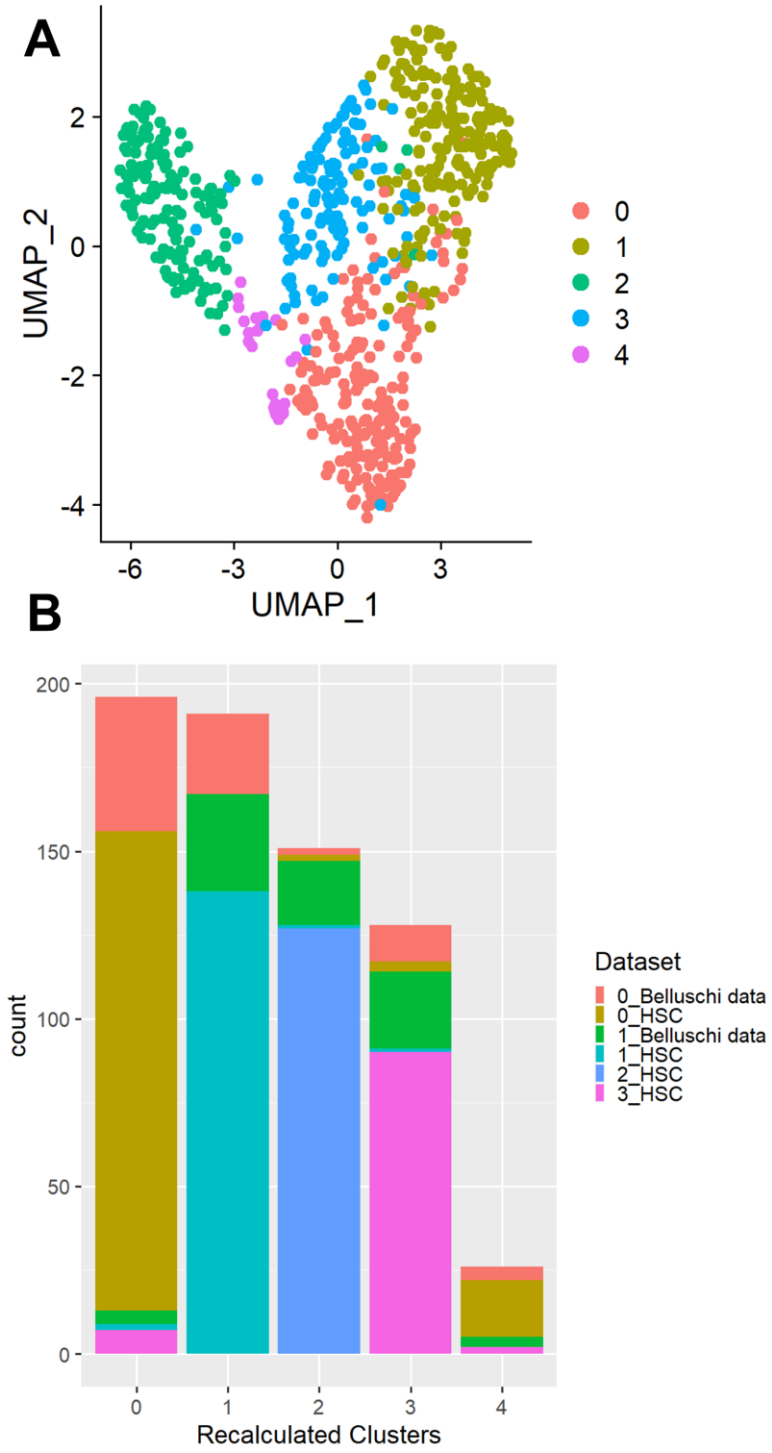

Figure 4.28. Reclustering of integrated datasets

A. UMAP projection of the integrated Belluschi dataset and the HSC subset with recomputed clusters.
B. Bar plot of the the compositions of the new clusters according to original dataset and cluster assignments.

By integrating our own HSC subset transcriptomic data with a published scRNAseq dataset of two sorted $CD34^{+}/CD38^{-}$ populations with robust repopulation and differentiation capacity, we have shown that gene expression similarities exist between the $CD34^{-}$ and $CD34^{+}$ HSCs. Additional information on the levels of maturity of the two populations within the published dataset has also enabled us to identify HSC0 as more immature cluster compared to HSC1. We have been unable to generate differentially expressed gene lists between the two datasets because of the fact these tests cannot be carried out on integrated data due to violation of the statistical assumptions these analyses rely on. But through the different distributions of the separated datasets across clusters, we have shown that the most immature B0 cluster has a number of cells that cluster together with HSC1 and HSC3, the latter of which we have identified as a more mature cluster. This may imply that even within this immature population another level of heterogeneity exists. For our own data, the integration also confirmed that HSC1 may share features with both published $CD34^{+}$ populations as it clustered together with cells originally assigned to both B0 and B1.

4.7 Chapter summary and discussion

In this chapter we have presented an analysis of single cell transcriptomics data of the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ population. Following QC, clustering and dimensionality reduction, we used known marker genes to assign identities to the different cell populations within our data. We identified a mature myeloid cluster (C6) expressing genes coding for cell surface markers such as CD14 and ANPEP as well as functional myeloid genes such as LYZ and myeloid transcription factors. We also looked in depth into a cluster that was partitioned in two distinct subclusters (C7). We showed that one these subclusters (C7.1) was expressing innate lymphoid cell gene signature, KLRB1, IL7R, GATA3 and RORA, while the other (C7.2) expressed genes such as GATA2, CD9 and HDC, which were indicative of mast cells. We also detected a maturing erythroid population (C5) that was expressing high levels of a number of globin genes. Two different clusters expressing megakaryocyte and erythroid genes were also identified (C3 and C4), with the distinction that C4 was highly enriched for S-phase genes, suggesting a maturing erythroid population. Expression of known HSC genes identified three distinct clusters (C0-2) as more

immature, which compelled us to subset these populations for downstream analysis. In this subset we also included C3, which in addition to megakaryocyte and erythroid genes, also promiscuously expressed a number of transcription factors associated with other haematopoietic lineages.

After isolating these four clusters, we rescaled and re-clustered the data, which resulted again in 4 new clusters. We identified a robust HSC signature in clusters HSC0 and HSC1. We also observed an MEP-like signature in HSC2. In addition, we also detected a number of lymphoid-associated transcription factors and genes expressed in HSC3. Some of these transcription factors were also expressed in HSC1 and to lesser degree HSC0. However, we were unable to detect expression of more mature lymphoid genes such as EBF1 and PAX5 or their targets, which could be indicative of early lymphoid priming events in these cells.

We next undertook gene ontology enrichment analysis using differentially expressed gene lists for each cluster. The defining features in HSC0 was a low number of upregulated genes, positive enrichment for mRNA catabolic processes and negative enrichment for oxidative phosphorylation, which were all indicative of a highly quiescent cell population with minimal cellular activity compared to the other clusters. HSC1 was enriched for pathways associated with myeloid and erythroid processes, and negatively enriched for alternative splicing processes. In HSC2, processes associated with erythroid and megakaryocyte cell maturation were enriched, such as protein folding and DNA synthesis, which also was suggestive of a more active cell state. Finally, HSC3 had positive enrichment for processes associated with leukocytes.

Our next step was to conduct pre-ranked GSEA using publicly available HSC datasets and population specific gene modules. Using this approach, we showed that HSC0 and HSC1 showed significant enrichment for available HSC gene sets, while the latter also showed enrichment for an LMPP gene set. HSC2 tightly associated with MEP gene sets and MEP gene modules. HSC3 was highly enriched for LMPP and MLP gene sets. In addition, both HSC2 and HSC3 had negative enrichment for HSC gene sets.

Finally, we integrated our HSC subsets with a published scRNAseq dataset of two sorted $CD34^{+}/CD38^{-}$ HSC populations. We showed that our HSC0 clustered predominantly with a population that was identified by the authors of the dataset as more immature. On the other hand, the more mature of the published populations

associated with HSC1, HSC2, and HSC3, which had also been identified as more mature in our earlier analysis. This integration analysis allowed us to visualise $CD34^+$ and $CD34^-$ HSC data together and reveal potential similarities between the two populations. We were interested into also identifying the differences between these two HSC populations, but differentially expressed gene lists can only be generated from raw normalized data. As the two datasets have different sequencing depth, such analysis would be skewed.

Unexpectedly, our scRNAseq revealed a high proportion of cells with HSC gene signatures present within the $CD34^-/CD38^-/CD93^+/CD117^{Hi}$. Furthermore, these cells with high expression of HSC-associated genes were subdivided into three distinct clusters suggesting a level of heterogeneity. While clustering algorithms can result in "over-clustering", we furthered the analysis by subsetting the suspected HSC clusters and conducting GO pathway and gene set enrichment analysis by using differentially expressed gene lists. Both analyses reasserted tangible differences between the different HSC clusters based on enriched biological processes and enrichment of different population-specific gene signatures. They also presented us with the evidence to identify two clusters as more mature (HSC2 and HSC3) compared to the more enriched for immature HSC signatures, HSC0 and HSC1. All these findings suggest high degree of HSC-level heterogeneity within the $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ population. Furthermore, increasingly maturing and cell-fate restricted gene signature enrichment identified in clusters HSC1, HSC2, and HSC3, could also be suggesting fate priming events or early cell fate restriction. More specifically, HSC0, HSC1, and HSC3 were expressing lymphoid transcription factors, while HSC3 also expressed higher levels of more mature lymphoid genes. To establish whether gradual lineage restriction is taking place within these three populations, we would have to carry out single-cell functionality assays and assess differentiation outputs. One of the features of lymphoid fate restriction is loss of erythroid and megakaryocyte potential, but $CD34^-/CD38^-/CD93^{Hi}$ cells have been shown to generate BFU-Es, *in vitro* indicating the presence of cells with erythroid differentiation capacity (Anjos-Afonso et al., 2013). These cells however could represent the MEP population we also identified. We never assayed mice repopulated with $CD34^-$ SRCs for human erythroid cells as human erythropoiesis is highly inefficient *in vivo.* This limitation can be circumvented in future experiments by administering mice with EPO post-transplantation (Belluschi et al., 2018).

Another surprise in our scRNAseq analysis was the substantial amount of cells that were sequenced and passed QC, expressed genes associated with HSCs. The original 3 clusters we identified as HSC-like (C0-C2) came to a total of 410 cells, while the whole dataset consisted of 697 cells. Our *in vivo* repopulation experiments estimated the SRC frequency of the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ population at 1/52 (19%). Our scRNAseq data estimates the fraction of cells with stem cell gene signature at almost 59%, which excludes the MEP-like population we included in the HSC-like subset. While technical steps required for the preparation of cell suspensions for IV injections could result in some cell loss, this cannot fully account for the discrepancy. Another possible explanation for the high number of cells expressing HSC gene signatures but low estimated HSC frequency could be that all of these cell populations could represent true HSCs, but fail to home and establish a niche in the murine bone marrow. This is supported by the fact that the NSG mouse model only supports low levels of engraftment of this population unless cytokines such as SCF and G-CSF are administered to the mouse recipients post-transplantation. While we have shown that the $C\text{-}Kit^{Mutant}$ and NSG-S mouse models are more permissive to $CD34^{-}$ SRC engraftment, it is possible that they also can support only some of the $CD34^{-}$ HSC populations. Another possibility is that some cells within the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi}$ population have more short-term repopulation capacity, which peaks soon after transplantation and quickly declines before our first bone marrow aspiration at 6 weeks. However, in previous repopulation studies of the $CD34^{-}/CD38^{-}/CD93^{Hi}$ population, we did not observe at three weeks post-transplantation any repopulation, even when the whole bone marrow was analysed rather than just sampled by bone marrow aspiration (Anjos-Afonso et al., 2013). This analysis was only conducted in NSG with administration of G-CSF and SCF, not in $C\text{-}Kit^{Mutant}$ or the NSG-S mouse strains.

Another curious finding was the presence of multiple populations with megakaryocyte, erythroid and mature erythroid gene expression patterns. One of these clusters, we included in our HSC subset analysis, HSC2, was of interest, because MEP populations have been previously characterised as expressing both CD34 and CD38 (Doulatov et al., 2010; Edvardsson et al., 2006). This population, however, lacked expression of both genes, while also maintaining expression of all the heptad transcription factors. It was also largely devoid of a more mature lymphoid gene signature. The differentially expressed genes in HSC2 tightly associated with

both the erythroid and megakaryocyte cell fates, suggesting this cluster is more likely to represent a bipotent MEP population rather than multipotent HSCs. While it is possible that this might be a contaminating population that expresses CD34 and CD38, this is not supported by the gene expression data, where HSC2 expressed the lowest levels of CD34 in the entire HSC subset, and CD38 is not expressed at all in the entire dataset.

On the topic of CD34, we found that this gene was actually expressed at low level in HSC0, HSC1, and HSC3. A potential explanation for this could be that the CD34 gene is transcribed, but not translated, which could indeed be happening in a cluster like HSC0, where mRNA catabolic processes were positively enriched. CD34 has also been reported to be internalised by endocytosis (Krauter et al., 2001). Depending on the CD34 protein turnover following internalisation, intracellular CD34 expression can be checked by immunofluorescence staining. A third option is that CD34 is indeed expressed at low levels by the population we characterise as $CD34^-$. $CD34^{Low}$ human HSCs have been reported before, however, relative to the $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ population we sort, $CD34^{Low}$ cells have a significantly higher expression and fall within a gating strategy traditionally used for isolating $CD34^+/CD38^-$ cells (Fig. 4.29D). By back-gating of the $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ population we observe that these cells are mostly scattered along the positive CD34 edge of the $CD34^-/CD38^-$ gate (Fig. 4.29D).

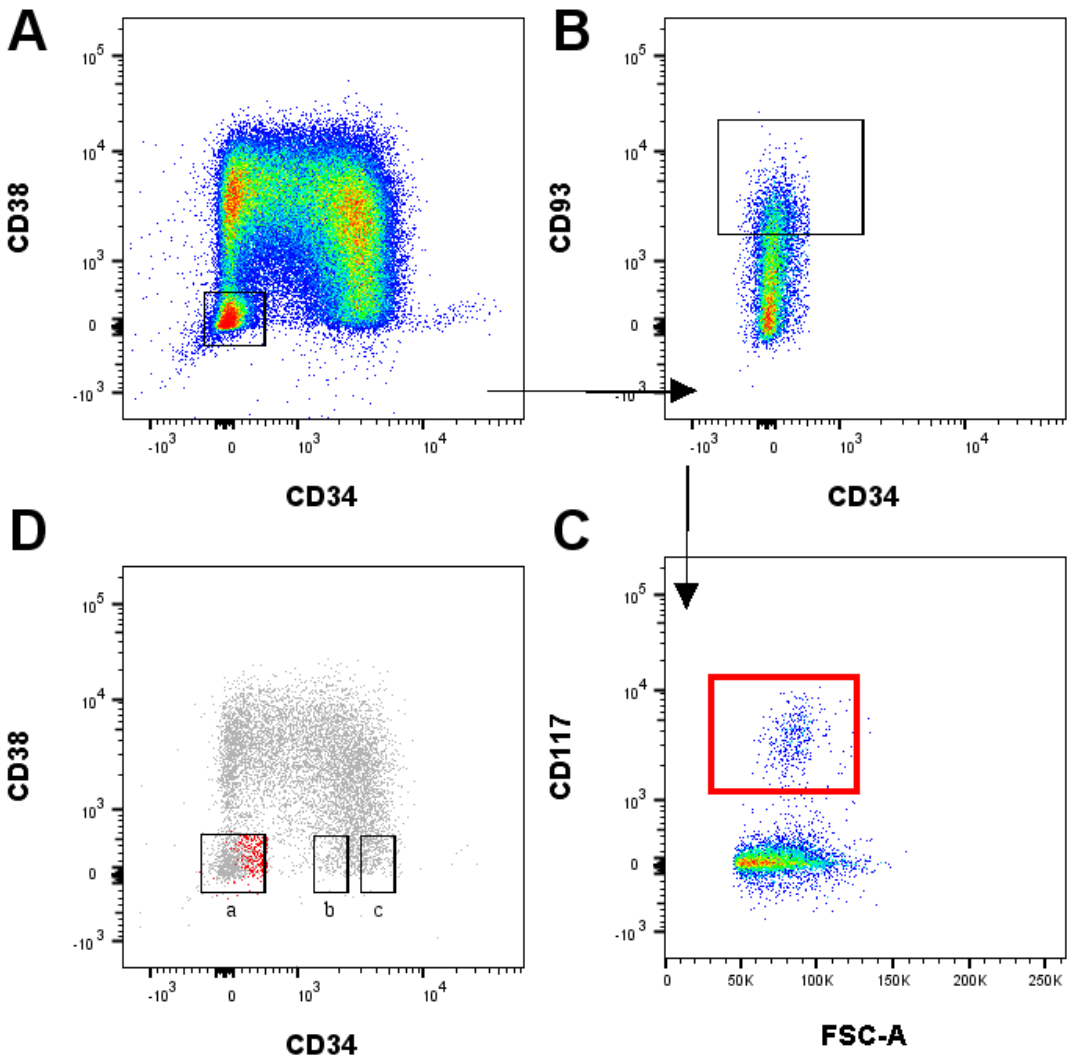

Figure 4.29. Back-gating of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} population

A-C. Gating strategy for CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} cells.
D. Back-gated CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} cells onto the CD34 vs CD38 plot (**a** gate, red points). Gates **b** and **c** represent gating for the CD34^{Low} and CD34^{+} HSCs, respectively.

Chapter 5. Prenatal haematopoietic stem cell heterogeneity

5.1 Introduction

Our understanding of definitive HSCs and progenitors in foetal haematopoiesis has been mainly advanced by studies in mouse, which have been extrapolated to the human system. But some major differences between the genesis of the haematopoietic system between the two species have also been identified.

The primitive wave of human haematopoiesis originates in the yolk sac, which differs structurally between mice and humans. The major yolk sac haematopoietic output at Carnegie Stage (CS) 7-8 is nucleated erythrocytes and to lesser degree macrophages and megakaryocytes (Fukuda, 1973; Takashina, 1987; Wang et al., 2020). These cells can be detected in the embryo proper from CS 10 onwards (Tavian, Hallais, & Péault, 1999). Definitive HSCs emerge from the aorta-gonad-mesonephros (AGM) region of the dorsal aorta. HSCs arise from intra-aortic haematopoietic clusters and express KDR (VEGF receptor 2) and CD34 (Cortés et al., 1999; Tavian et al., 1996, 2001; Tavian, Hallais, & Péault, 1999). Transplantations of cells from different haematopoietic regions derived from CS 12-17 human embryos has shown that multipotent repopulating HSC are primarily detected in the AGM region starting from CS 14 (Ivanovs et al., 2011). These repopulating cells are CD34, CD144 (VE-cadherin), and CD45 positive, and also $CD45RA^{-}/CD38^{Low/-}$ (Ivanovs, Rybtsov, Anderson, Turner, et al., 2014; North et al., 2002). Yolk sac HSCs were only detected after CS 17, and repopulation of mouse recipients tended to be poorer and rarer with yolk sac cells. Repopulating foetal liver (FL) HSCs are detected later, 7-8 post-conception weeks (pcw) and still express VE-cadherin (CD144) (Ivanovs et al., 2011; Oberlin et al., 2010).

Foetal liver HSCs are less well defined. Similarly to adult and UCB-derived HSCs, the $CD34^{+}/CD38^{-}/CD90^{+}$ fraction has been shown to be enriched for repopulating cells (Doulatov et al., 2012; R. Majeti et al., 2007). Over the years various markers have been reported as potentially useful in the positive enrichment of human FL HSCs such as CD143, GPI-80, EPCR, and CD117 (Calvanese et al., 2014; Jokubaitis et al., 2008; Maillard et al., 2020; Subramaniam et al., 2019).

Other purifying markers have only been reported in murine foetal haematopoiesis such as Mpl (CD110) (Petit Cocault et al., 2016).

Although compelling evidence of the stem cell heterogeneity in foetal haematopoiesis exists, and scRNAseq and mass cytometry data of foetal haematopoietic tissues is quickly becoming available, still little is known about the cell surface markers phenotype of human foetal HSCs. While several potential purifying markers have been reported, most of them have been tested at isolated developmental stages and mostly on FL-derived HSCs. How the expression of these markers changes with gestational age and colonisation of the bone marrow niches has not yet been explored.

In this chapter, we report the design and optimisation of an 18-colour flow cytometry panel, aiming to assay a list of HSC markers in late first trimester and second trimester foetal liver (FL) and foetal bone marrow (FBM). We are also presented multiparameter analysis of the data that helped us identify a population enriched for cells capable of multilineage repopulation in C-Kit^{Mutant} recipients.

5.2 Flow cytometry antibody panel optimisation

Our first step was to compile a list of the HSC cell surface markers we wanted to investigate in early haematopoietic development. We started by including markers of general interest and markers that have been reported to be common between adult and foetal haematopoiesis such as CD45, CD34, CD38, CD45RA, CD90 and GPI-80. We also included markers that are of personal research interest such as CD93 and CD117. Earlier in this thesis, we described a population of CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} HSCs in postnatal haematopoiesis and were interested in investigating whether a population with similar phenotype exists in early development. CD49f has been shown to purify umbilical cord blood HSCs, but whether it is also expressed in early development remains a question. In the differentiation of HSCs from human embryonic stem cells (hESCs), CD43 is an early HSC marker that precedes CD45 appearance (Vodyanik et al., 2006). The rest of the markers, CD143, CD144, CD110, EPCR and CD7 have been previously reported either in human or murine foetal haematopoiesis. We identified all the relevant

reports and picked our antibody clones in accordance with published methodologies (Table 5.1).

Antigen (Clone)	Function
CD45 (HI30)	Haematopoietic identity marker
CD143 (BB9)	Angiotensin-converting enzyme, meadiates production of angiotensin II
GPI-80 (3H9)	Vannin-2, a mediator of transendothelial migration and neutrophil adhesion
CD110 (1.6.1)	Mpl, receptor for thrombopoietin
EPCR (RCR-252)	Endothelial protein C receptor, role in anticoagulation
CD41 (HIP8)	Integrin alpha chain B, cell-to-cell signalling
CD43 (IG10)	Leukosialin
CD144 (BV9)	VE-cadherin, cell adhesion molecule
CD117 (104D2)	Receptor for SCF
CD93 (VIMD2)	Complement receptor
CD38 (HB7)	Cyclic ADP ribose hydrolase
CD34 (G812)	Glycoprotein and a cell-cell adhesion molecule
CD90 (5E10)	Glycophosphatidylinositol, adult stem cell marker
CD49f (GOH3)	An integrin, adult stem cell marker
CD45RA	CD45 isoform associated with more mature cell types
CD7	Naïve T cell marker

Table 5.1. List of foetal HSC cell surface markers, corresponding antibody clones and respective functions.

We also required an exhaustive list of lineage-committed cell surface markers in order ensure that we would be eliminating the majority of lineage-positive cells in our analyses. Our finalized list included standard lineage cocktail (Table 5.2). In addition, HSCs in the foetal liver have been shown to associate with blood vessels, and endothelial cells can express CD31 as well as a number of other markers, which are present in our main panel (e.g. EPCR, CD144 and CD143) (Table 5.2). Therefore, we also included CD31 in the lineage panel. We didn't include any

hepatocyte markers as one of our processing steps of FL samples includes density gradient centrifugation, where hepatocytes are excluded.

Marker	Cell type
CD2	T cells/NK
CD3	T cells
CD10	NK cells
CD11b	Myeloid
CD14	Macrophages/Monocytes
CD16	Gran./Mono./NK/Macr.
CD19	B cells
CD20	B cells
CD24	B cells/Granulocytes
CD31	Endothelial
CD56	NK cells
CD66b	Granulocytes
CD235a	Erythrocytes

Table 5.2. List of lineage cocktail antibody panel.

After selecting our markers of interest, our next step was to allocate the fluorochrome conjugates to the different antibodies. For some we were restrained by commercial availability. For example, we were compelled to assign PE to GPI-80 as we wanted to use the same format of this antibody as it has been reported in the literature. For the rest of the antibodies, we were faced with the challenge of having to assign different fluorochromes with limited understanding of the antigen abundance across FL and FBM haematopoietic cells. This is potentially confounding as there was the issue that if a dimmer fluorochrome conjugate was assigned to an antibody recognizing a rare cell surface marker, it could be impossible to distinguish the positive signal. This is also one of the major challenges of flow cytometry – revealing dim events. Therefore, we undertook quantification of the average abundance of molecules present per cell for each marker in FL. As little information was available for how cell surface marker expression may change in cells across different developmental stages, we picked FL samples from three stages, 6, 12, and 18 post-conception weeks (pcw).

To quantify cell surface marker abundance per cell in these tissues, we took each sample and stained it individually with a PE conjugated antibody against each of our HSC and lineage markers of interest. PE (R-phycoerythrin) is a unique

fluorophore, not only because is bright, but also because of its large size which only allows one PE molecule to be bound to a single antibody. When this feature of PE antibodies is used together with BD Quantibrite™ Beads, which can be used to calculate a linear regression for fluorescence intensity against PE molecules per bead, we can accurately estimate the average abundance of bound antibodies per cell. We stained the three FL samples for each PE antibody and calculated the average number of bound antibodies per cell. We found that the abundance of different antibodies tended to follow the same pattern in the three different stages of FL. One marker, CD235a (Glycophorin A) was significantly more abundant on the cell surface of 6 pcw FL. This could be possibly explained by higher presence of mature erythrocytes in this sample due to incomplete red cell blood lysis during the initial sample processing. Otherwise, the average bound antibodies per cell were similar between samples. CD43 was the most intensely expressed marker followed by CD31, CD45, and CD38. The majority of HSC markers, CD144, CD143, CD90, CD117, & CD93 had an overall lower average marker density per cell. The average marker abundance per cell varied between lineage markers with CD31 and CD24 being relatively abundant, while CD66b and CD3 had very low average numbers of bound antibodies per cell.

Using the average abundance of different markers, we assigned fluorochromes to different antibodies, by picking brighter dyes for markers with lower expression intensity and dimmer dyes for the more highly expressed antigens (Table 5.3). We assigned FITC to all lineage markers as it is a moderately bright fluorochrome and available for all the lineage markers in our panel. We proceeded with assigning bright fluorochromes such as BV786, BV711, BV421, PEcy7 and PE-CF594 to the cell surface markers with the lowest antigen density per cell. We also assigned dim fluorochormes (APCH7, APC700, BUV395, and BUV805) to the most abundant markers. Due do commercial availability, we had to assign some bright fluorochromes to relatively abundant markers such as CD7 and CD38. As our CD93 staining protocol relies on a biotinylated antibody, we assigned the secondary antibody to a moderately bright fluorochrome, BUV496. To minimise occupation of other channels, we decided to use Sytox Green as a Live/Dead dye, which has similar emission spectrum as FITC.

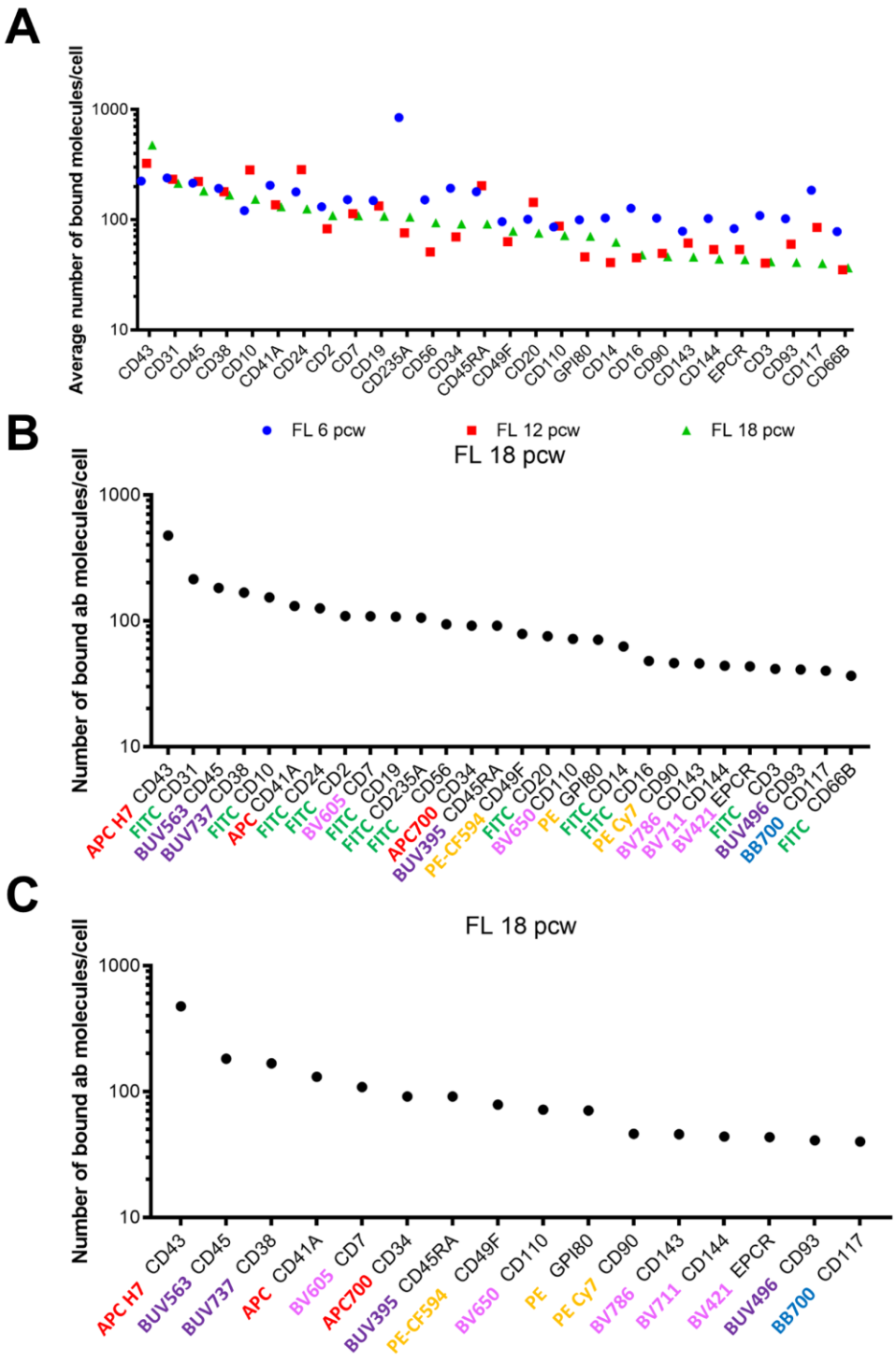

Figure 5.1. Quantification of bound antibodies per cell for HSC and lineage markers

A. Quantification of average number of bound antibodies per cell for the HSC and lineage markers of interest in FL samples at stages 6, 12, and 18 pcw.
B-C. Quantification of bound antibodies per cell for both HSC and lineage markers (B) and HSC markers only (C) and respective fluorochrome assignments.

Laser	Very Bright	Bright	Moderate	Dim
Ultraviolet (355 nm)		BUV737	BUV395 BUV496	BUV805
Violet (405 nm)	BV421 BV650 BV711	BV605 BV786		
Blue (488 nm)	BB700		FITC	
Yellow/Green (561 nm)	PE PE-CF594 PE.cy7			
Red (640 nm)		APC		APCR700 APCH7

Table 5.3. Fluorochromes grouped according to respective excitation lasers and relative brightness (adapted from BD website).

After we finalised the antibody panel, we proceeded to titrate all of the antibodies. This is a necessary step in the use of any new antibody as overstaining can lead to negative cells appearing positive. This has negative impacts on resolution and can also increase spillover into other channels. Proper titration is also highly important in large panels, because it can help reduce the staining volume as large staining volumes can result in poorer cell staining. We used cord blood mononuclear cells for the titrations and tested each antibody at volumes of 1, 2, 3, 4 and 5 µl in a staining volume of 100 µl and stained $10x10^6$ cells per titration. We stained cells for 30 minutes at 4°C. We excluded dead cells and doublets before concatenating data for each antibody to visualise the best staining. For a number of antibodies, we observed two distinct populations of positives and negatives (e.g. CD45RA, CD93, CD7, CD45, CD38, CD41, CD34, CD43, CD49f and GPI-80) (Fig. 5.2). We also computed titre curves based on staining index, but staining indexes were fairly similar for most of the antibodies.

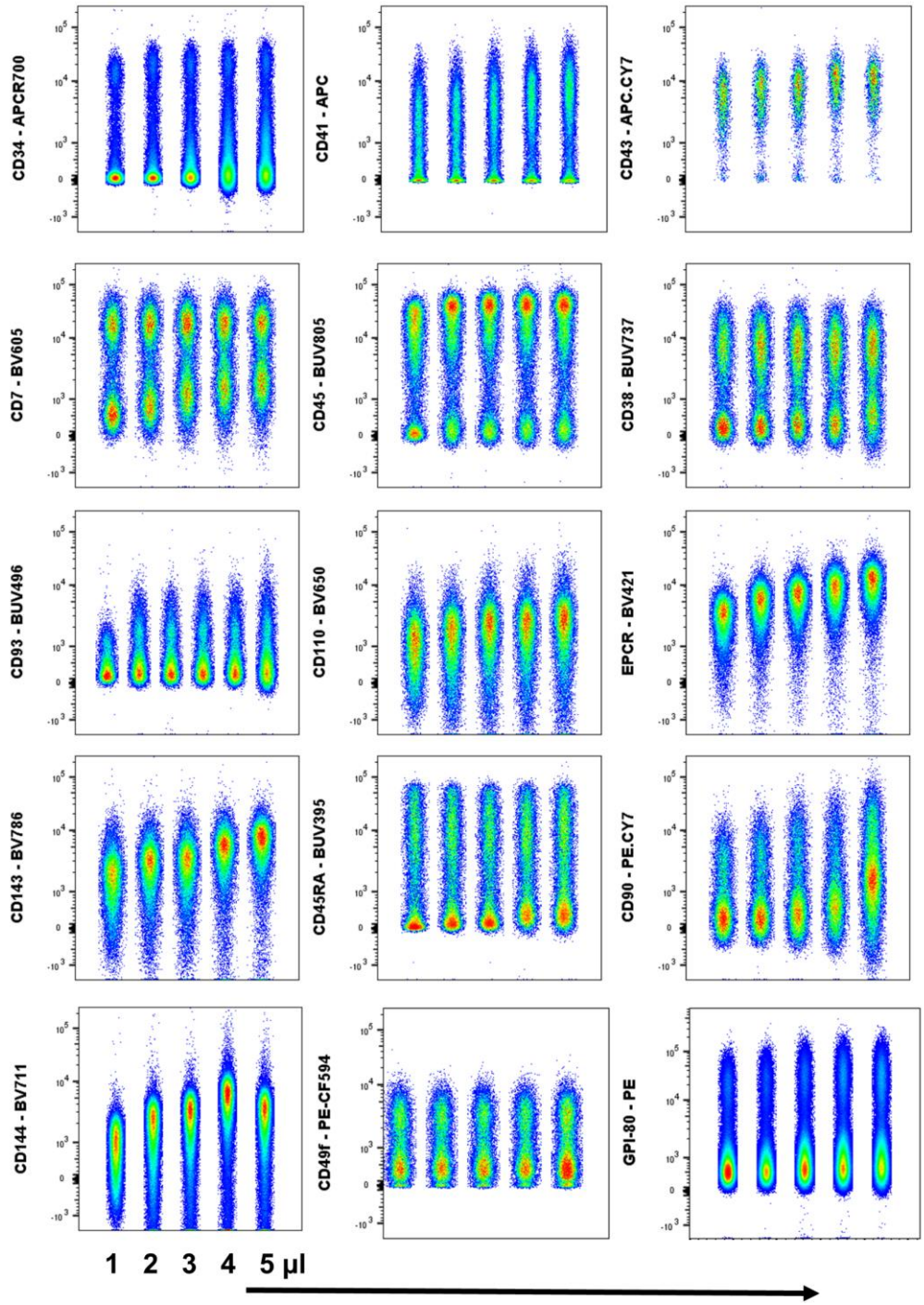

Figure 5.2. Titrations for HSC antibodies.

Concatenated flow plots of titrations with different antibody volumes. Prior to concatenation files were gated for single, live and Lin^- cells with the exception of GPI-80, where Lin^+ cells were included to include higher numbers of positive cells.

A few antibodies such as CD110, CD143, and EPCR appeared to be staining a fair amount of cells, even at the lowest antibody volume, which we found unrealistic and possibly the result of overstaining as these antigens are fairly rare in cord blood

mononuclear cells (Fig. 5.2). Therefore, we repeated these titrations with lower antibody volumes on the same type of cells. (Fig. 5.3). For all antibodies, we again observed overstaining at volumes of 1 µl, and for CD143 at 0.5 µl as well. We determined that overstaining was no longer a problem if we used volume of 0.25 µl for CD143 and 0.5 µl for EPCR, CD144, CD110 and CD117.

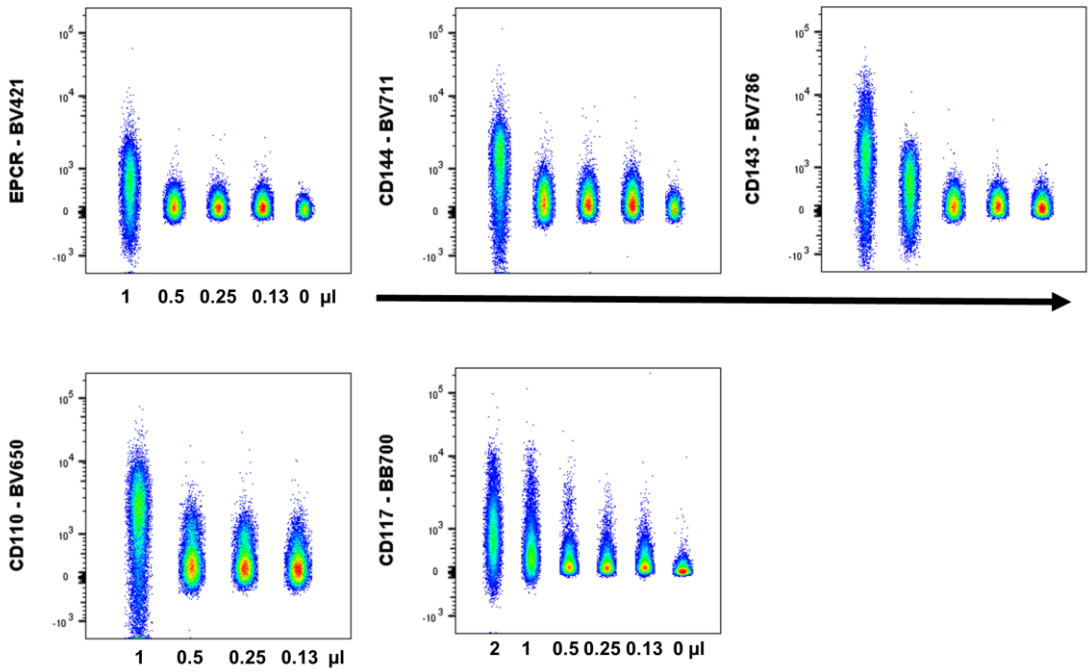

Figure 5.3. Low volume antibody titrations for EPCR, CD144, CD143, CD110 and CD117

Concatenated flow plots of titrations with respective antibody volumes. Prior to concatenation individual samples were gated for single, live and Lin^- cells. Antibody volumes used are 2 µl (CD117 only), 1 µl, 0.5 µl, 0.25 µl, 0.125 µl but have been rounded in the figure.

These experiments provided a good indication of which antibody volumes no longer present a risk of overstaining. But as very few cells were positive for the markers in mononuclear cells, we proceeded to stain samples that have higher expression frequency for the respective markers. We lineage depleted mononuclear cells to test the CD117 and CD110 antibodies, and cultured HUVEC (an endothelial cell line) to test the antibodies for CD143, CD144 and EPCR.

We stained $10x10^6$ HUVEC cells with an antibody volume of 0.5 µl for CD144 and EPCR, and 0.25 µl for CD143. We also carried out staining for the respective PE antibodies at equivalent amounts to compare the frequency of positives. We

observed large degree of similarity in both the frequency of the positives as well as the pattern of staining between PE and BV conjugates for all antibodies we tested. Almost all HUVEC cells were positive for CD144 and EPCR (Fig. 5.4A & 5.4C). A smaller fraction of HUVEC cells was positive for CD143, but the frequency of positives as well as pattern of staining was similar between PE-CD143 and BV786-CD143 (Fig. 5.4B). For CD110, we used lineage-depleted cord blood mononuclear cells and found very limited expression, but at similar frequencies (Fig. 5.4D). We attempted to look more closely into the $CD34^{+}/CD38^{-}$ fraction but found very limited number of positive cells.

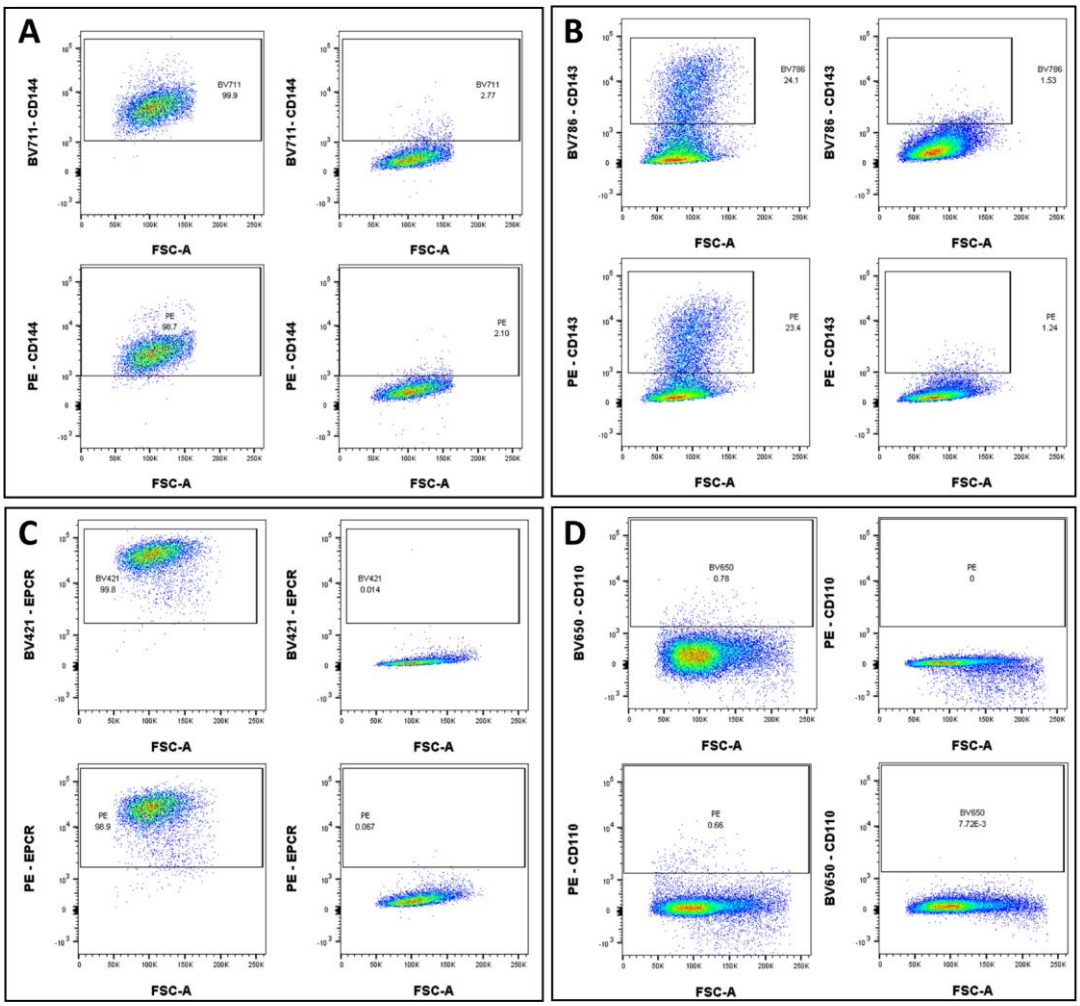

Figure 5.4. Testing antibody volume and specificity

A. Flow cytometry plots for single live HUVEC cells stained for CD144 conjugated to BV711 (top left panel) or PE (bottom left panel) and respective FMO controls (BV711 – top right; PE – bottom right).

B. Flow cytometry plots for single live HUVEC cells stained for CD143 conjugated to BV786 (top left panel) or PE (bottom left panel) and respective FMO controls (BV785 – top right; PE – bottom right).
C. Flow cytometry plots for single live HUVEC cells stained for EPCR conjugated to BV421 (top left panel) or PE (bottom left panel) and respective FMO controls (BV421 – top right; PE – bottom right).
D. Flow cytometry plots for single live lineage-depleted cord blood mononuclear cells stained for CD110 conjugated to BV650 (top left panel) or PE (bottom left panel) and respective FMO controls (BV650 – top right; PE – bottom right).

For the validation of the BB700-CD117 antibody we used an APC-conjugated antibody of the same clone, we had high confidence in. We stained lineage-depleted mononuclear cells gated for single, live, $Lin^-/CD34^+/CD38^-$ cells and visualised CD117 for the respective antibody formats (Fig. 5.5). We found both CD117-APC and CD117-BB700 were giving us a similar staining and the expected frequency of positive cells, which we have shown is between 90 and 100% for the $CD34^+/CD38^-$ population.

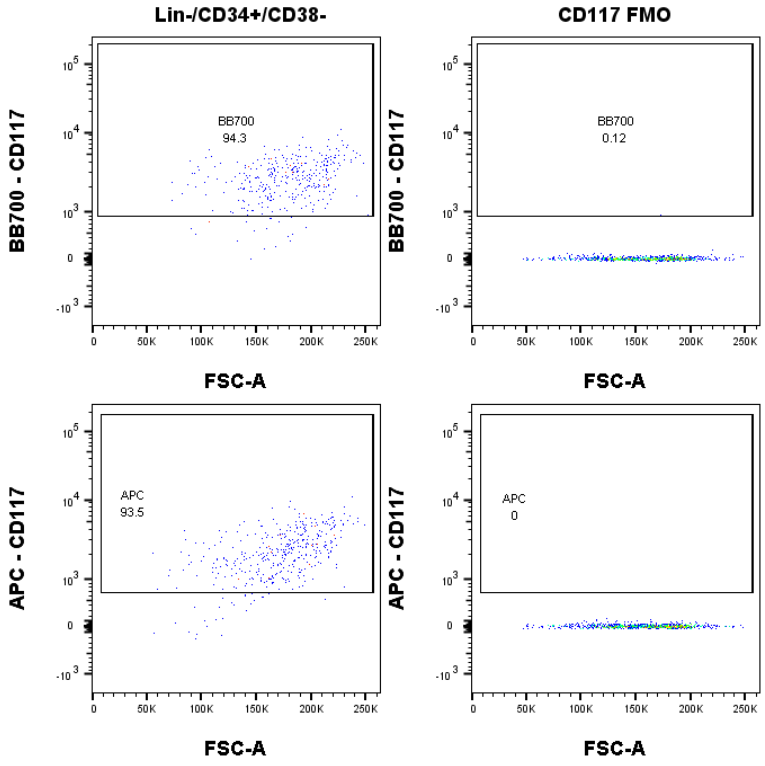

Figure 5.5. Validation of BB700-conjugated CD117 antibody

Staining of lineage-lineage depleted cord blood mononuclear cells with BB700-CD117 (top left panel) and APC-CD117 (bottom left panel) and respective FMO controls (top right and bottom right panels).

To summarise, we have designed and optimised a 17-colour flow cytometry panel for the study of HSC cell surface marker dynamics during ontogeny. We have quantified the antigen density of these markers in three different FL stages, 6, 12 and 18 pcw. We assigned fluorochromes according to cell surface abundance by assigning brighter dyes to rarer cell surface markers and dimmer ones to more abundant antigens. We then titrated our antibodies and further validated CD143, CD144, EPCR, CD110, and CD110 using cells with higher frequency of positive expression.

5.3 Multidimensional analysis and clustering

We selected a total of 13 foetal samples for analysis, consisting of 8 FL (staged between 7 and 18 pcw) and 6 FBM (staged between 12 and 20 pcw) samples (Table 5.4).

Sample Type	**Stage (pcw)**	**Sample ID**
Foetal liver	7	L07
Foetal liver	8	L08
Foetal liver	10	L10
Foetal liver	11	L11
Foetal liver	12	L12
Foetal liver	14	L14
Foetal liver	18	L18
Foetal bone marrow	12	BM12
Foetal bone marrow	13	BM13
Foetal bone marrow	15	BM15
Foetal bone marrow	17	BM17
Foetal bone marrow	18	BM18
Foetal bone marrow	20	BM20

Table 5.4. List of foetal samples used for analysis

All samples were thawed after having been stored in liquid nitrogen. Cells were spun down and counted. All samples were stained and analysed on the same

day to minimise batch effects. For samples, where more than $40x10^6$ cells were counted, amount of antibodies was adjusted. Data was collected at approximately 5, 000 events per second. Compensation was calculated automatically using single-colour bead controls prior to data collection. Pre-processing of files was undertaken with FlowJo to select for single, live, Lin^- and $CD45^+$ cells. Gated populations were exported and used in downstream analysis.

We used several packages developed for high-parameter flow cytometry analysis to import, transform and scale the data (see Material and Methods). We plotted per-sample, markers expression distribution (Fig. 5.6). We observed some distinguishing markers between FBM and FL samples such as CD45RA, which was more highly expressed in bone marrow samples. Despite selecting for CD45-expressing cells, we kept the marker as an analysis variable, as CD45 expression levels can vary between different populations.

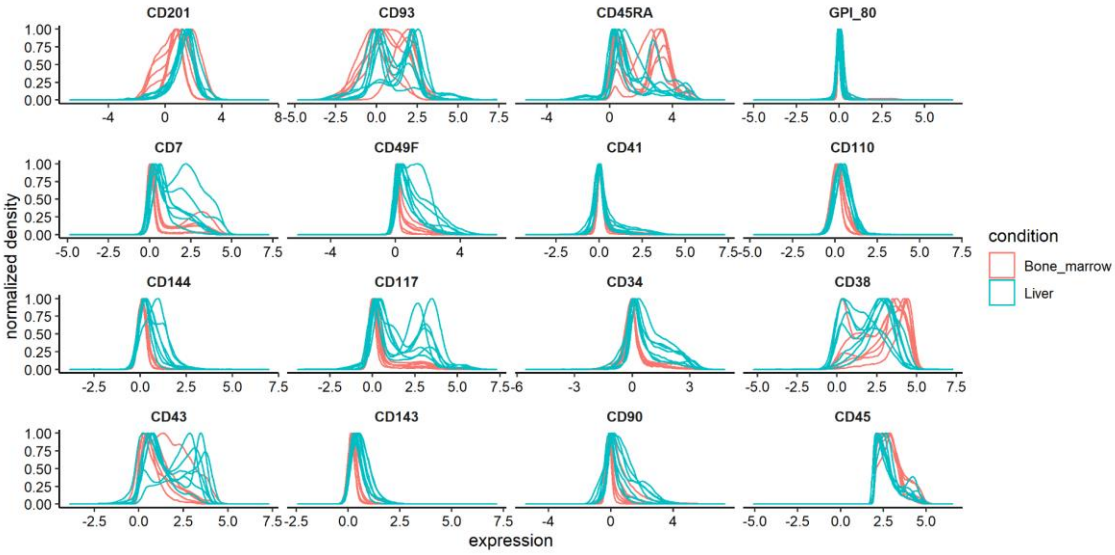

Figure 5.6. Per-sample marker expression distribution

Per-sample densities of marker expression (arcsinh-transformed with a cofactor of 250) of 16 cell surface markers. Sample types are presented as conditions, Bone_marrow (coral) and Liver (turquoise).

To visualise global similarities and differences between samples, we used a multi-dimensional scaling (MDS) plot, which performs unsupervised clustering of samples using median marker expression across all cells per sample (Fig. 5.7). In our MDS plot, we observed that the first dimension (MDS dim. 1) separates the data

by the tissue types. The second dimension separates samples by gestational age to some degree. This is particularly prominent for the FL samples, where first trimester samples (L07, L08, L010 and L011) cluster around the top of the chart, while second trimester (L14 and L18) are localised in the bottom corner. Interestingly, FL12 and BM12 clustered closely together. At this stage of development, haematopoietic cells from the FL are likely to be actively colonising the FBM. It is therefore possible that the $Lin^{-}/CD45^{+}$ composition of the two tissues is similar. Separation of FBM samples by MDS dim. 2 was significantly less pronounced, which indicates higher similarity in measured variables in these samples.

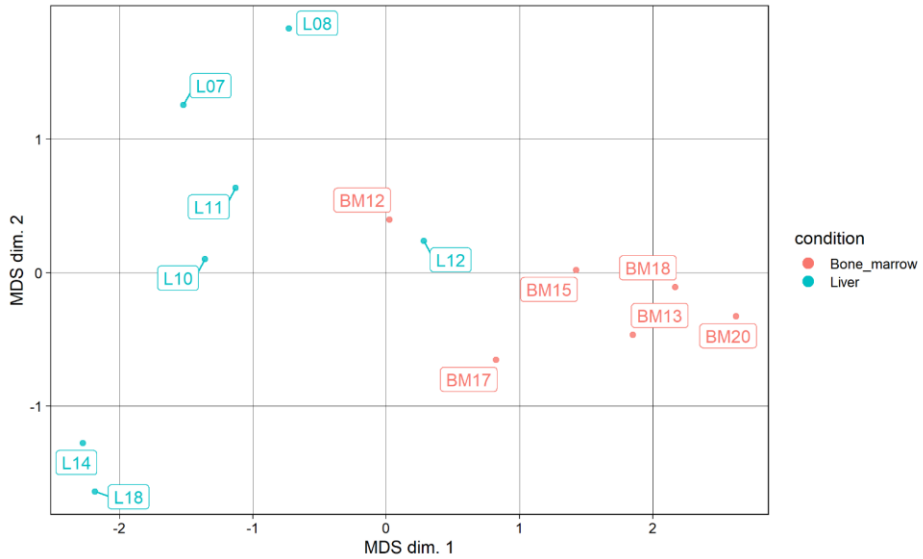

Figure 5.7. MDS plot for individual FL and FBM samples

MDS plot base on transformed median marker expression of 16 cell surface markers across all cells per sample. Plot labels represent sample IDs.

To gather more insight into the structure of the data by specific markers, we generated a heatmap of median marker intensities. In addition, we undertook hierarchical clustering by Euclidean distance (Fig. 5.8). This plot is useful for visualising which markers may best explain similarities and differences between samples. We observed that, again, samples clustered according to condition. Notably, FBM samples had higher median CD93, CD45RA and CD45 expression. On the other hand, CD201, CD90, CD143, CD110, and CD144 had higher median expression in FL samples. In terms of sample clustering, second trimester FL samples (L14 and L18) clustered separately from first trimester samples. In FBM, samples also clustered by developmental age with BM17, BM18 and BM20

clustering, while BM15 clustered closer to BM13. Similarly, to the MDS plot, BM12 clustered closer to L11 and L12. The markers that differentiated L11 and L12 from BM12 were CD41 and CD45 (highly expressed in BM12) as well as CD144 and CD34 (lowly expressed in BM12). BM12 also has the lowest expression of all FBM samples for the markers, CD45RA and CD38, which appeared to be responsible for the stratification of FBM and FL samples in the clustering.

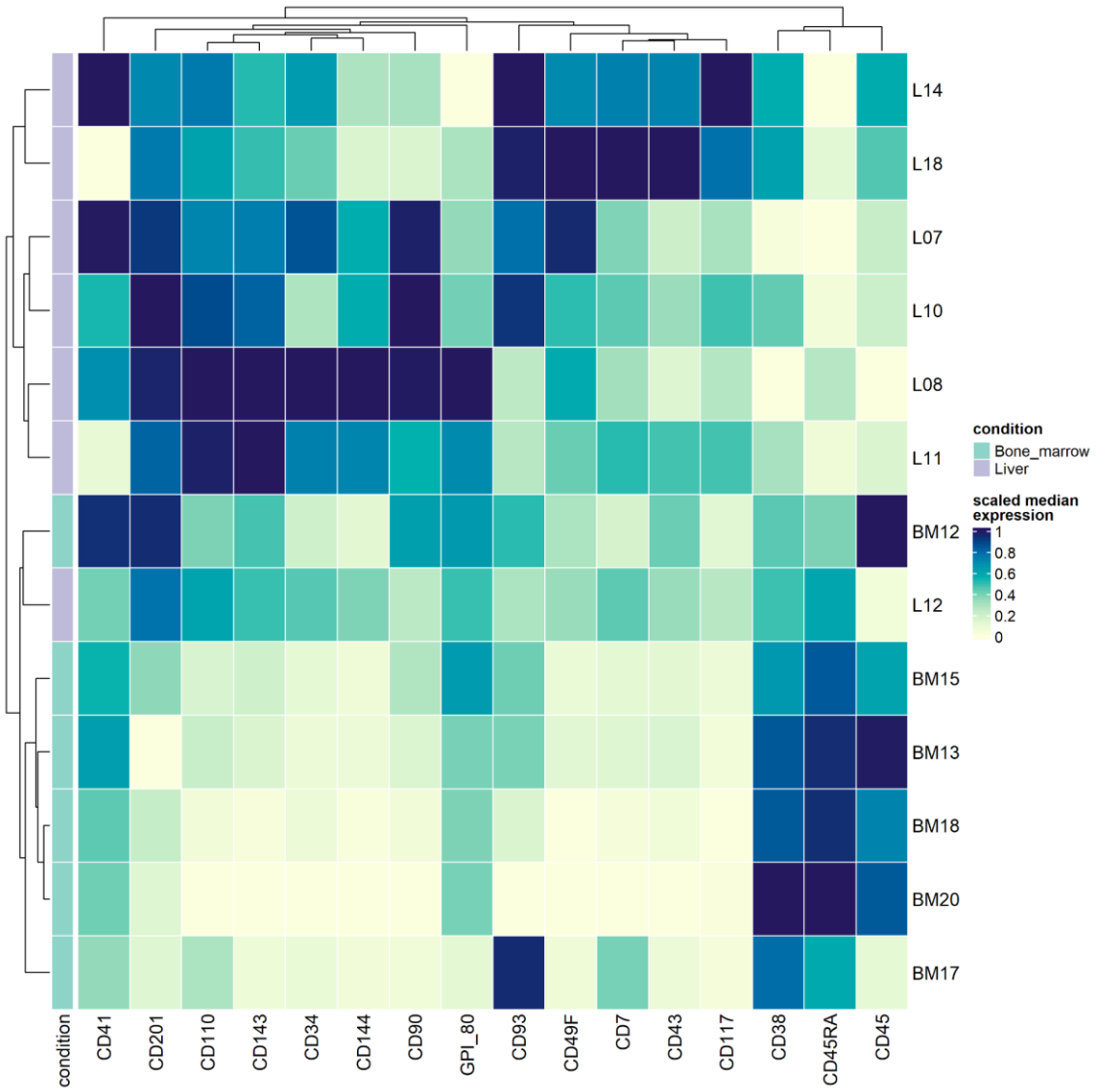

Figure 5.8. Heatmap of the median marker expression across samples

The heatmap represents scaled median expression (yellow-low expression; blue-high expression). Dendograms represent hierarchical clustering of samples and markers. Row annotation (lilac and green) represent sample conditions (Liver and Bone_marrow, respectively).

To identify which markers, contribute to the variance, within individual samples, we calculated non-redundancy scores. Higher scores are assigned to markers that are responsible for larger proportion of the intra-sample variability (Levine et al., 2015). The marker with the highest NRS score was CD45RA, followed by CD117, CD93 and CD38 (Fig. 5.9). CD45RA and CD38 expression define more mature populations, so likely represent the mature-immature population variability within individual samples.

The lowest NRS scores were assigned to CD110, CD143, CD144, GPI-80 and CD41, which appeared to have little contribution to the variability within individual samples. This is expected for markers that are expressed by few cells.

For quite a few of the markers, we noted individual NRS scores appeared to be higher in FL samples (Liver). It is possible that this is due to higher population diversity in FL samples compared to FBM ones. To check this, we proceeded with carrying out clustering analysis.

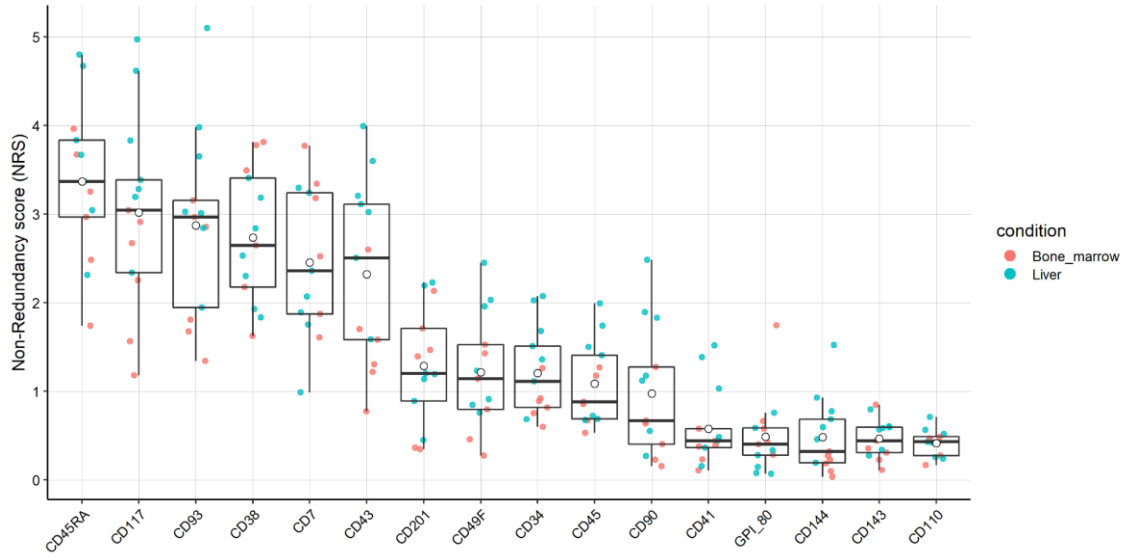

Figure 5.9. Non-redundancy scores for the cell surface markers

Box-plots of non-redundancy scores where full points represent individual samples (coral points – FBM; turquoise – FL). Empty circles represent mean NRS values per markers.

5.4 FlowSOM clustering of HSC cell surface marker expression reveals population diversity in foetal liver and foetal bone marrow samples

While cell population identification is typically carried out by manual gating, unsupervised clustering has the advantage of being less subjective, efficient and reproducible in the analysis of large datasets. We applied FlowSOM clustering to the dataset, which results in the mapping of all cells from every sample to a cluster. For flow cytometry data, automatic approaches of determining ideal cluster number are not always optimal, so it is recommended that the initial number of clusters supplied to the algorithm is relatively high. In our case, we tested 30 and 20 clusters and found that no new clusters, with significantly different distribution of median scaled marker expression emerged with higher number of clusters. Therefore, we performed the initial analysis using a setting of 20 clusters. We also performed meta-clustering of the 20 FlowSOM clusters to help us determine similarity between clusters.

The most prominent clusters with a relative size adding up to approximately 62% of all clusters were C2, C4 and C19 (Fig. 5.10). In C19, the scaled median expression for CD38 was high, with intermediate levels for CD45RA, and to lesser degree CD201. On the other hand, C2 was expressing intermediate levels of CD38 and higher levels of CD43. Other interesting clusters we identified, were C1 and C15. C1 expressed the highest levels of CD117 of all clusters, while C15 appeared to express high levels of markers CD144, CD90, CD49f and CD143. Both clusters appeared to be $CD45RA^{-}$. We also noted a set of clusters that the meta-clustering grouped together by what appeared to be shared expression of CD45RA, CD38 and to lesser degree CD7. These clusters, C11, C12, C17, C19, C20, we marked as potentially committed.

As human foetal HSCs have been reported to be $CD34^{+}/CD38^{-}$, we looked for a cluster with an expression pattern that could define such a population. The only clusters that we could note using globally scaled median marker expression (Fig. 5.10) were C3, C10 and C15. But while scaled median marker expression can be useful in giving an overview of a dataset, it is inadequate in representing the expression specifics of a cluster.

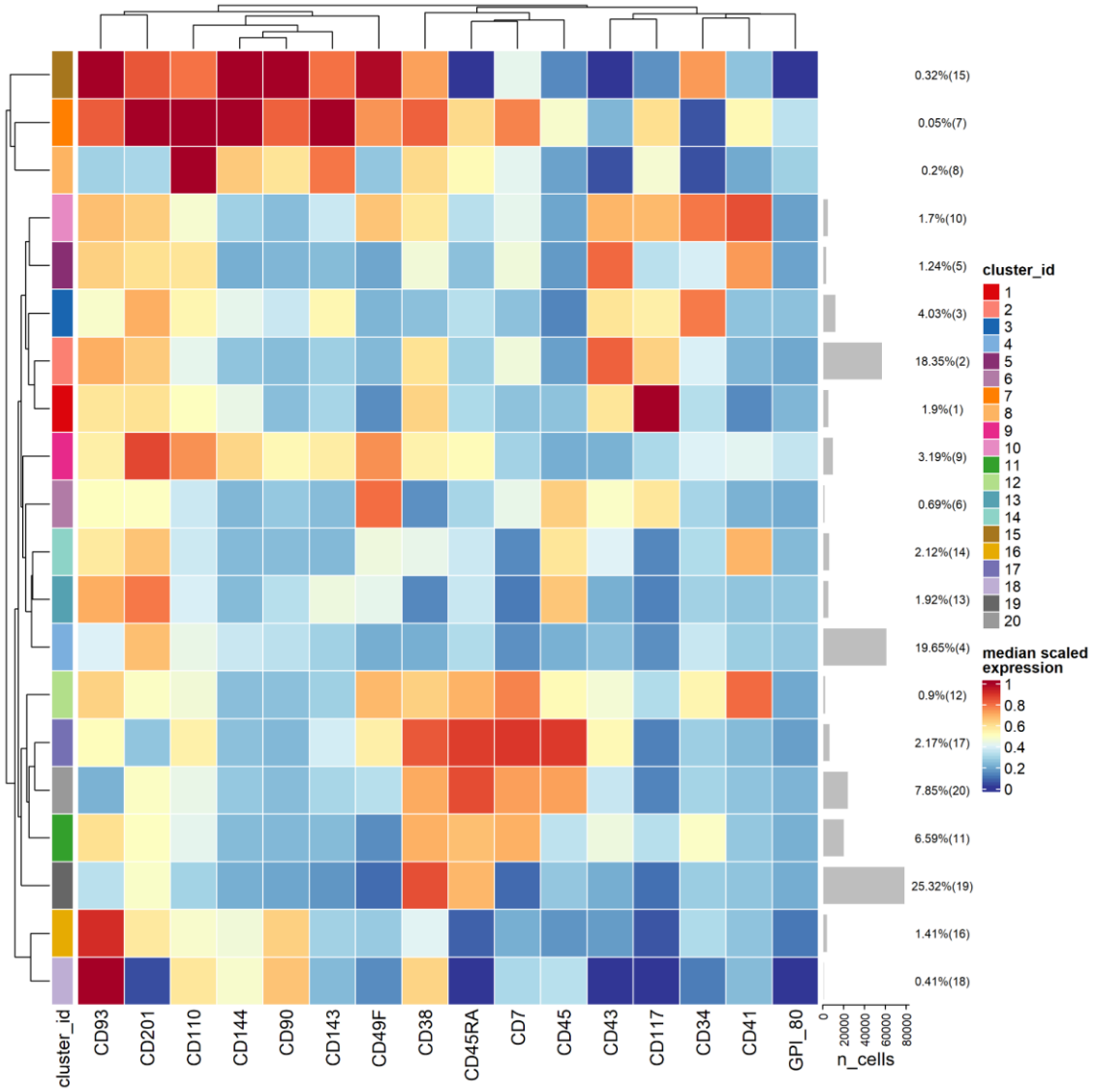

Figure 5.10. Heatmap of the median marker expression across FlowSOM clusters

The heatmap represents scaled median transformed marker expression across all cells and samples. Dendograms represent hierarchical similarity between all clusters (rows) and markers (columns). Bar plots on the right side represent relative sized of clusters with absolute values in brackets.

We plotted the marker expression distribution for each individual cluster. What immediately stood out was C7, as it had abnormal expression distribution for every marker (Fig. 5.11). In the heatmap (Fig. 5.10), it appeared positive for most markers. Such pattern of expression can be observed in dying cells or debris. As this cluster represented only 0.05% of the dataset, we excluded it from further analysis.

We used the average expression distribution (blue ridge plots) to determine whether a marker had a lower or higher relative expression distribution in a particular

cluster (Fig. 5.11). The average expression distribution for CD38 was higher compared to other markers, suggesting that scaled median expression values may not accurately represent the actual data. As a reference, the cluster with the lowest expression was C13 with peak falling between 0 and 1. The highest expression was in C19 (peak falling around 4). In comparison, the clusters with the highest CD34 expression (C15 and C10) had expression peaking well below 4, while cluster, C18 had CD34 expression distributed below zero.

C1, which we highlighted for its high CD117 expression, did indeed have exceedingly bright CD117 expression distribution. Its CD34 expression was widely spread around 0, which indicated the presence of some $CD34^{Low}$ cells in this population. Its CD38 expression distribution was below 4 and spread toward 0, which could indicate the population is $CD38^{Low}$. This cluster also had positive expression distribution for CD144 and CD143.

C3 had high CD34 and low CD38 expression distribution, while also being positive for CD117 and CD144, and to lesser degree CD143 and CD90. C15, in addition to being highly positive for CD34 was one of the clusters with the highest expression for CD144, CD143, and CD90, while also being positive for CD117. Both of these clusters are quite interesting as they potentially represent $CD34^{+}/CD38^{-}$ populations while expressing a number of potential HSC markers.

We were also interested whether a population with the $CD34^{-}/CD38^{-}/CD93^{+}/CD117^{+}$ phenotype could be identified, but no such clusters stood out at this stage of the analysis. C10 had a CD34 expression with two peaks, suggesting that this population may contain both $CD34^{+}$ and $CD34^{Low}$ cells. But while it did express CD117, its expression for CD93 was low.

Visualisation of expression distribution suggested that data transformation had succeeded at enabling us to visualise all captured cells. It also helped us identify a cluster potentially containing debris or dying cells. What also stood out was that some markers had a positive shift in the expression distribution compared to others. This was the case with CD38. The possible explanation for this is that due to constraints in commercial availability, we assigned this relatively abundant marker to a bright fluorochrome. This can be resolved computationally by increasing the cofactor in the arcsin transformation, which would result in bringing the CD38 expression distribution closer to zero. However, as the transformation is applied globally, this could result in loss of expression resolution for the rest of the markers.

This would have impact on markers such as GPI-80 and CD41 by further narrowing down their expression distribution. Therefore, we proceeded with the current settings. However, insufficiently high co-factor in arcsin transformation can lead to low expression being presented as intermediate and even high.

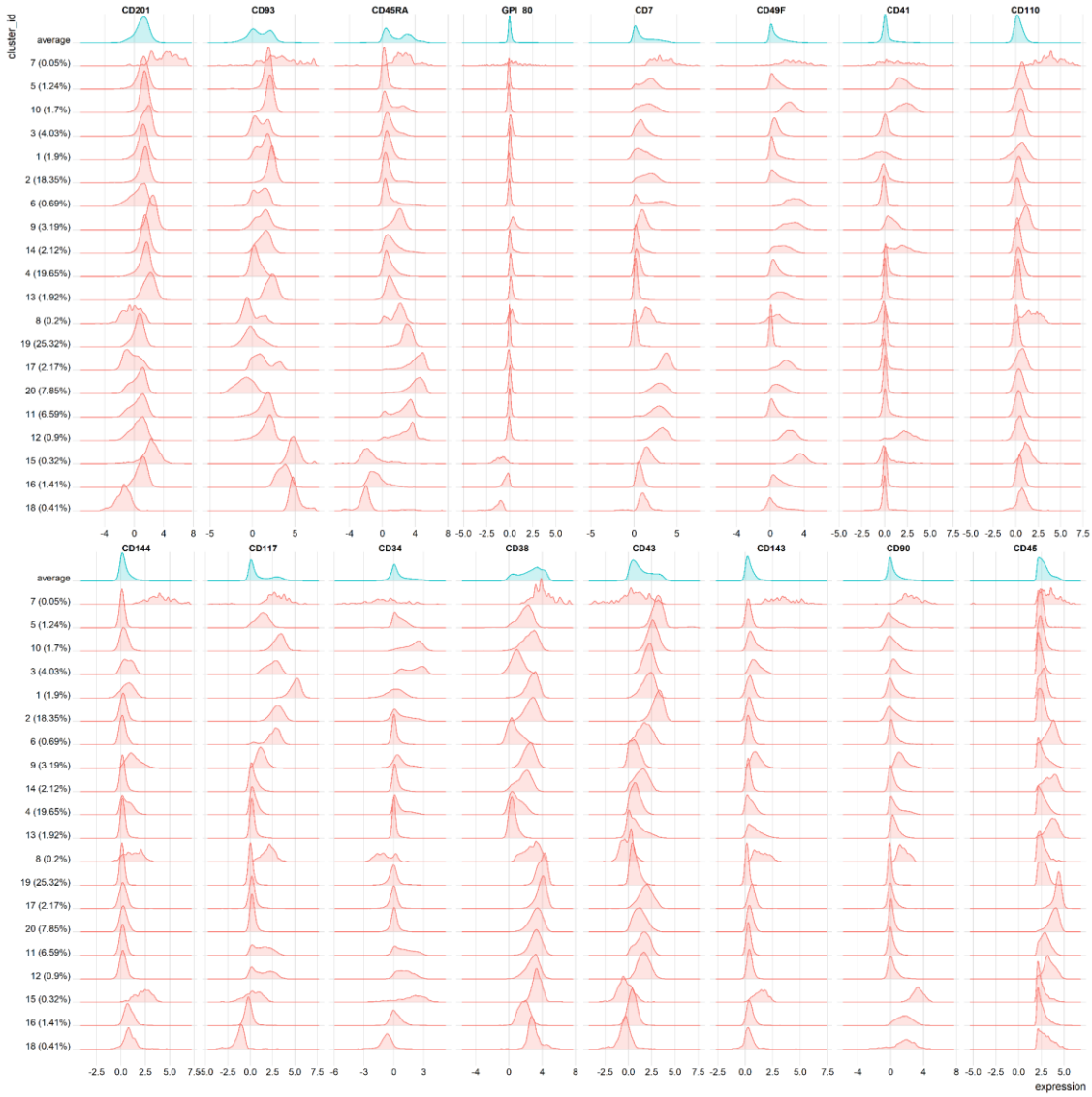

Figure 5.11. Distribution of non-scaled marker intensities across FlowSom clusters

Red ridge plots represent marker expression distribution per individual clusters. Blue ridge plots represent average expression distribution of marker across all clusters.

5.5 Dimensionality reduction analysis

To help us asses our clustering, we proceeded with dimensionality reduction analysis. The dimension reduction method we chose was UMAP and it was computed independently of FlowSOM clustering. We were also interested in visualising potential differences between FBM and FL samples, so we plotted UMAP projections separately for each tissue type, overlaying the plots with the computed FlowSOM clusters.

The differences between FBM and FL that stood out were a few clusters that appeared to be differentially abundant between the two conditions (Fig. 5.12). For example, C13 and C19 seemed more abundant in FBM, while C2 and C9 appeared more abundant in FL.

"Spillover" of FlowSOM clusters could be observed from one UMAP dimension area into another. This was true for C17 which was separated by UMAP dim.2. This was also observed to a lesser degree in C3, C4 and C11. This can sometimes be due to under-clustering or insufficient amount of variables, however, we hypothesized that this may be due to some clusters being absent at some stages of development, adding a layer of complexity UMAP cannot resolve.

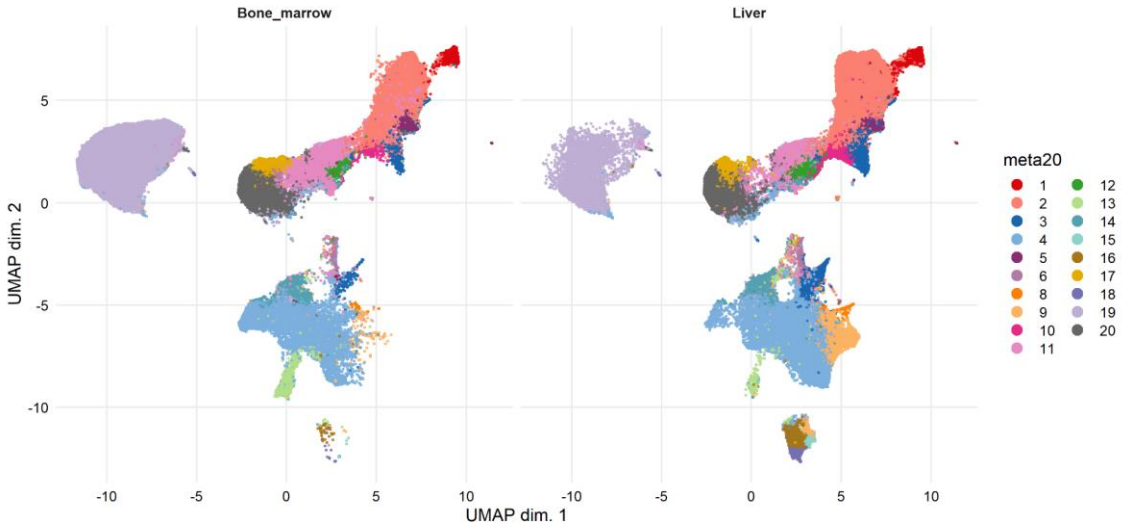

Figure 5.12. UMAP based on marker expression across FBM (Bone_marrow) and FL (Liver) samples

UMAP projections coloured according to the 20 FlowSOM clusters (excluding C5 and C19) and stratified by conditions (Bone_marrow & Liver). Cluster colour codes have been amended as a result of the removal of two clusters, but the numbering remains the same.

To check this, we plotted UMAP projections for each individual sample (Fig. 5.13). Interestingly, we observed an abundance of clusters already present at early stages of FL development. The majority of L07 and L08 clusters could be detected in more advanced FL as well. A notable exception is C19, which is a relatively large cluster. It is mostly absent in first trimester samples but steadily increasing and reaching a peak at L14 before beginning to diminish again. In contrast, this cluster is present in all FBM samples. Another notable feature of the data was C2. In FL samples, this was an abundant cluster. In FBM, while present, it appeared significantly sparser.

Another trend that encompassed both FL and FBM samples, were the group of clusters in the lower section of the UMAP dim. 2 (clusters 3, 4, 8, 9, 13, 14, 15, 16 and 18). These clusters were present in the earliest FL samples and appeared to be relatively stable in the first trimester. However, in second trimester this group of clusters steadily diminished with C15, C16 and C18 disappearing almost entirely. These three clusters, in addition to C9 appeared to be almost entirely absent from the FBM.

The remaining clusters of this group, C3, C4, C13, and C14 were present at the earliest FBM stage of our dataset, BM12, but grew sparser with advancing developmental age.

Some clusters seemed to remain stable over different stages. C11, C17 and C20 were present in all samples, while others like C1, C10 and C12 seemed to be relatively stable in the FL, but disappeared in late stage FBM samples.

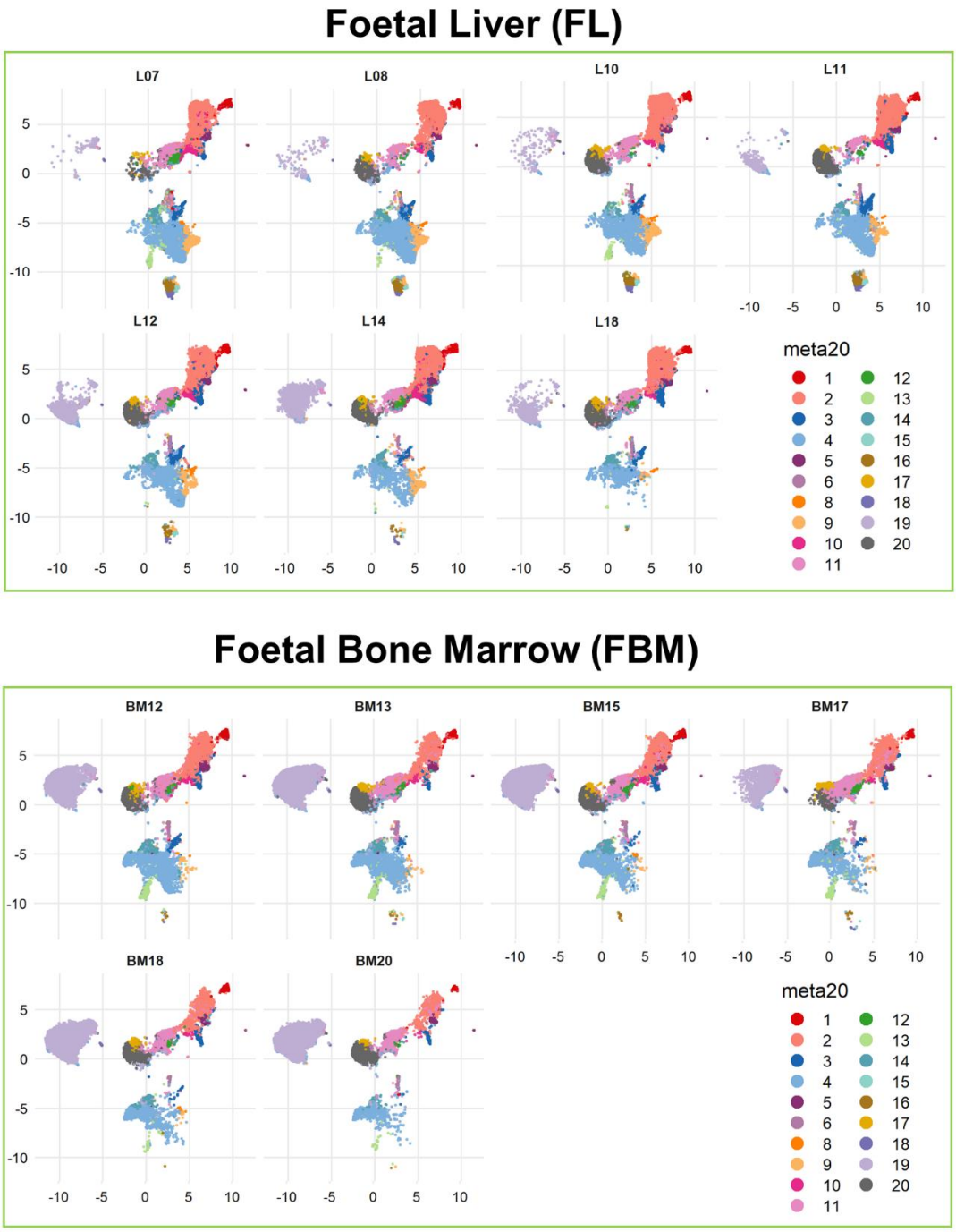

Figure 5.13. UMAP projections stratified by sample

UMAP projections of individual FL (top panel) and FBM (bottom panel) samples coloured by FlowSOM cluster.

5.6 Cluster dynamics across samples and developmental stages

Seeing that there were differences between both sample types and developmental stages, we were interested in capturing the changes in cluster dynamics at a finer level. While UMAP projection proved to be useful, high cell density doesn't allow accurate estimation of changes in individual cluster abundances over developmental age. Therefore, we plotted the proportions of each cluster per sample.

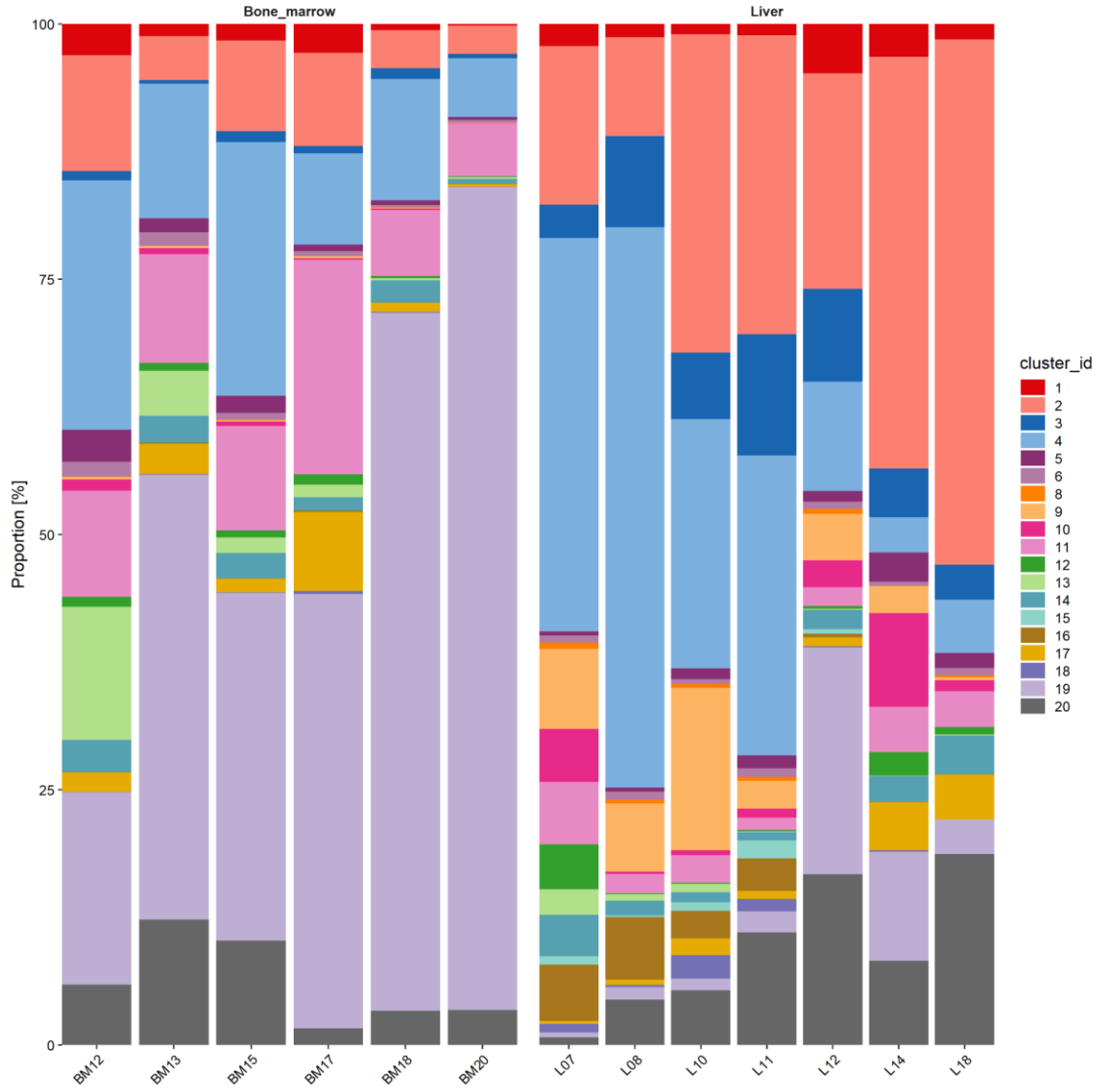

Figure 5.14. Relative FlowSOM cluster abundance

Stacked barplots of 18 FlowSOM clusters across samples stratified by tissue type (left panel – FBM; right panel – FL).

Some of the clusters already highlighted in the dimensionality reduction stood out again. One of them was C19, which steadily increased in relative size in the FBM with advancing stages. In FBM20 this population accounted for over 75% of the $Lin^-/CD45^+$ bone marrow cell content. Interestingly, this cluster constituted less than 1% in first trimester FL samples, only reaching a notable size at L12 before decreasing again. While the FBM is known to be actively colonised by FL haematopoietic cells from 11 pcw onwards, little is known about whether some of the cells that colonise the FBM can shuttle back to the FL. C19 isn't expressing many of the markers on our panel. Most notably it's highly positive for CD38 and CD45RA, suggesting it is either a committed or progenitor population (Fig. 5.15). This is further confirmed by the rapid relative growth of this cluster in the FBM. Different foetal haematopoietic organs have different capacities to support haematopoietic progenitors in different developmental stages, which may explain why it is not abundant in the FL. Seeing that the levels of this population are quite high it probably plays an important role in foetal homeostasis.

C11 was similar to C19 with intermediate to high CD38 expression, and intermediate CD45RA expression. C11, however, had wide expression distributions for both CD117 and CD34, suggesting the population encompasses $CD34^+$ and $CD117^+$ cells. C11 was also notable with its dynamics over gestational stages, being one of the clusters that was present in both FL and FBM samples across all stages. It did, however, remain quite stable in the FL, while in the FBM it was quite abundant at 12 pcw only to slowly decrease in size with advancing stages.

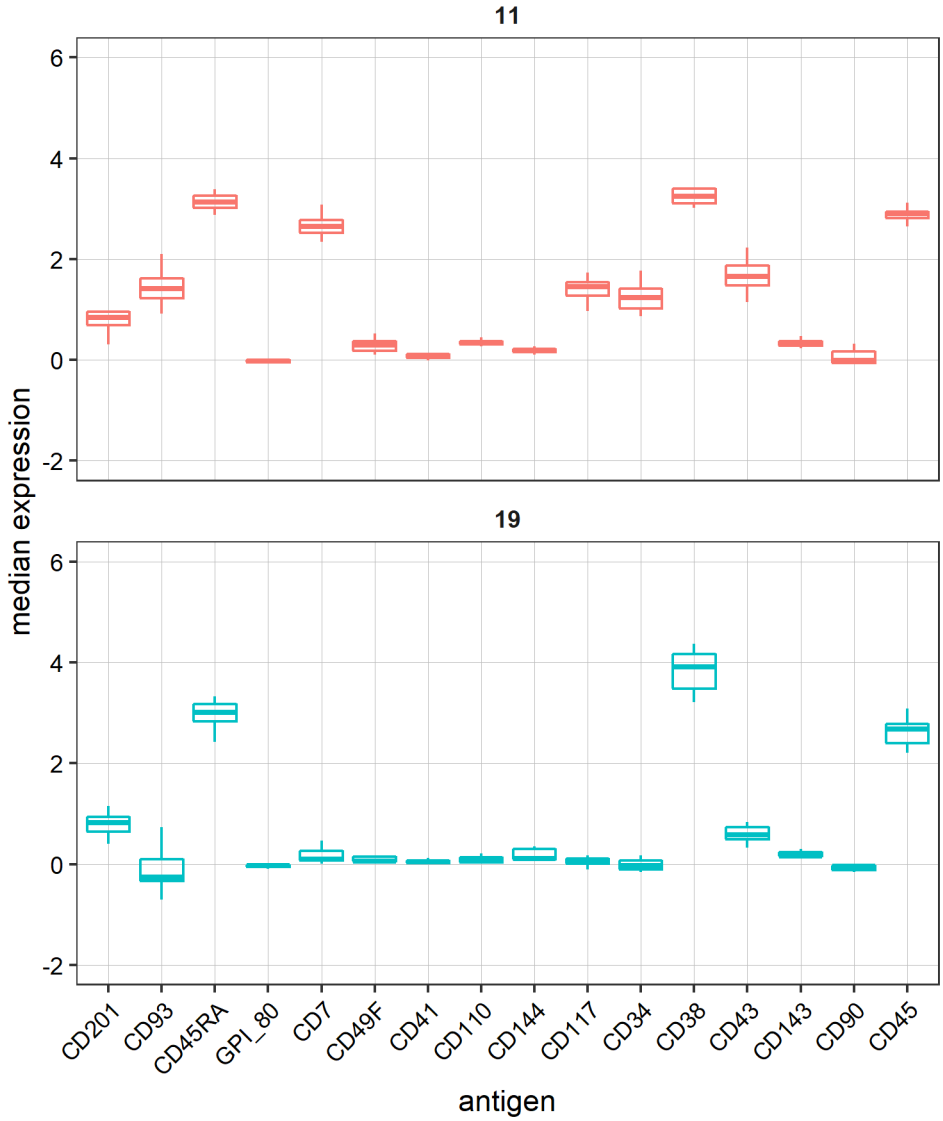

Figure 5.15. Median marker expression in FlowSOM clusters 11 and 19

Boxplots of median marker expression in C11 (top panel) and C19 (bottom panel). Horizontal line across boxes represents median expression across all samples.

A smaller cluster that was noticeable in the FBM, but was mostly absent from the FL, was C13. C13 gradually decreased in size in FBM samples with advancing stages, but it was one of the most abundant bone marrow clusters in BM12. It was negative for most markers with the exception of CD201, CD93 and CD49f (Fig. 5.16). The presence of this cluster in the FBM at 12 pcw suggests it is either one of the first populations to populate the bone marrow or one of the first to be generated by the newly residing HSCs and progenitors. Some mature haematopoietic cells like macrophages have been shown to play a role in the homing and retention of HSCs during development in zebrafish (D. Li et al., 2018). It is possible that other mature

lineages also participate in the priming of the foetal bone marrow niches and retention of migrating HSCs in mammals. It is also possible that this population is produced in the FL but rapidly expands once it reaches the FBM. In the mouse foetal liver, $CD93^{+}/CD45RA^{Low}$ cells have also been shown to be lymphoid progenitors (Montecino-Rodriguez et al., 2006; Yamane et al., 2009).

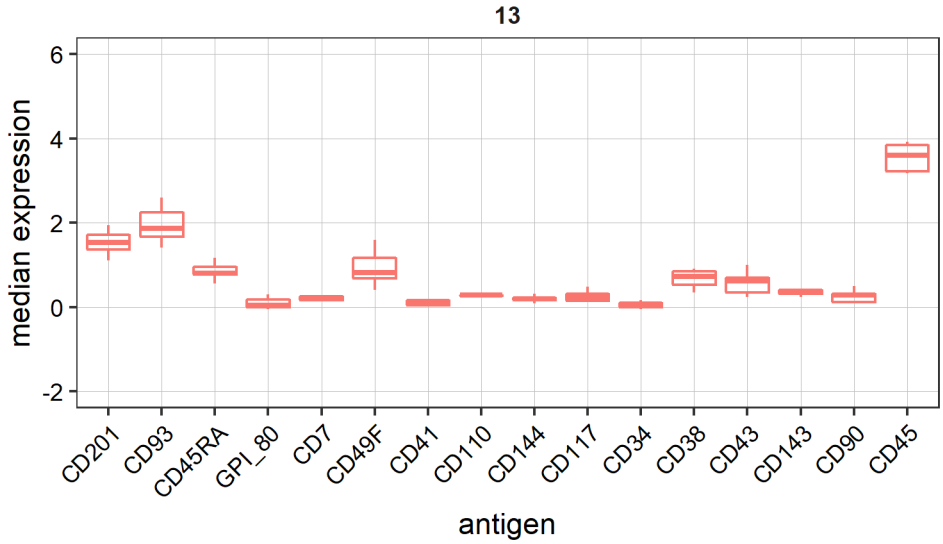

Figure 5.16. Median marker expression in FlowSOM cluster 13

Boxplots of median marker expression in C13. Horizontal line across boxes represents median expression across all samples.

Other clusters like C2 and C9 were more abundant in FL samples. C2 was one of the most abundant clusters, steadily increasing in size with advancing gestational age. It was observed at lower levels in the FBM, where it decreased with advancing stages. C2 expressed relatively high levels of CD38 (Fig. 5.17), and its expression distribution for CD34 was spread out, suggesting the presence of some $CD34^{+}$ cells (Fig. 5.11). Its median CD49f and CD93 expression was also relatively high (Fig. 5.17 & 5.11). This cluster was quite large, and therefore unlikely to represent an immature HSC population. However, its relative expansion in the liver is more controlled compared to the expansion of C19 in the bone marrow, which could hint at an immature progenitor.

C9 was also a notable cluster in FL samples with relatively large proportions, but decreased in size in late second trimester samples and was absent from L18. In the FBM, it was mostly absent. With higher expression of CD45RA than C2, C9 was more likely to represent a maturing population (Fig. 5.11 & 5.14).

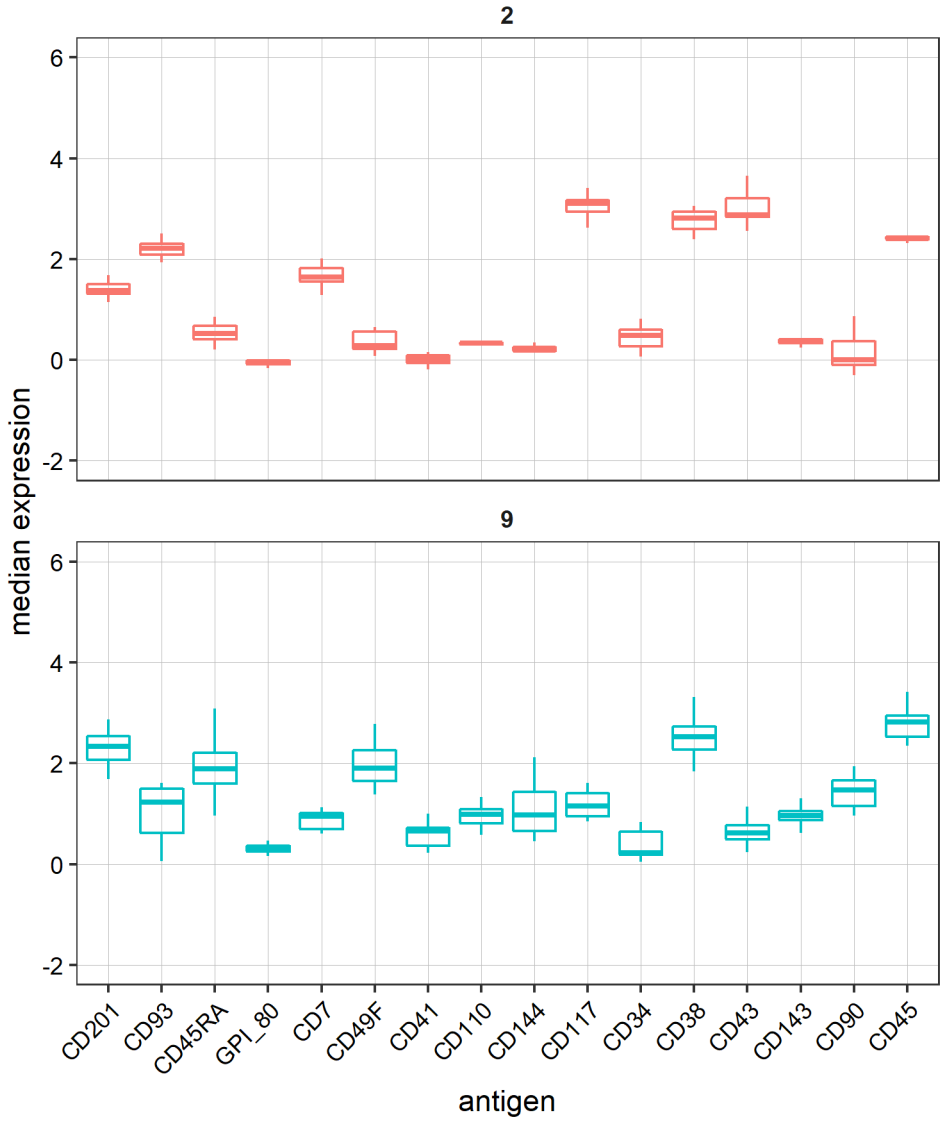

Figure 5.17. Median marker expression in FlowSOM clusters 2 and 9

Boxplots of median marker expression in C2 and C9. Horizontal line across boxes represents median expression across all samples.

We observed that some clusters were most abundant in early stage FL and FBM samples and gradually decreased in relative size with developmental age. C3 and C4 were two clusters that were relatively big in early FL but gradually decreased in proportion. Both clusters were present in the FBM at 12 pcw. C3 was relatively small, while C4 was of significant size, but gradually decreased. Of the two clusters, C3 was quite interesting appearing to be $CD34^{+}/CD38^{-}/CD117^{+}$. C4 appeared to have low expression for all markers with the exception of CD201 (Fig. 5.18). It is possible that these populations play a role in the early colonisation of the FL and

FBM, but are no longer needed, once the haematopoietic niches are established. It could also be that the cell surface marker phenotype of populations changes following FBM colonisation, but persists in the foetal liver until later in gestation. C4 could represent a mature haematopoietic population that seeds other haematopoietic organs or enters circulation.

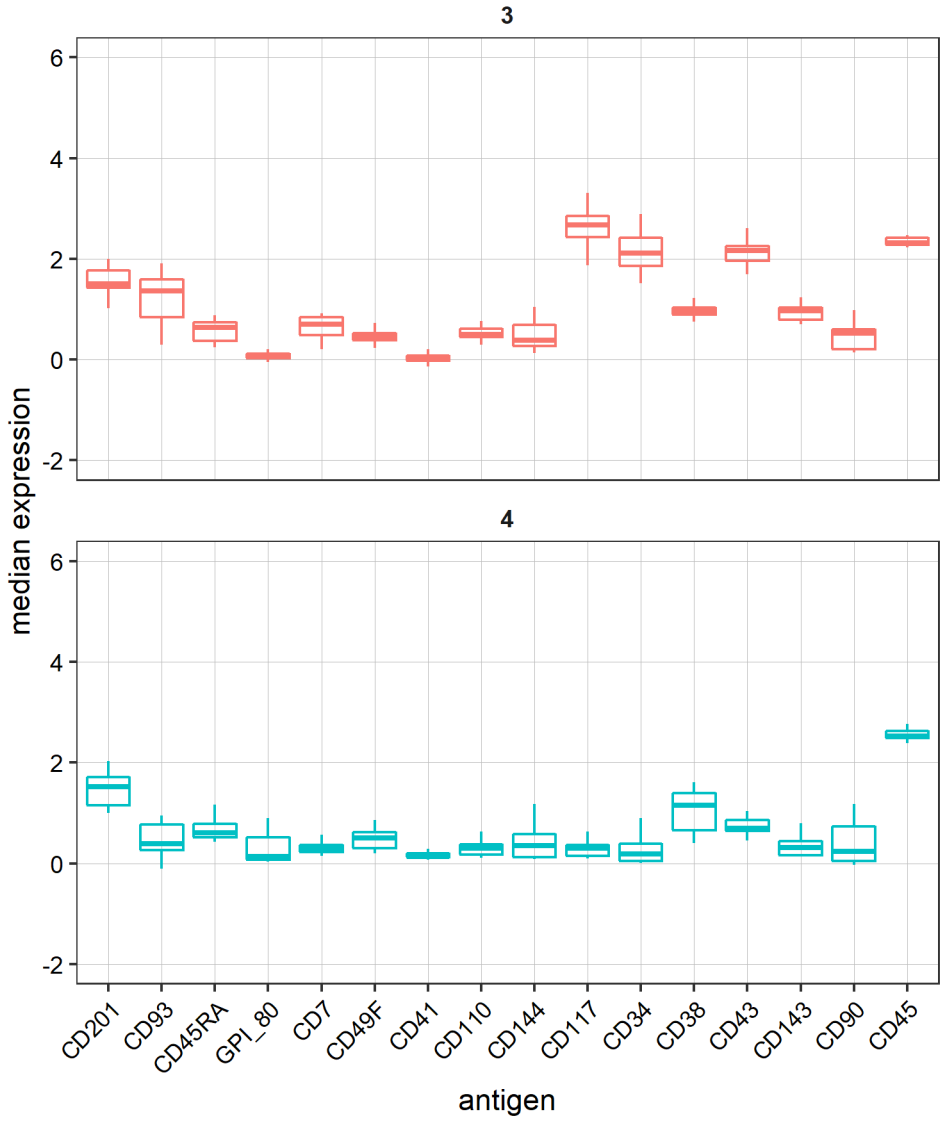

Figure 5.18. Median marker expression in FlowSOM clusters 3 and 4

Boxplots of median marker expression in C2 and C9. Horizontal line across boxes represents median expression across all samples.

Three clusters of relatively small size appeared to be most prominent only in first trimester FL samples. C15, C16, and C18 were detectable in L07, L08, L10, L11.

C15 and C18 could also be observed in L12 but at very small proportions (Fig. 5.14). These clusters were mostly absent from FBM samples.

C15 had a widely spread distribution for CD34 expression but was one of the most positive clusters for CD34 (Fig. 5.11). In addition, C15 expressed the highest levels of CD144, C143 and CD90, while also being $CD45RA^-$ (Fig. 5.11 & 5.19). In terms of CD38 it expressed more intermediate levels.

C16 and C18 were highly similar in expression patterns. Both were negative for CD34 and CD45RA. Both clusters expressed CD93, with C18 being the cluster with the highest expression for CD93. C18 also had slightly elevated expression for CD38. Due to the high degree of similarity between these clusters, it is possible that they actually represent the same biological population that has been split up in the analysis because of over-clustering. We also noted the transformation cofactor may not have been suitable for CD38, which may have also contributed to them being split up.

The defining feature of C15, C16, and C18 is that they are restricted to first trimester FL samples. C15 was of particular interest with high expression of a number of markers. However, this population is not maintained in the FBM. It is possible that once it colonises the FBM, its cell surface marker phenotype changes in response to the new conditions of the bone marrow niche. C16 and C18, on the other hand, being restricted to early liver, can potentially be haematopoietic cells originating from the yolk sac (YS). Murine YS-derived haematopoietic cells have been reported to express both CD38 and CD93 (Dagher et al., 1998; Yoder et al., 1997).

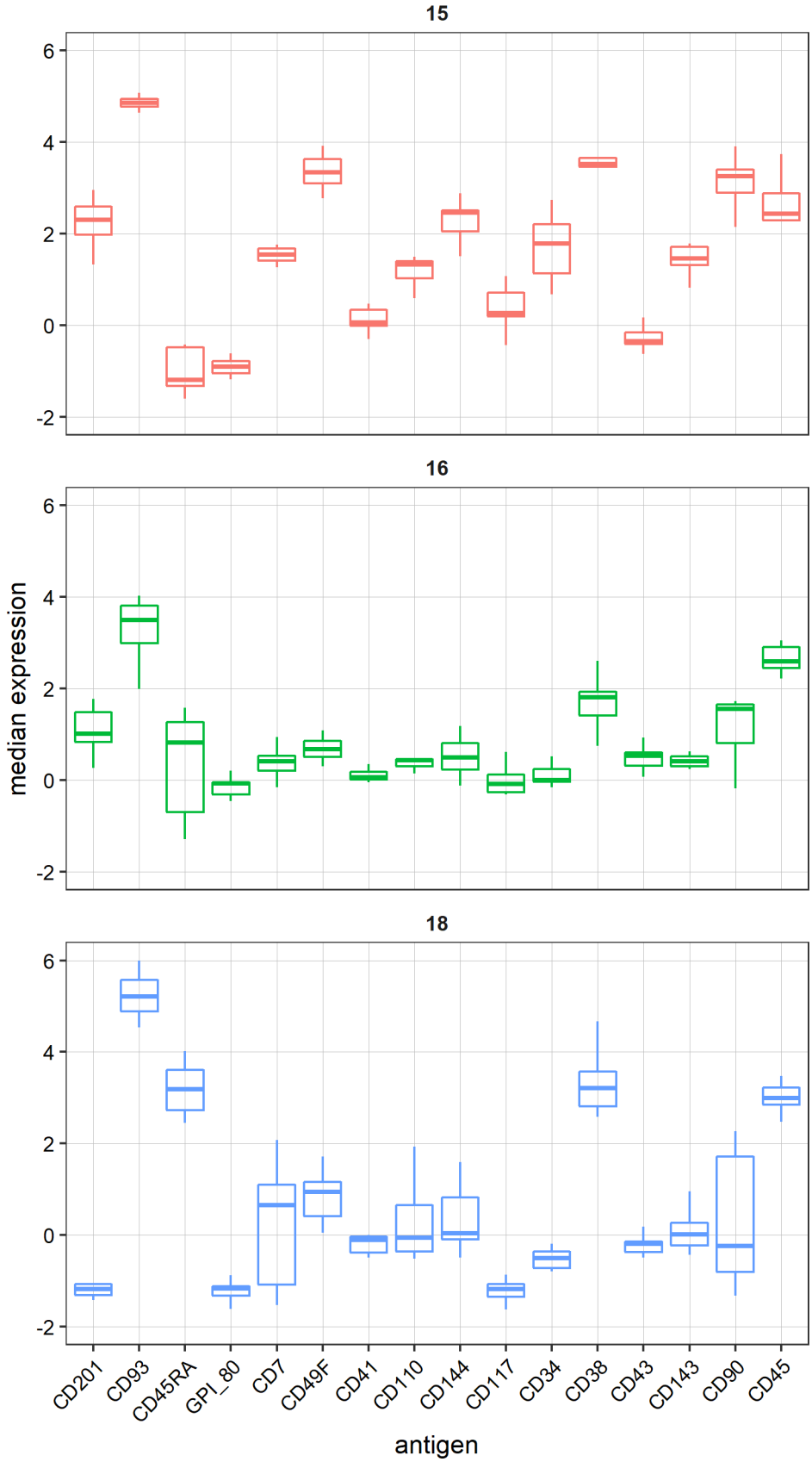
15
16
18
median expression
6
4
2
0
-2
CD201
CD93
CD45RA
GPI_80
CD7
CD49F
CD41
CD110
CD144
CD117
CD34
CD38
CD43
CD143
CD90
CD45
antigen

Figure 5.19. Median marker expression in FlowSOM clusters 15, 16, and 18

Boxplots of median marker expression in C15, C16 and C18. Horizontal line across boxes represents median expression across all samples.

C1 was a cluster of relatively small size that was observed in all samples in all stages (Fig. 5.14). Its size fluctuated in both tissues. It decreased in size during first trimester FL samples, only to increase in L12 and steadily decrease again until L18. Interestingly, in the FBM, this cluster increased from 13 pcw until 17 pcw. In BM18 and BM20 it has decreased again. The dynamics of this cluster could indicate two waves of migration from FL to FBM at approximately 12 pcw and 17 pcw, which are preceded by population expansion in the FL. This cluster had widely spread expression distribution for CD34 which potentially encompasses $CD34^{-}$ and $CD34^{Low}$ cells (Fig. 5.11). It was also highly positive for CD117 and had relatively high expression for CD144, while lacking CD45RA expression (Fig. 5.20). The CD38 expression distribution in C1 was wide as well and potentially includes $CD38^{Low}$ cells as well (Fig. 5.11)

Another potential HSC cluster, C10, also fluctuated in size. Like C1, it went through waves of shrinking and expansion in FL samples (Fig. 5.14). It was relatively small in FBM12 and grew smaller with increasing stages. This cluster expressed CD41 and CD49f at high levels (Fig. 5.20). CD41 has been shown to be expressed by murine foetal HSCs and play a role in their trafficking across haematopoietic tissues (Emambokus & Frampton, 2003). CD49f has been shown to mark human cord blood and adult HSCs, but whether it has similar roles in foetal haematopoiesis is still unknown. Integrins such as CD49f have been reported to play a role in cell anchoring to niches (Corbel & Salaün, 2002; Yoshitake et al., 2003). Blocking of CD49f, on the other hand, has been reported not to affect trafficking of haematopoietic cells from the FL to the FBM (Qian et al., 2007). This, however, has been only shown in mouse.

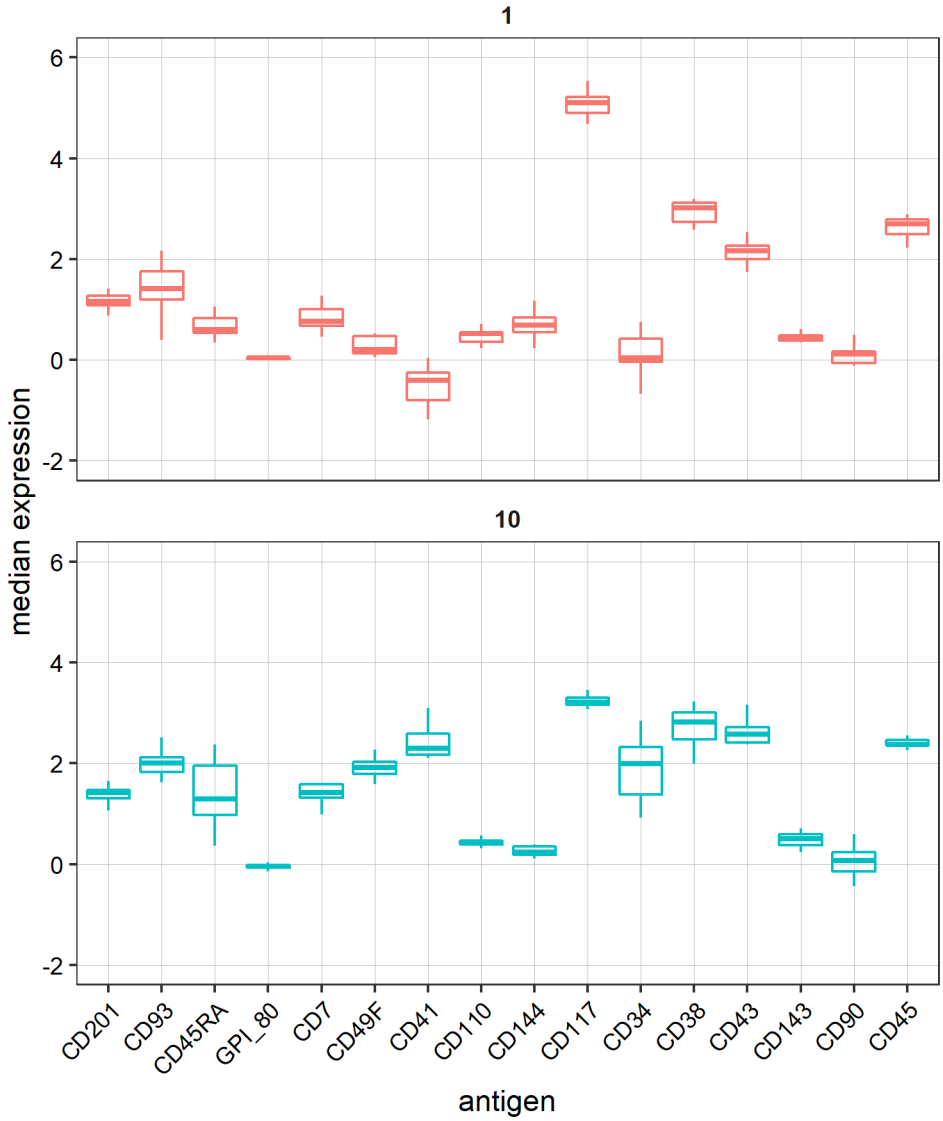

Figure 5.20. Median marker expression in FlowSOM clusters 1 and 10

Boxplots of median marker expression in C1 and C10. Horizontal line across boxes represents median expression across all samples.

Two other clusters with an interesting dynamic were C17 and C20. C10 was present in early FL samples and gradually increased with progressive gestational stages with the exception of L14, where it marked a decrease (Fig. 5.14). C17 remained relatively small in proportion until L14, when it grew in size. Both clusters were also notable in FBM samples, BM12, BM13, BM15, and BM17, but their relative proportions decreased in later samples. In terms of markers, these populations are likely to represent progenitors as they have high CD38 and CD45RA expression (Fig. 5.21). They may play a role in early colonisation of the FBM, but also be required in

the FL in later gestational stages. Both clusters defined by the same pattern of marker expression but slightly different intensities. These two clusters could represent the same population split by over-clustering.

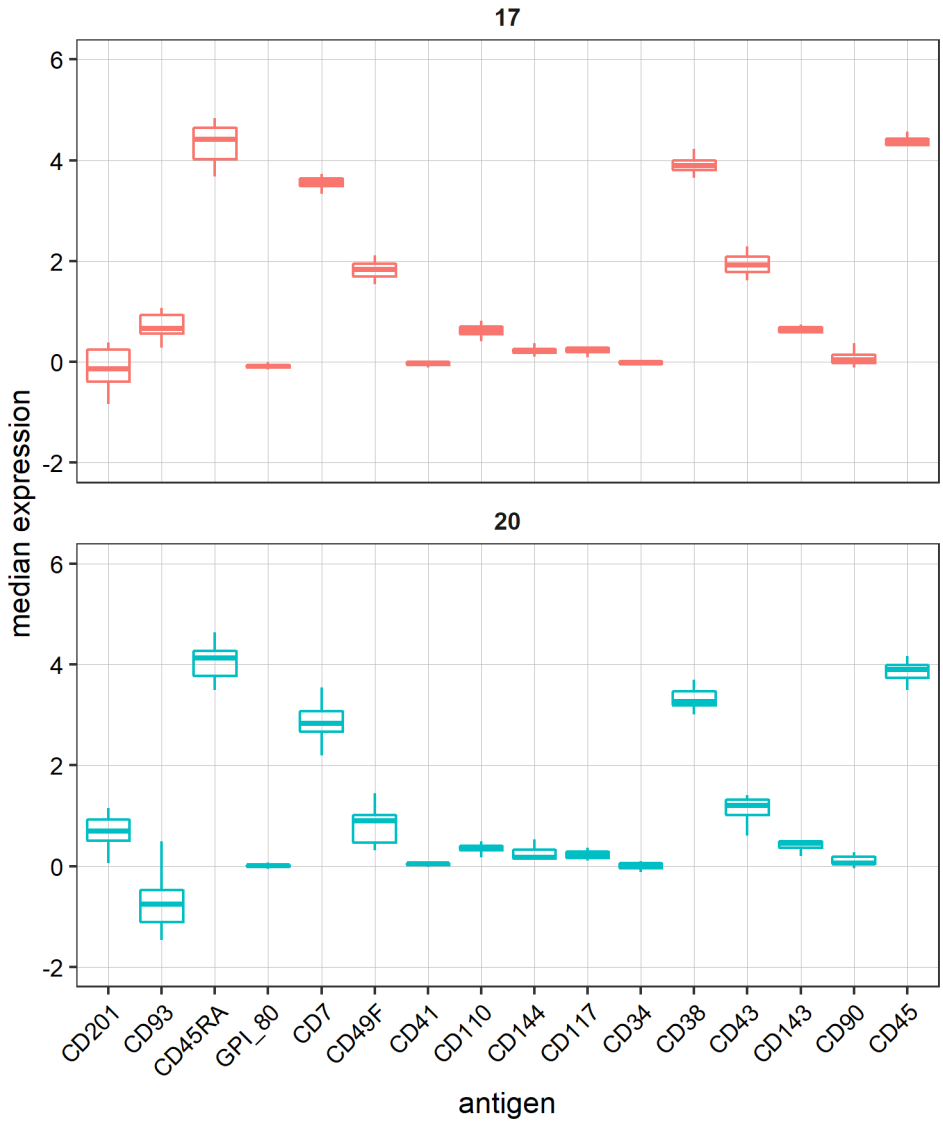

Figure 5.21. Median marker expression in FlowSOM clusters 16 and 20

Boxplots of median marker expression in C16 and C20. Horizontal line across boxes represents median expression across all samples.

C12 was a peculiar cluster as it was present in high proportion in L07 and then was abruptly lost until L14 (Fig. 5.14). It was observable in early FBM samples, BM12, BM13, BM15 and BM17, but decreased in relative size in BM18 and BM20. It appears to represent a $CD34^{Low}/CD38^{+}/CD45RA^{+}$ population (Fig. 5.22). CD41,

CD49f and CD7 were also expressed here. These group of markers are strange combination. $CD41^{+}/CD49f^{+}$ megakaryocytes have been described in the murine FL, but this population is also reported as $CD45^{-}$ (Cortegano et al., 2019). CD7, however, is a lymphoid marker, which has also been reported to be expressed by FL T cell precursors (Bárcena et al., 1993; Haddad et al., 2006). The sudden loss of C12 from FL after L07 may be due to these cells migrating to the foetal thymus (Haynes et al., 1988). This strange combination of cell surface markers in C12 may suggest a mix of populations that haven't been successfully resolved in the clustering.

C14 was present in all samples and stages but wasn't one of the most abundant clusters (Fig. 5.14). In the FBM, it became smaller with progressive gestational stages. It appeared mostly negative with low expression for CD38, CD41, and CD49f, bearing some resemblance to C12.

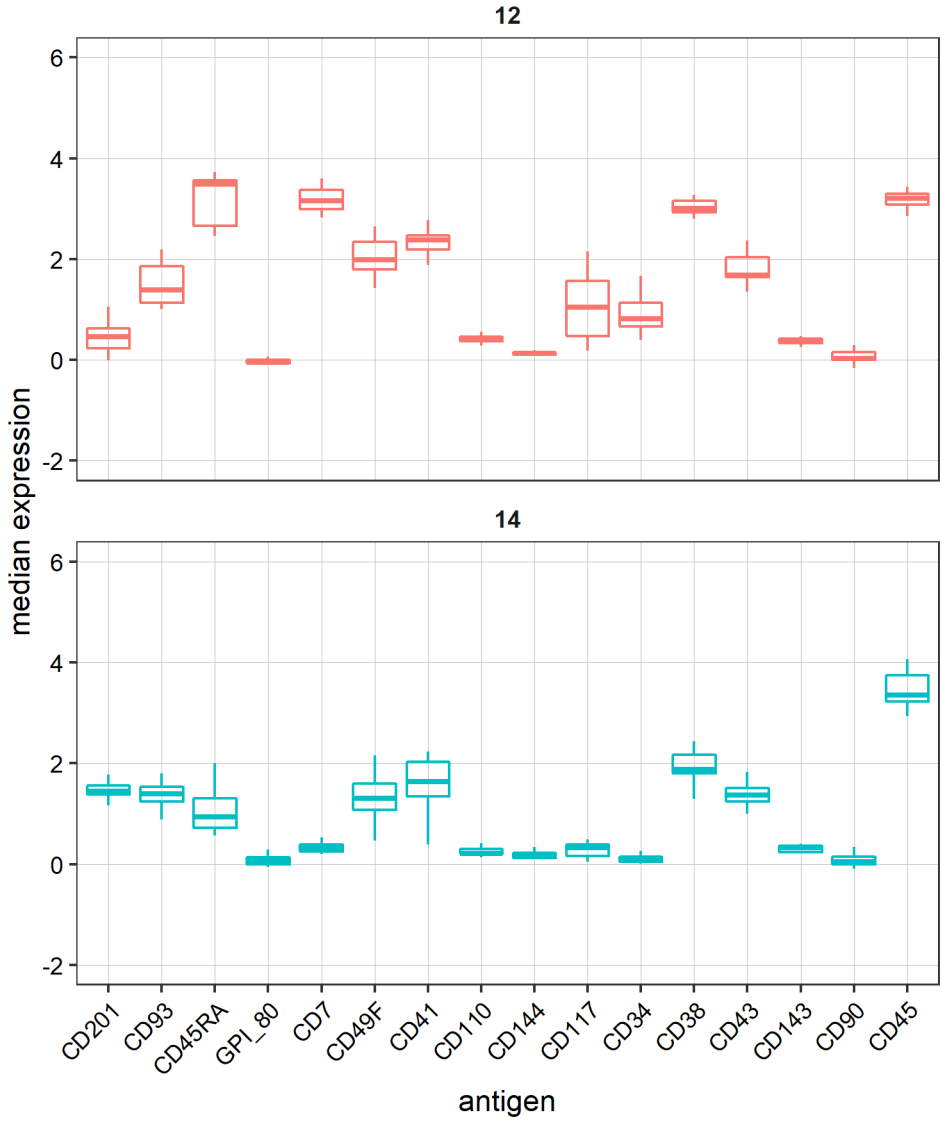

Figure 5.22. Median marker expression in FlowSOM clusters 12 and 14

Boxplots of median marker expression in C12 and C14. Horizontal line across boxes represents median expression across all samples.

5.7 Patterns of cluster marker expression across gestational stages

After establishing that populations defined by the markers in our panel dynamically changed over developmental stages in specific tissues, we proceeded to explore if marker expression also varies in a stage-dependent manner. We were particularly interested in looking at CD34, CD144 and CD117, which have been

reported to be expressed by HSCs in the early stages of FL colonisation. We also looked at CD143, another marker associated with endothelial lineages. In order to identify the top and bottom expressing clusters for each marker, we used median expression values scaled for each individual marker.

We observed CD144 and CD143 tended to be more highly expressed in FL samples. This was particularly true for CD144 which had high to intermediate expression in FL clusters 1, 3, 4, 8, 9, 15, 16 and 18. Its expression appeared to decrease in corresponding BM samples, but most of the clusters it was expressed in were either absent or consisted of very few cells in FBM samples (Fig. 5.23). CD143 expression persisted even in FBM samples for cluster 3, which persisted across FBM samples. In C3, we also observed that CD144 was quite high in early FL samples (L07, L08 and L10), but decreased in later stages, while CD143 expression increased and continued to be highly expressed in this cluster across FBM samples (Fig. 5.23). CD143 has been shown to be expressed in cord blood, adult and foetal human HSCs, while CD144 marks early foetal HSCs. It is possible that CD144 is steadily downregulated over time after HSCs colonise the FL in the first trimester, and upregulation of other markers begins to take place. C3 was also positive for CD117 and CD34. While their expression dynamic varied across samples, we noted CD144 and CD143 tended to be co-expressed. This was true for C3, 8, 9, and 15. However, CD144 appeared to be specific to C1, while CD143 was expressed at intermediate levels in C17.

C1 stood out as one of the clusters with the highest levels of CD117 expression across samples and stages. This cluster was also positive for CD144 in early FL samples, but appeared to express CD34 at low levels (Fig. 5.23).

C3 was of particular interest as it appeared to be $CD34^{+}/CD117^{+}/CD143^{+}$ and also $CD144^{+}$ in early FL samples. C10 lacked high expression for CD143 and CD144 but was double positive for CD34 and CD117. Finally, C15, which was observed to be mostly absent from FBM samples in the previous section, appeared to be $CD34^{+}/CD144^{+}/CD143^{+}$.

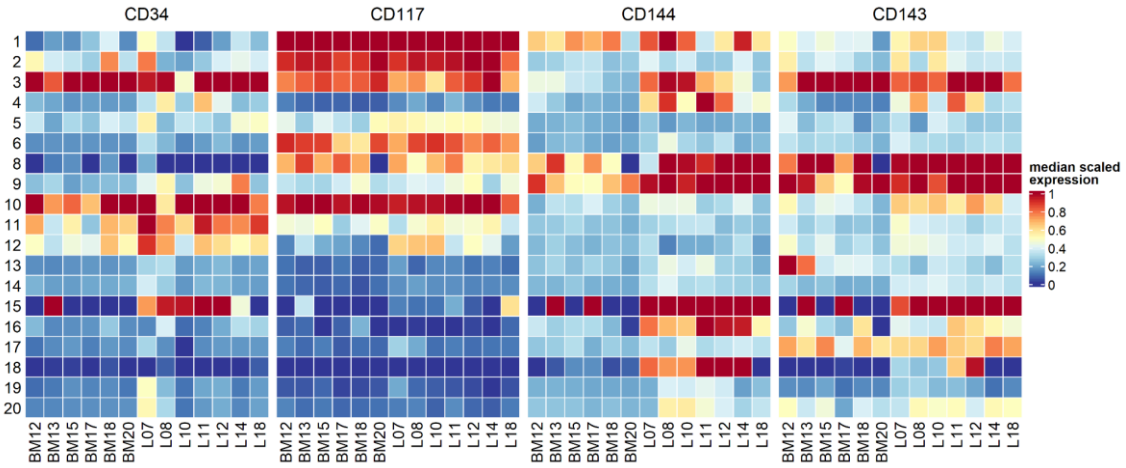

Figure 5.23. Scaled median marker expression for CD34, CD117, CD144 and CD143 across clusters and individual samples

Heatmap of the scaled median marker intensities of CD34, CD117, CD144, and CD143 for each cluster (row) and individual samples (columns). Expression is scaled per each individual marker. Zero expression values (dark blue squares) represent absent data points where clusters were not present in respective sample.

CD41 has been reported as one of the earliest markers of definitive HSCs in the murine embryo (Ferkowicz et al., 2003; Mikkola et al., 2003; Mitjavila-Garcia et al., 2002; Robin et al., 2011). In humans, however, reports have suggested CD43 may precede CD41 expression (Ivanovs, Rybtsov, Anderson, & Medvinsky, 2014; Vodyanik et al., 2006). CD41 was highly expressed in clusters 5, 10, 12 and to lesser degree C14 (Fig. 5.24). It overlapped with other early HSC markers CD34 and CD117 in C10, where CD43 was expressed as well (Fig. 5.24). In C3, CD43 expression overlapped with CD34, CD117 and CD143. However, CD43 was most highly expressed in C2 and C5. C2 was a highly abundant cluster that steadily increased in relative size in FL samples.

CD90, a well described marker of human foetal HSCs, correlated with CD144 in cluster 8, 9, 15, 16 and 18 (Fig. 5.24 and 5.23) (Ivanovs et al., 2011). Unlike CD144, CD90 continued to be highly expressed in FBM samples in C8 and C9 (Fig. 5.24). All of these clusters, however, were quite small across FBM samples.

Positive selection for CD110 has been shown to purify cord blood and adult bone marrow $CD34^{+}/CD38^{-}$ HSCs (Petit Cocault et al., 2016). Little is known about its role in human foetal haematopoiesis, but separation of murine FL HSCs by CD110 expression demonstrated it positively selects for *in vivo* repopulating cells (Solar et al., 1998). It was most highly expressed in clusters 8, 9 and 15. CD41 and CD110

expression both define megakaryocyte-erythroid lineages, but the two markers correlated only in one cluster, C9.

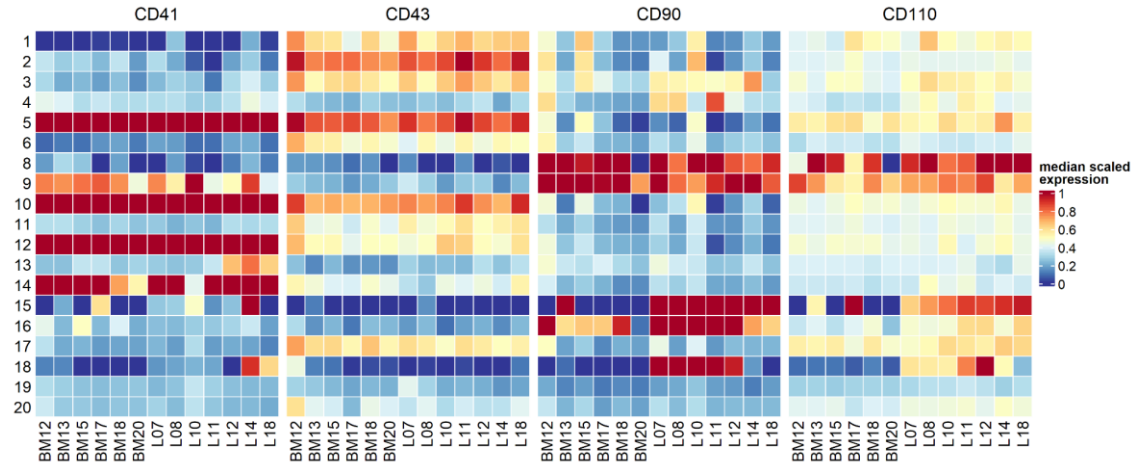

Figure 5.24. Scaled median marker expression for CD41, CD43, CD90 and CD110 across clusters and individual samples

Heatmap of the scaled median marker intensities of CD41, CD43, CD90 and CD110 for each cluster (row) and individual sample (column). Expression is scaled per marker. Zero expression values (dark blue squares) represent absent data points where clusters were not present in respective sample.

CD201 (EPCR), has been widely reported as an HSC marker in both mouse and human. In human FL, it has been shown to separate the $CD34^{+}/CD38^{-}/CD90^{+}/GPI\text{-}80^{+}$ population. Furthermore, $CD34^{+}/CD38^{-}/CD90^{+}/GPI\text{-}80^{+}/EPCR^{+}$ cells repopulated conditioned mouse recipients, while $CD34^{+}/CD38^{-}/CD90^{+}/GPI\text{-}80^{+}/EPCR^{-}$ cells did not (Subramaniam et al., 2019). We observed robust EPCR expression in C9 as well as C15 in FL samples (Fig. 5.25). This cluster was also positive for GPI-80, as well as CD90 (Fig. 5.25).

While well-described in postnatal haematopoiesis, the importance of CD49f in marking human foetal HSCs remains a question. Human mesodermal pre-haematopoietic cells have been reported to co-express CD49f and CD143, however, CD49f expression does not appear to persist in definitive HSCs (Julien et al., 2016). While CD49f can be expressed by YS-derived haematopoietic cells, these are $CD45^{-}$, and have been excluded from the analysis. In our dataset, CD49f was highly expressed in C6 and C15 and moderately expressed in C9, C10 and C12 (Fig. 5.25). Another marker expressed in C6 was CD117, while C15 expressed also a range of other markers (Fig. 5.23 & 5.24).

CD93, first identified in the murine foetal liver as a purifying HSC marker, was also highly expressed in C15 as well as C16 and C18. Both C16 and C18 also co-expressed CD144 and CD90, but the expression was restricted to FL samples only, while CD93 persisted in the FBM samples for these two clusters. C16 and C18, however, were quite small in FBM samples, and insufficient cell numbers could make expression readouts unreliable.

GPI-80 was most highly expressed in C8 and C9. However, both of these clusters were mostly absent from the FBM. GPI-80 has been well-reported as a purification marker for both FL and FBM HSCs (Prashad et al., 2015). In clusters 4, 13, and 14, GPI-80 expression was restricted only to FBM samples. While all three of these clusters decrease in size in FBM samples, C4 remains relatively large. The scaled expression values for C4 suggested the cluster had low CD34 expression (Fig. 5.23), but the spread of the unscaled CD34 expression distribution suggests this cluster may also contain $CD34^{+}$ cells (Fig. 5.11). C4 also expressed CD117 and CD143.

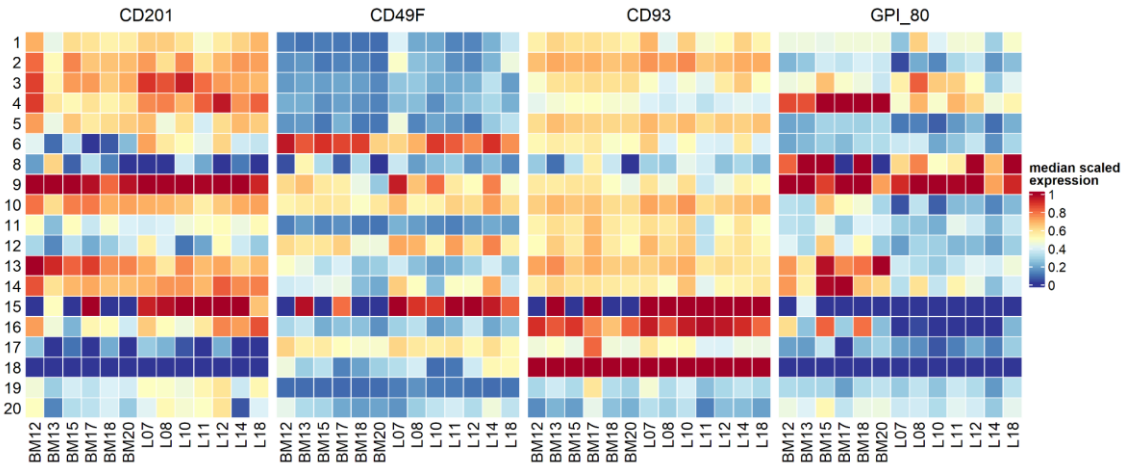

Figure 5.25. Scaled median CD201, CD49f, CD93 and GPI-80 marker expression

Heatmap of the scaled median marker intensities of CD201, CD49f, CD93 and GPI-80 for each cluster (row) and individual sample (column). Expression is scaled per marker. Zero expression values (dark blue squares) represent absent data points where clusters were not present in respective sample.

Expression of mature markers CD45RA and CD38 tended to overlap, encompassing clusters C11, C12, C17, C19, and C20. Intermediate CD38 expression also extended to clusters 1, 2, 8, 9 and 10. CD45RA had low to intermediate expression in C8 and C9, as well as FBM samples fir C10. Interestingly in C18, CD45RA expression was absent in most FL samples, but high in FBM samples. C18 was one of the smaller clusters in the analysis and also decreased in size with progressive stages both in the FL and FBM. By comparing relative size to CD45RA expression, we observed CD45RA expression was only high in samples where the size of this cluster was reduced to just a few cells (Fig. 5.26 & 5.14).

C19 was one of the largest clusters in FBM samples and increased in relative size through progressive stages in the FBM. We observed that CD38 expression was present in FL samples for this cluster at intermediate levels, but increased in the FBM (Fig. 5.26). CD34 appeared to be expressed at low levels, so it is likely this cluster represents multiple mature populations that cannot be resolved by the markers in our panel.

CD7 expression was restricted to clusters 11, 12, 17, and 20, which were also CD45RA^{+} and CD38^{+}, suggesting these represent mature populations potentially associated with the lymphoid lineage (Fig. 5.26).

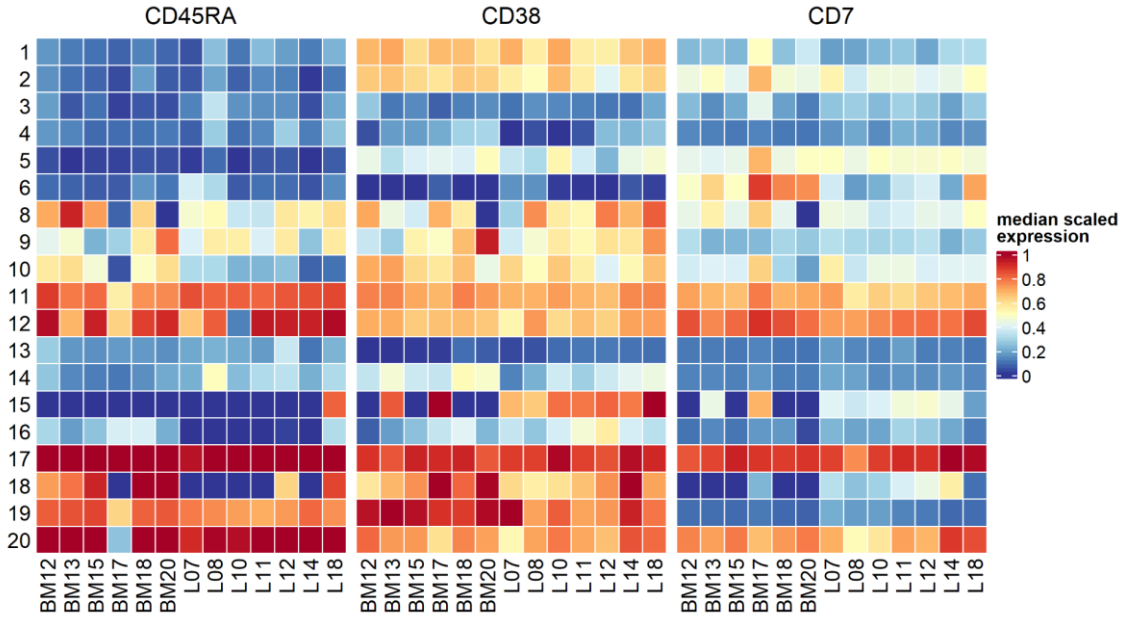

Figure 5.26. Scaled median CD45RA, CD38 and CD7 expression across clusters and individual samples.

Heatmap of the scaled median marker intensities of CD45RA, CD38 and CD7 for each cluster (row) and individual sample (column). Expression is scaled per each individual marker. Zero expression values (dark blue squares) represent absent data points where clusters were not present in respective sample.

By analysing the scaled expression of each marker in our panel across clusters and samples, we identified a number of populations of interest. C3 was quite an abundant cluster in both FL and FBM samples, but gradually decreased in relative size with progressive stages. Marker expression analysis suggested this population is $CD34^{+}/CD38^{-}/CD117^{+}/CD143^{+}$. CD144 was also expressed by these cells in early FL.

C10 appeared to be $CD34^{+}/CD117^{+}/CD41^{+}$, but there was potentially moderate expression of CD38 in this population as well. Furthermore, this cluster was quite rare in FBM.

C1 was present in both FL and FBM at similar levels. But while it co-expressed CD117 and CD144, this population also appeared to be $CD34^{Low}$ and $CD38^{+}$. As CD38 expression levels may be overestimated, and we are interested in potential candidate populations that may represent the equivalent of cord blood $CD34^{-}$ HSCs, C1 remains a population of interest.

5.8 Analysis of $CD38^{-}/CD45RA^{-}/CD7^{-}$ cell subsets in FL and FBM samples

As it was established by the non-redundancy scores in the previous section, the variability within individual samples was mostly attributed to the expression of CD45RA, CD38 and CD7. These are also markers associated with more committed cells and in order to focus the analysis onto potential HSC populations, we repeated the analysis by gating out all $CD45RA^{+}/CD38^{+}/CD7^{+}$ cells. While the $CD34^{-}$ fraction possibly contains large proportions of committed cells, we retained these to enable us to look for potential $CD34^{-}$ HSC candidate populations. We analysed approximately 5, 200 cells per sample.

Multidimensional scaling of the data was carried out, revealing strong changes in the intra-sample variability. The removal of $CD45RA^{+}/CD38^{+}/CD7^{+}$ cells resulted in bringing all bone marrow samples closely together (Fig. 5.27). In comparison FL samples were spread out in MDS dim. 1 by trimester. First trimester

FL samples appeared to cluster closer to FBM samples, but the two conditions were still separated by MDS dim. 2. The FL and FBM samples with the highest proximity were BM12 and L10. While L12 had been the most closely related sample to BM12 in the initial analysis, here it clustered away from FBM samples (Fig. 5.7 & 5.27). This indicates that the similarities between BM12 and L12 are likely driven by cells expressing CD45RA, CD38 or both.

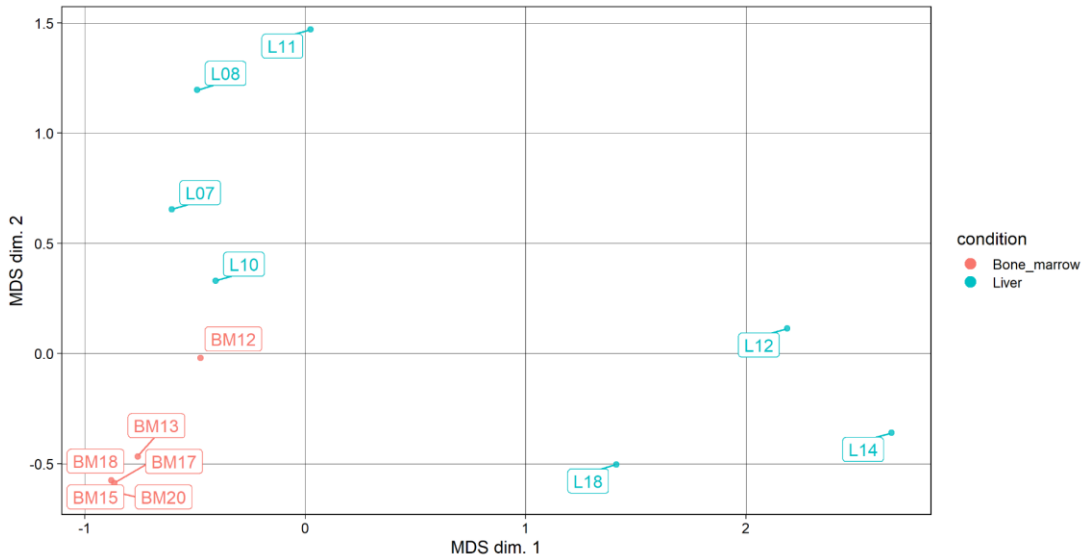

Figure 5.27. MDS plot for individual FL and FBM samples

MDS plot base on transformed median marker expression of 16 cell surface markers across all cells per sample. Plot labels represent sample IDs.

Next, we clustered the data, supplying a number of 12 clusters. One of the clusters amounted to approximately 56% of the whole dataset (Fig. 5.28A). This cluster was $CD34^-$ and had low median expression for CD201. By increasing the number of clusters, we could further explore it, however, as many as 20 clusters failed to resolve the population, suggesting that either the measured variables are not sufficient for this task, or even higher number of clusters may be required. Therefore, we filtered this population out.

In this analysis we noted four $CD34^+$ clusters (C1, C2, C11, and C12) (Fig. 5.28A). The unscaled expression distribution of CD34 across the different clusters suggested that C3 and C4 may also contain CD34 expressing cells (Fig. 5.28B). In addition, C12 also expressed CD201, CD90, CD144 and CD49f Fig. 5.28A & B). C11 had similar expression profile but at lower intensity, while C1 also expressed CD117

and CD41. C2 and C11 were also two of the biggest clusters (Fig. 5.28A). The rest of the clusters could be defined as $CD34^-$. C7 had one of the highest median expressions for CD90, while C5 was highly positive for CD49f. C8 was positive for both CD41 and CD43.

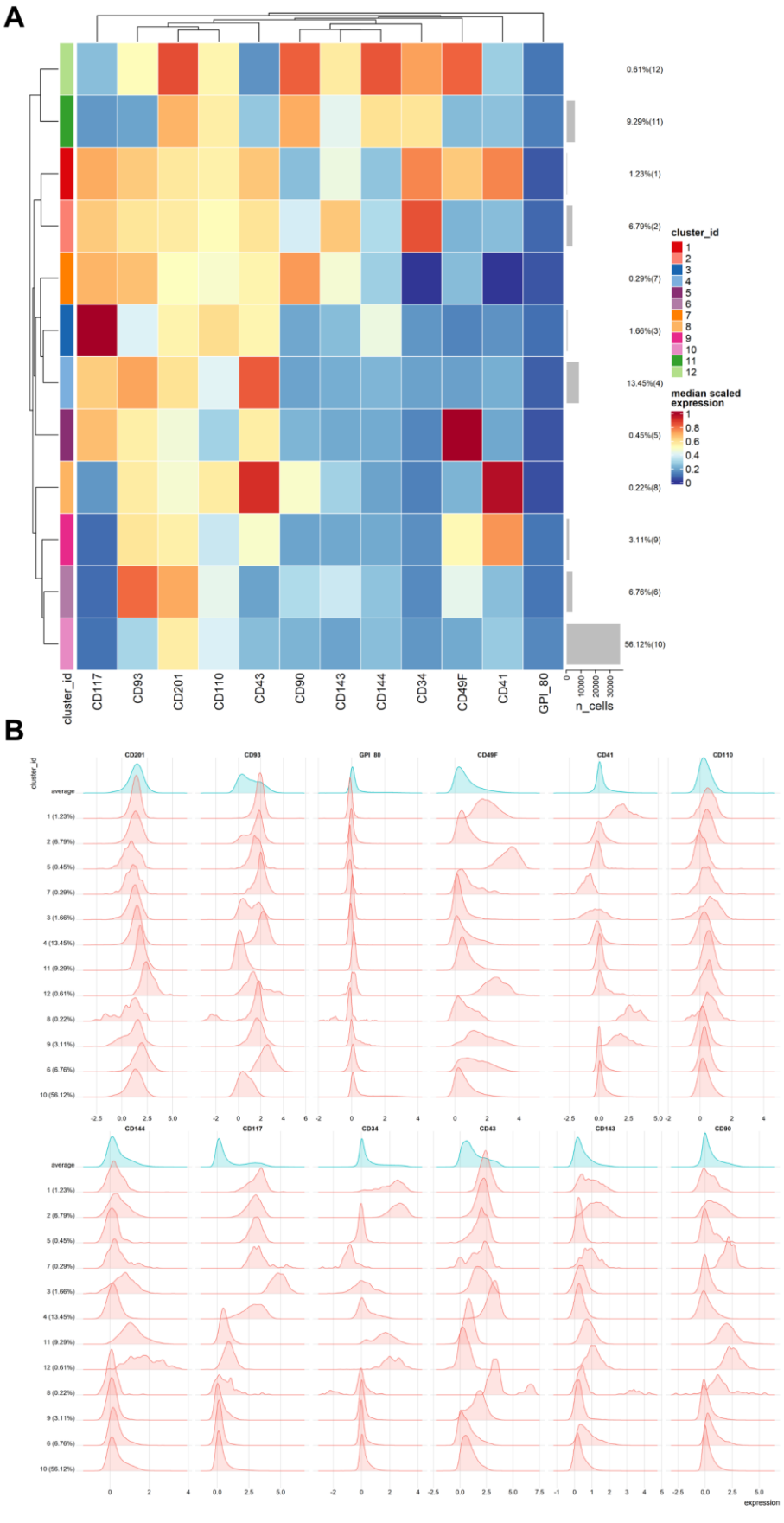
A
0.61%(12)
9.29%(11)
1.23%(1)
6.79%(2)
0.29%(7)
1.66%(3)
13.45%(4)
0.45%(5)
0.22%(8)
3.11%(9)
6.76%(6)
56.12%(10)
cluster_id
1
2
3
4
5
6
7
8
9
10
11
12
median scaled expression
1
0.8
0.6
0.4
0.2
0
cluster_id
CD117
CD93
CD201
CD110
CD43
CD90
CD143
CD144
CD34
CD49F
CD41
GPI_80
0
10000
20000
30000
n_cells
B
CD201
CD93
GPI 80
CD49F
CD41
CD110
CD144
CD117
CD34
CD43
CD143
CD90
cluster_id
average
1 (1.23%)
2 (6.79%)
5 (0.45%)
7 (0.29%)
3 (1.66%)
4 (13.45%)
11 (9.29%)
12 (0.61%)
8 (0.22%)
9 (3.11%)
6 (6.76%)
10 (56.12%)
expression

Figure 5.28. FlowSOM clusters and corresponding marker expression profiles

A. Heatmap of scaled median transformed marker expression across all cells and samples. Dendograms represent hierarchical similarity between all clusters (rows) and markers (columns). Bar plots on the right side represent relative sized of clusters with absolute values in brackets.
B. Red ridge plots represent marker expression distribution per individual clusters. Blue ridge plots represent average expression distribution of marker across all clusters.

Similarly to the unfiltered dataset, cluster dynamics varied across stages and tissue types. A few clusters however appeared to have a similar dynamic in both tissues. C2, C3, and C4 were present at similar proportions in both tissue types and increased with progressive stages (Fig. 5.29A). C11, the biggest cluster in first trimester FL sampels, increased until L11, before dropping sharply in size. In the FBM, it constituted approximately 25% of all cells before it was gradually outsized by other clusters. It sharply drops in size at L11 in the FL while it is quite large in BM12, which could indicate that C11 constitutes a population that rapidly colonises the FBM between 10 and 12 pcw. This kind of dynamic was also observed for C6. While C11 was $CD34^+$, C6 was $CD34^-$ and expressed CD93 and CD49f (Fig. 5.28A & B).

C9 was absent in most first trimester FL samples, but was surprisingly abundant in the FBM (Fig. 5.29A & B). From 12 pcw it expanded in relative size, peaked in FBM17 and began decreasing again. C9 is $CD34^-/CD41^+$ with some expression of CD43 and CD49f (Fig. 5.28A & B).

While smaller in size, C7 and C12 stood out as two clusters that were specific to the FL (Fig. 5.29A & B). C12 was highly positive for CD34, CD144, CD49f and CD201 (Fig. 5.28A & B). On the other hand, C7 was the most negative cluster for CD34, but was positive for CD117 (Fig. 5.28A & B). The unscaled expression distribution also suggested there may be some $CD144^+$ and $CD143^+$ in this cluster (Fig. 5.28B).

C1 was relatively small cluster present in the FL (Fig. 5. 29A & B). In the FBM is apeared to be a rarer population. C1 was a cluster of special interest as scaled median marker expressions suggested its phenotype as $CD34^+/CD144^+/CD143^+/CD90^+/CD201^+$ (Fig. 5.28A). Unscaled expression distribution also indicated this cluster is likely to be CD117-expressing as well (Fig. 5.28B).

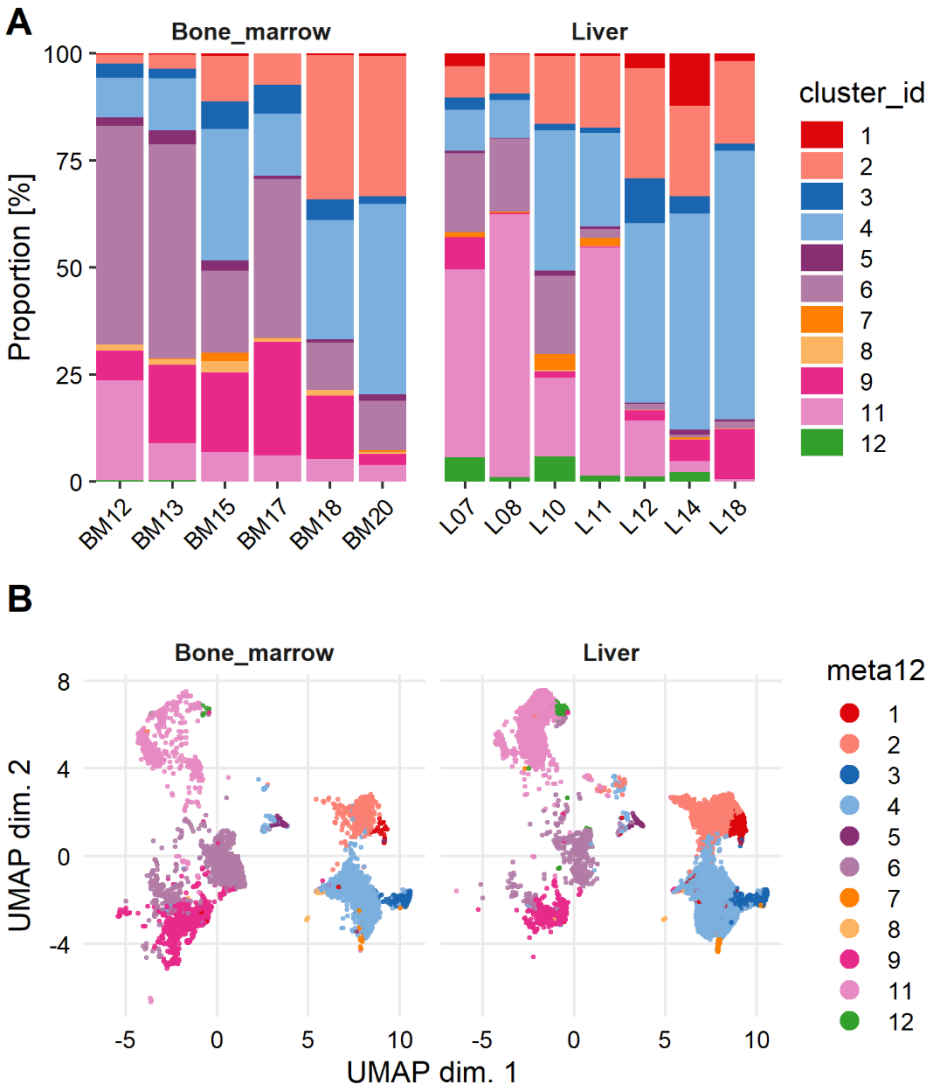

Figure 5.29. Cluster dynamics across sample sand dimensionality reduction

A. Stacked barplots of 11 FlowSOM clusters across samples stratified by tissue type (left panel – FBM; right panel – FL).
B. UMAP projections coloured according to the 11 FlowSOM clusters stratified by conditions (Bone_marrow & Liver). Cluster colour have been amended as a result of the removal of two clusters, but the numbering remains the same.

As we observed some potential HSC populations in FL, but not FBM samples, we supposed these populations may change their marker phenotype prior or at the time of FBM colonisation. We undertook multidimensional scaling analysis of the different clusters across samples in order to estimate the degree of similarity between clusters.

C12, which was specific to the FL, clustered closely together with C11 (Fig. 5.30). C7 clustered tightly with C2, indicates the two clusters may represent the same population. The main differences between that C7 was a very small cluster and also appeared to be $CD34^-$ (Fig. 5.28A & 5.29A). C8 was quite spread out across to MDS dimensions, which is probably due to the high spread of expression for multiple markers (Fig. 5.30 & 5. 28B). Overall the data was split in two by MDS dim. 1. One group contained clusters 1, 2, 3, 4, 5, and 7, while the other was composed of clusters 6, 11 and 12 with cluster 9 localising apart (Fig. 5.30). The first group contained all the clusters that explanded in relative size over time or remained stable. The second group contained the clusters that decreased in relative size over time.

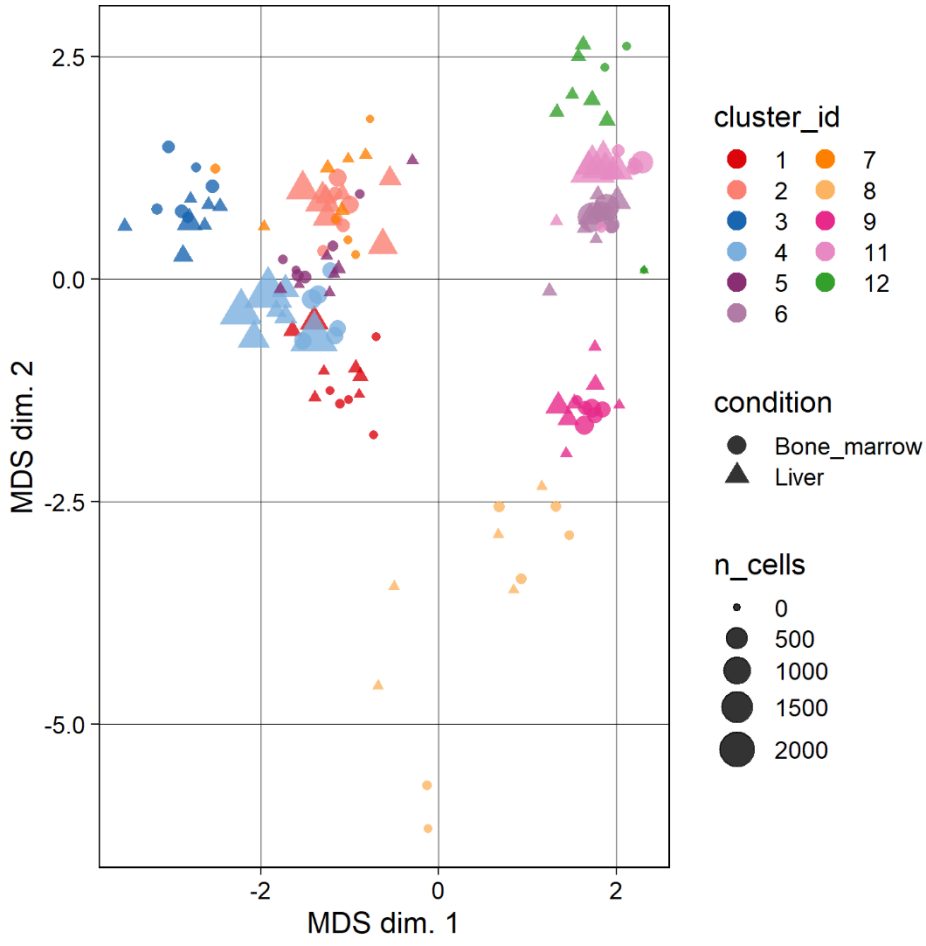

Figure 5.30. Multidimensional scaling plot for FlowSOM clusters

MDS plot on pseudobulks computed for each cluster and sample. Points are coloured by their cluster identity and shaped by condition.

By subsetting the data collected across FL and FBM of different stages for $CD45RA^-/CD38^-/CD7^-$ cells, we carried out a more focused search for potential HSC populations. We identified clusters that co-expressed a number of the markers in our panel in different combinations. By subsetting the data, we also showed that CD45RA and/or CD38 expressing cells are likely to be the populations contributing to the similarity between second trimester FL samples and the FBM. When we excluded these cells from the analysis, we revealed that L12, L14, and L18 were quite different from the rest of the samples, which strongly suggests that rather in waves, immature HSC and progenitor populations colonise the liver rapidly before the onset of the second trimester.

5.9 *In vivo* validation of FL and FBM populations identified by multiparameter flow cytometry

We identified a number of clusters of interest in the previous section. We noted both in the full and subset data the existence of clusters that appeared to co-express a range of markers from our panel. One of such marker, CD117, was one of the brightest markers in the dataset (Fig. 5.28B). We also observed that CD144 and CD143 were often co-expressed, but CD144 median expression tended to decrease in later stages. The three markers were also co-expressed with C1, C2 and C3 in the analysis of the $CD45RA^-/CD38^-/CD7^-$ data subset.

To begin with, we decided to target one population, C3 of the subset analysis. We chose this cluster because it has the brightest CD117 population. Based on previous experience on using this marker, we were confident we could identify this population using our standard gating strategy. We ran two independent experiments, using a 13 pcw FL and 17 pcw FBM sample. We decided on these samples as C3 appeared in both tissue types and across all stages. We gated for live, single, $CD45^+/Lin^-/CD45RA^-/CD34^+/CD38^-$ and gated for the highest-expressing CD117 cells in this population, using a CD117 FMO as a guide (Fig. 5.31). This population will be referred as $CD34^+/CD38^-/CD117^{Hi}$ from here on. It should be noted that the CD117 gates in this part of the project were higher compared to the ones we used to sort for $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ population in Chapter 3.

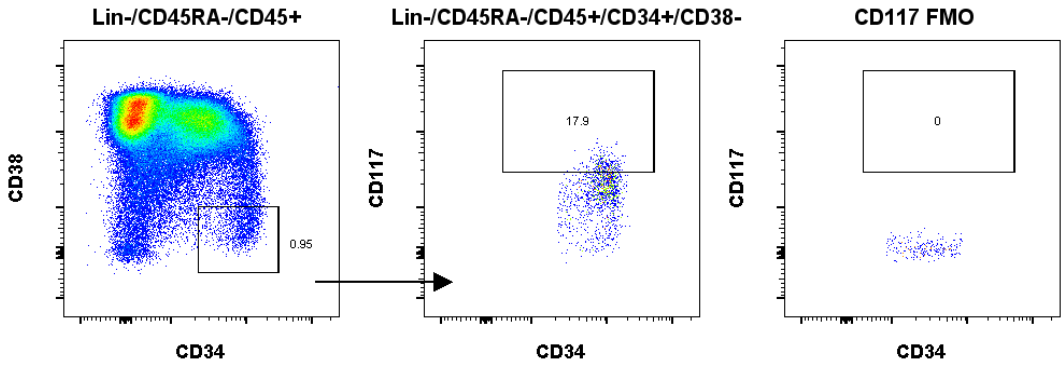

Figure 5.31. Gating strategy for sorting $CD34^{+}/CD38^{-}/CD117^{Hi}$ cells

Representative flow cytometry plots of gating strategy for $CD34^{+}/CD38^{-}/CD117^{Hi}$ cells (left and middle panel) and CD117 FMP control (right panel). Sample used in plots is 18 pcw FBM.

Cells were transplanted in $C\text{-}Kit^{Mutant}$ mice as previous work with this strain had suggested, they can sustain human repopulation at high levels and over long periods of time. Engraftment and lineage distribution were tracked via bone marrow aspirations every six weeks and assayed using flow cytometry. The first experiment we conducted was with 13 pcw FL, and we culled the mice 12 weeks post-transplantation. All subsequent experiments were carried out for 24 weeks.

We observed high engraftment in all mice injected with $CD34^{+}/CD38^{-}/CD117^{Hi}$ cells from both 13 pcw FL and 18 pcw FBM. The percentage of human $CD45^{+}$ cells in all assayed mice was over 60% as early as 6 weeks post-transplantation (Fig. 5.32A). Mice transplanted with 13 pcw FL were culled at 12 weeks, but engraftment persisted up to 24 weeks post-transplantation in mice repopulated with 18 pcw FBM cells. Lineage distribution in mice that received 13 pcw FL cells indicated higher proportion of $CD33^{+}$ myeloid cells compared to $CD19^{+}$ B cells (Fig. 5.32B). Almost no $CD3^{+}$ cells were present in the bone marrow, but these cells could be detected in the spleens which were also highly repopulated with human $CD45^{+}$ cells (Fig. 5.32A & B). In mice transplanted with 18 pcw FBM cells, the proportions of $CD19^{+}$ B cells were higher than the myeloid cells in the first two time points at 6 and 12 weeks following injections (Fig. 5.32C). At later stages, $CD33^{+}$ myeloid cells became more abundant than the B cells. In the spleen we observed a significantly higher engraftment of $CD3^{+}$ T cells compared to 13 pcw FL mice, but this difference may be due to the longer duration of the experiment of 17 pcw FBM cells (Fig. 5.32B & C).

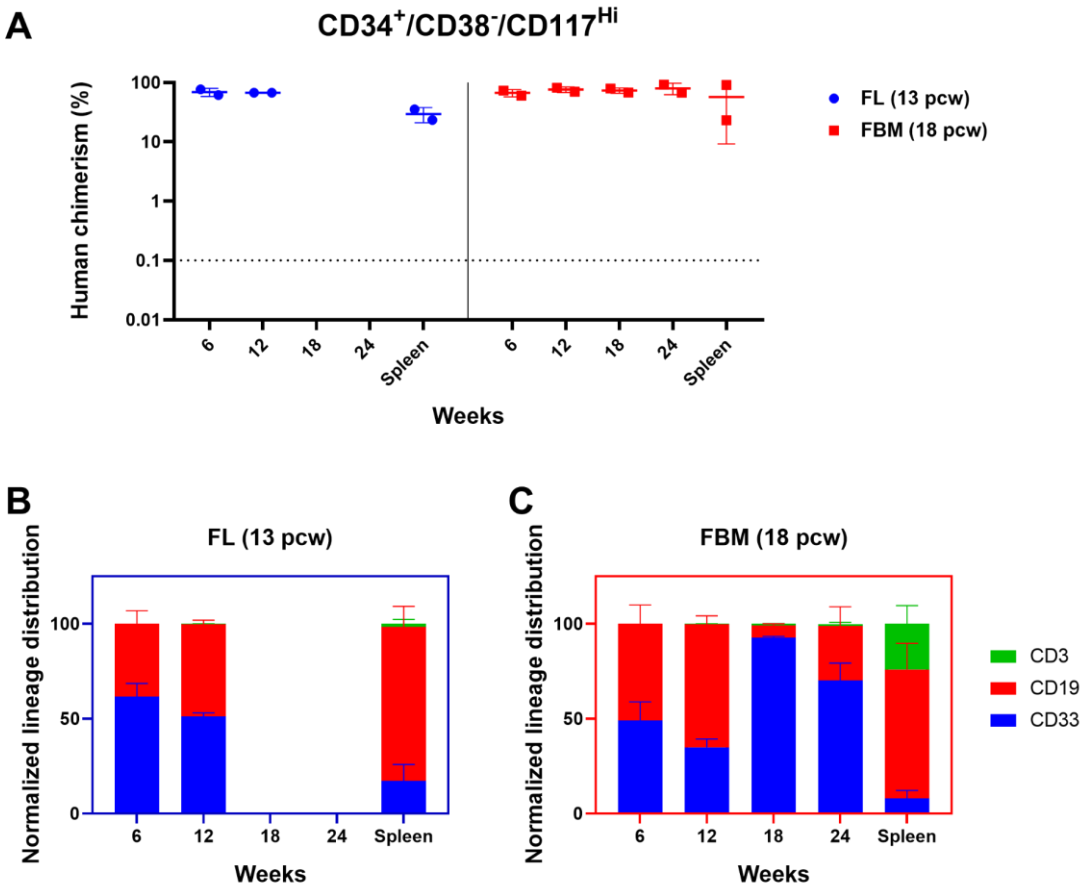

Figure 5.32. CD34^{+}/CD38^{-}/CD117^{Hi} FBM and FL cells repopulate C-Kit^{Mutant} mice

A. Quantification of human chimerism as percentage of human CD45^{+} (hCD45^{+}) cells of all single live cells. NSG-c-Kit^{Mutant} mice were injected with 4, 000 CD34^{+}/CD38^{-}/CD117^{Hi} cells from 13 pcw FL (n=2; blue) or 3, 000 CD34^{+}/CD38^{-}/CD117^{Hi} cells from 18 pcw FBM (n=2; red).
B-C. Quantification of lineage distribution normalised to total lineage-positive cells in NSG-c-Kit^{Mutant} mice injected with CD34^{+}/CD38^{-}/CD117^{Hi} cells from 13 pcw FL (B) and 18 pcw FBM (C).

After culling the mice transplanted with 13 pcw FL cells 12, we undertook secondary transplantation to assay the long-term reconstitution capacity of the foetal CD34^{+}/CD38^{-}/CD117^{Hi} population. We only used one of the transplanted mice, for the human CD45 selection. We split the yielded cells and injected two secondary mouse recipients, effectively splitting the graft from one donor into two recipients. We also carried out staining for CD34, CD38 and CD117 to assess the marker phenotype of the population after 12 weeks of primary reconstitution and found that the expression distribution of CD34 and CD38 had been conserved through the

duration of the primary reconstitution, and CD117 expression in the $CD34^{+}/CD38^{-}$ population had been maintained (Fig. 5.33A).

Both of the secondary $C\text{-}Kit^{Mutant}$ mouse recipients were repopulated with human cells at 6 and 12 weeks following the secondary transplantation (Fig. 5. 33C). We also observed multilineage engraftment pattern similar to that observed in the primary recipients (Fig. 5. 33D, 5.33B). Furthermore, the distribution of CD34 and CD38 expression following 12 weeks of secondary transplantation still resembled the phenotype at the end of the primary reconstitution and primary samples (Fig. 5.31A-B).

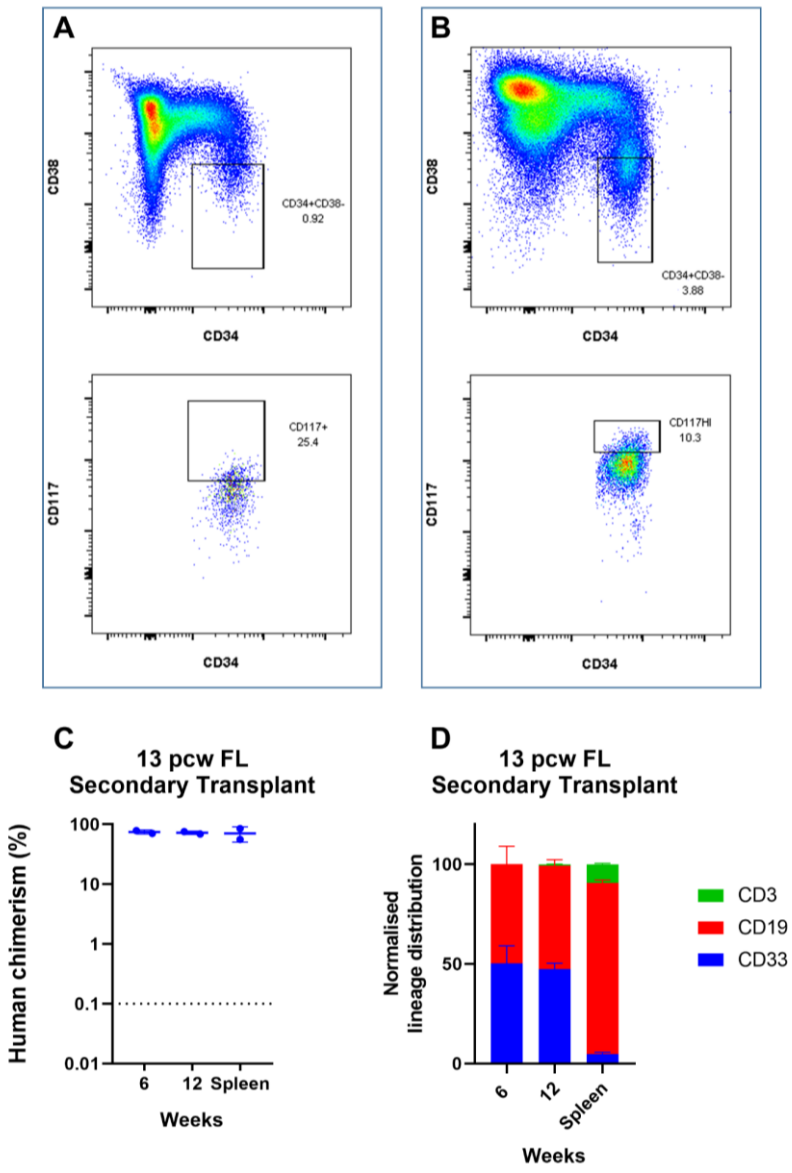

Figure 5.33. $CD34^{+}/CD38^{-}/CD117^{Hi}$ 13 pcw FL cells have long-term reconstitution capacity

A-B. Representative flow cytometry plots for CD34, CD38 and CD117 expression at the end of primary (A) and secondary (B) reconstitution with 13 pcw FL $CD34^{+}/CD38^{-}$ $CD117^{Hi}$ cells.
C. Quantification of human chimerism as percentage of human $CD45^{+}$ ($hCD45^{+}$) cells of all single live cells. C-Kit^{Mutant} mice injected with 25, 000 $hCD45^{+}$ from a single primary xenograft generated by 4, 000 13 pcw FL $CD34^{+}/CD38^{-}/CD117^{Hi}$ cells (n=2).
D. Quantification of lineage distribution normalised to total lineage-positive cells in secondary C-Kit^{Mutant} recipients.

We successfully demonstrated that high CD117 expression successfully purified $CD34^{+}/CD38^{-}$ repopulating cells derived from 13 pcw FL and 18 pcw FBM. In addition, we showed these populations are capable of multi-lineage engraftment. For the population from 13 pcw FL, we also carried out a 12-week plus 12-week serial transplantation, which demonstrated robust multilineage reconstitution in secondary recipients. This experiment, however, was based on a single FL sample and one primary xenograft. In order, to confirm the long-term repopulation capacity of this population more experiments are required. In addition, the duration of the serial transplantations would have to be extended to truly test the long-term self-renewal potential of the $CD34^{+}/CD38^{-}/CD117^{Hi}$ foetal HSCs.

As a first step, in verifying the hypothesised HSC clusters from our multi-parameter marker analysis in the FL and FBM, we began by sorting for a $CD34^{+}/CD38^{-}$ population expressing high levels of CD117. This specific cluster was identified as C3. In addition to being present in both FBM and FL samples, it also expressed CD144. We added CD144 to our panel and sorted for $CD34^{+}/CD38^{-}/CD117^{Hi}/CD144^{+}$ and $CD34^{+}/CD38^{-}/CD117^{Hi}/CD144^{-}$ to investigate if CD144 expression could further stratify this population (Fig. 5.34).

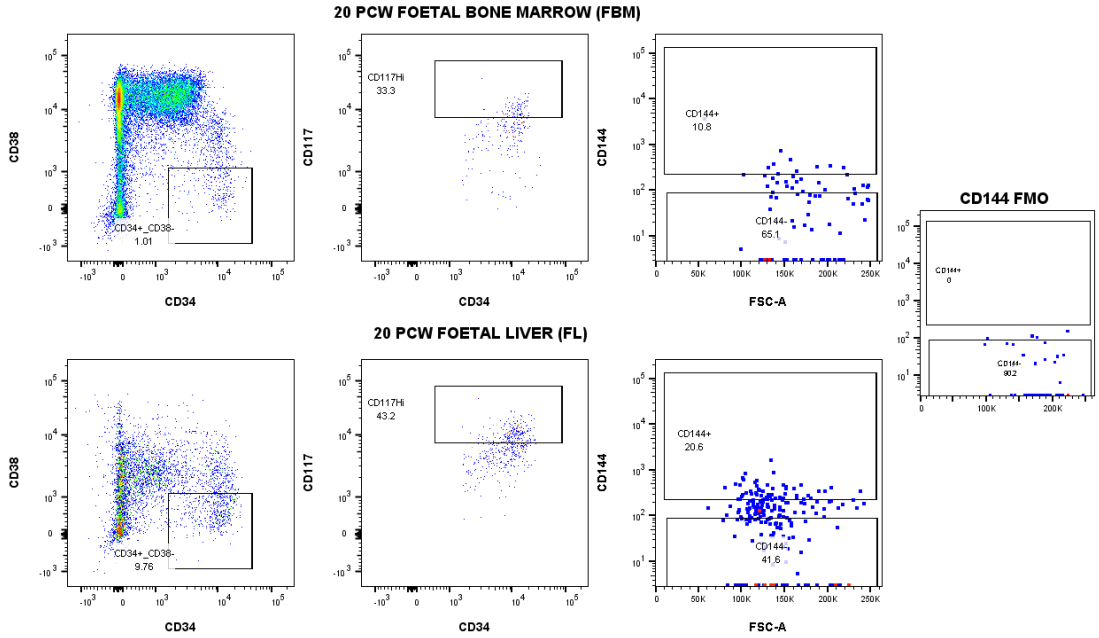

Figure 5.34. Gating strategy for the isolation of CD34+/CD38-/CD117Hi/CD144+/- cells from FBM and FL samples

Representative flow cytometry plots for gating strategy $CD34^{+}/CD38^{-}/CD117^{Hi}/CD144^{+/-}$ cells in 20 pcw FBM (top panels) and 20 pcw FL (bottom panels) plus CD144 FM0 control.

We used 20 pcw and 18 pcw FBM and FL samples for these repopulation experiments, but the mice injected with 18 pcw FL cells had to be culled soon after the start of the experiment. Following the transplantation of 20 pcw FBM populations, we tracked the engraftment every 6 weeks from the twelfth-week time point. We observed engraftment in all injected mice. FBM $CD144^{+}$ cells appeared to have the lowest engraftment level, though we lost one mouse before the 24-week time point (Fig. 5.35). However, these mice were injected with very low cell doses (Table 5.5), which could also explain the lower engraftment. $CD144^{-}$ FBM cells repopulated C-Kit^{Mutant} mice in both experiments, but the engraftment was only detectable from 18 weeks onwards. At 24 weeks the engraftment in these mice was quite high, with presence of myeloid, B and T cells (Fig. 5.35A-B). In the FL, we detected robust high engraftment at every time point for both cell populations, but the lineage distributions were distinct. The mouse injected with $CD144^{+}$ FL cells had high proportion of $CD33^{+}$ myeloid cell repopulation, which was similar to what we observed in the $CD144^{+}$ FBM experiments. But unlike the mice repopulated with $CD144^{+}$ FBM cells, $CD3^{+}$ T cell proportions were quite high in the spleen of the $CD144^{+}$ FL mouse (Fig. 5.35B). We

also found high degree of similarity between the lineage distribution of CD144$^-$ FBM and CD144$^-$ FL repopulated mice. The mouse injected with FL cells, however, established engraftment earlier.

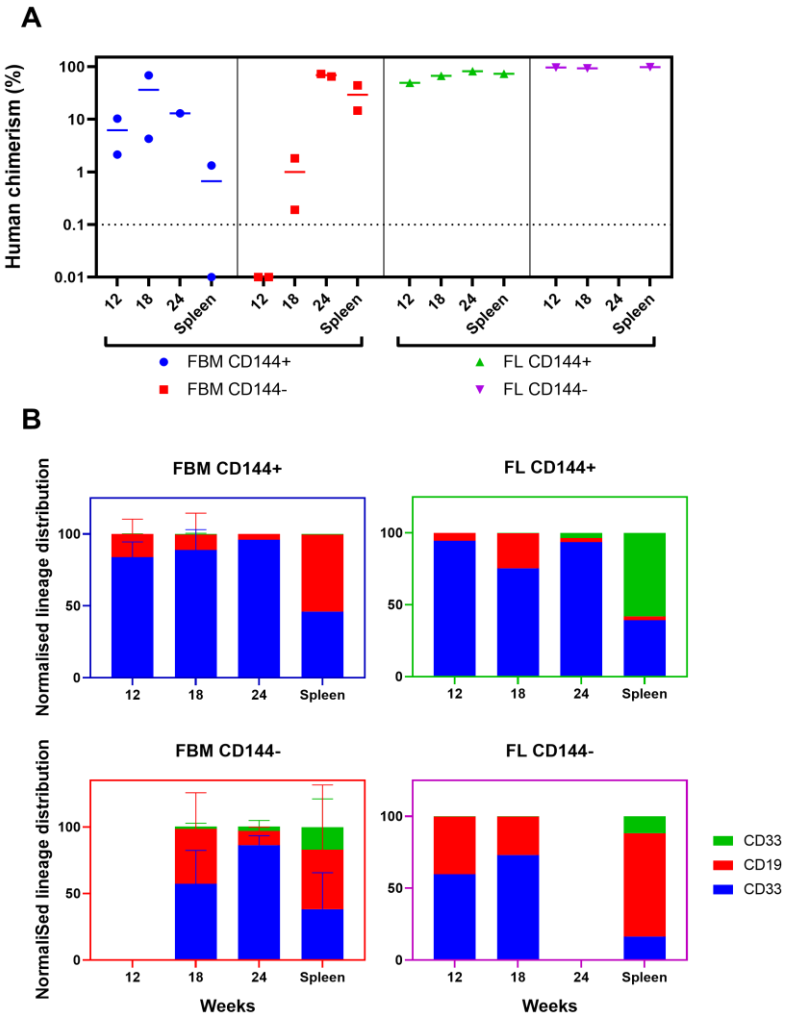

Figure 5.35. Repopulation kinetics and lineage distribution of second trimester FBM and FL CD34$^+$/CD38$^-$/CD117^{Hi}/CD144$^{+/-}$ cells

A. Quantification of human chimerism as percentage of human CD45$^+$ (hCD45$^+$) cells of all single live cells. C-Kit^{Mutant} mice injected with FBM and FL CD34$^+$/CD38$^-$/CD117^{Hi}/CD144$^{+/-}$ cells. See Table 5.5 for cell doses and sample staging.
B. Quantification of lineage distribution normalised to total lineage-positive cells in secondary C-Kit^{Mutant} recipients. Colours of plot borders correspond to panel A.

Phenotype	**Tissue Origin**	**Stage**	**Cell Dose**
CD144^{+}	FBM	20	160
	FBM	18	170
	FL	20	1450
CD144^{-}	FBM	20	831
	FBM	18	200
	FL	20	3400

Table 5.5. Cell dose numbers for repopulation experiments

By conducting *in vivo* repopulation assays in C-Kit^{Mutant} mice, we have demonstrated that high CD117 expression selects for a population with primary and secondary repopulation potential. We have also demonstrated successful multi-lineage engraftment of as little as 160 CD34^{+}CD38^{-}/CD117^{Hi}/CD144^{+} from second trimester FBM. The long-term self-renewal capacity of these cells, however remains to be tested. Additional replicates and samples from earlier stages would also be required to confirm the existence of these populations in haematopoietic development.

5.10. Chapter Summary and Discussion

In this chapter we have presented the optimisation and execution of high-parameter flow cytometry antibody panel, aiming to advance our understanding of the expression patterns of reported HSC markers in human foetal haematopoiesis. Using clustering algorithms and dimensionality reduction, we successfully identified a CD34^{+}/CD38^{-}/CD117^{Hi}/CD144$^{+/-}$ FBM and FL population with *in vivo* repopulation potential. We also conducted a secondary transplantation experiment that provided preliminary evidence CD34^{+}/CD38^{-}/CD117^{Hi} FL cells may have long-term self-renewal capacity. Further validations will be required to confirm this. Our data analysis of the subset data in section 5.8 of this chapter also indicated the presence of two clusters that appeared to contain CD34^{-} populations with some expression of CD93 and CD117. Compelled by the work conducted in Chapter 3 of this thesis, we attempted repopulation assays by sorting for different populations within the CD34^{-}/CD38^{-} compartment of FL and FBM samples in parallel with the work in section 5.9.

No multilineage human engraftement was detected in any of the mice transplanted with $CD34^-$ cells 24 weeks post-transplantation (Table. 5.6). These experiments suggested that the $CD34^-$ compartment in human FL and FBM may not be enriched for $CD34^-$ SRC between 13 and 20 pcw. Further experiments with larger cell doses may be required to confirm this. For $CD34^-/CD38^-/CD93^+/CD117^+$ derived from CB, we pooled MNCs from multiple donors and carried out high-efficiency lineage depletion before FACS. For our experiments on FL and FBM, samples were not pooled and committed cells were excluded only during FACS using a standard Lineage antibody cocktail. While it is possible that the $CD34^-$ HSCs emerge later in gestation, it could be that they are too rare to be efficiently captured in individual foetal samples without pre-processing steps that exclude committed cells and without larger starting cell quantities.

Stage (pcw)	**Tissue**	**Phenotype**	**Cell dose**	**Number of mice**
18	FBM	$CD34^-CD38^-CD117^+$	18,000	2
20	FL	$CD34^-CD38^-$ $CD93^+CD117^+$	11000	1
20	FBM	$CD34^-CD38^-$ $CD93^+CD117^+$	170	1
17	FBM	$CD34^-CD38^-$ $CD93^+CD117^+$	94	1
18	FBM	$CD34^-CD38^-$ $CD93^+CD117^+$	165	1
18	FBM	$CD34^-CD38^-$ $CD93^+CD117^+$	500	2
13	FL	$CD34^-CD38^-$ $CD93^+CD117^+$	4000	2
20	FL	$CD34^-CD38^-CD93^-$ $CD117^+$	50000	1
20	FBM	$CD34^-CD38^-CD93^-$ $CD117^+$	18000	1
17	FBM	$CD34^-CD38^-CD93^-$ $CD117^+$	7290	1
18	FBM	$CD34^-CD38^-CD93^-$ $CD117^+$	4814	1

Table 5.6. List of CD34^{-} FL and FBM populations and respective cell doses tested in C-Kit^{Mutant} for repopulation capacity

While much more powerful techniques for studying single cell heterogeneity such as scRNAseq and mass cytometry have become the norm, we have demonstrated that high-parameter flow cytometry can still be an effective tool in identifying populations of interest. Still, we cannot discount some inherent limitations that come with such analyses. One major challenge in flow cytometric data is spectral overlap. Spectral overlap, or “spillover”, results from the use of fluorochromes that are measurable in more than one detector. Compensation is meant to correct for spillovers of all fluorochormes used in a flow cytometry run, so that the collected data are normalised and each parameter contains information from a single fluorochrome. Compensation is a powerful tool, but the challenge of spillover becomes more severe with every additional parameter and interaction between parameters cannot be discounted in large panels even with compensation.

While efforts have been made to expand the measuring capacity of flow cytometry instruments with the creation of new fluorochormes, little has been done to improve the mathematical methods for compensation corrections. Compensation also does not take into account cell-fluorochrome interactions that can interfere with excitement and emission spectrum of a dye. To complicate matters further, beads, which tend to stain more brightly than cells, are used in automatic compensations, resulting in a compensation matrix that does not always accurately normalise the actual data acquired.

New flow cytometry instruments capable of measuring over 30-parameters are now commercially available, but the aims of individual experiments and the nature of the parameters examined must be carefully considered in designing 30-plus panels. In this chapter we showed that even in a 17-colour panel, challenges can be encountered at pre-processing. For some markers, it was apparent different parameters may have been better suited for the data transformation. Some online tools developed for mass cytometry recommend that separate transformations are applied for each individual marker in a dataset. However, this is not the case in the packages used in this analysis. The use of different transformation settings for individual parameters in a dataset may also introduce a level of bias to the data.

Mass cytometry (MC) has several advantages over flow cytometry. It is more permissive of larger panels as the issue of “spillover” is not as pronounced. While sample runs can take significantly longer in MC, methods for sample indexing have been developed so multiple samples can be run at the same time. Some caveats in MC, however, is that choice of antibodies is still more limited and samples are lost during the data collection. Fluorescence-activated cell sorting remains the most reliable method to isolate cells based on cell surface marker expression. Undoubtedly, new developments in engineering and chemistry coupled with improvements in parameter normalisation methods will aid the development of instruments and software that will be capable of meeting the growing demand in biomedical research for multi-parameter measurements of cell markers.

Chapter 6. Discussion

Haematology research in the 1990s marked a significant advancement in the identification of human HSCs. Following the isolation of a $CD34^+/CD90^+$ population with long-term *in vitro* self-renewal and differentiation capacity from foetal bone marrow, similar populations were isolated from umbilical cord blood, adult bone marrow, and peripheral blood in patients after G-CSF or GM-CSF mobilisation (Baum et al., 1992b; Bhatia et al., 1997; L. Murray et al., 1994, 1995). Succeeding studies reported the existence of a $CD34^-$ human HSC with *in vivo* repopulation potential, providing the first clues of the heterogeneity of the human HSC compartment (Bhatia et al., 1998; Zanjani et al., 1998). Work over the next decades reaffirmed the $CD34^-$ compartment as a potent source of HSCs in both umbilical cord blood and adult bone marrow, while also making strides toward elucidating the molecular mechanisms underlying this population (Anjos-Afonso et al., 2013; Danet et al., 2002). Despite these advances, the study of this unique population has been hampered by lacking strategies to isolate it at higher purity and its quiescent nature, which adds an extra layer of challenge in testing its functionality both *in vitro* and *in vivo.* While the NSG mouse model has been used in the past to study $CD34^-$ SRCs, their engraftment has been highly inefficient without administration of G-CSF and SCF (Anjos-Afonso et al., 2013). Many questions regarding the $CD34^-$ HSC have remained unanswered due to the limitations posed by our limited capacity to assay its functionality. Furthermore, the low frequency of SRCs within the $CD34^-/CD38^-/CD93^+$ poses a significant challenge in attempts to understand their molecular profile at the gene expression level both by bulk and single cell RNAseq.

In Chapter 3 of this thesis, we have reported our efforts at addressing these challenges. By screening the expression of a number of potential purifying cell surface markers, we showed that CD117 demarcates a distinct population within the $CD34^-/CD38^-/CD93^+$ compartment. By adding CD117 to our gating strategy, we showed first in the NSG mouse model that $CD34^-/CD38^-/CD93^+/CD117^+$ population reconstitutes conditioned recipients with multilineage engraftment. This was followed up by testing both $CD34^-/CD38^-/CD93^+/CD117^+$ and $CD34^-/CD38^-/CD93^+/CD117^-$ populations in enhanced NSG mouse models, NSG-S and C-Kit^{Mutant}. In the NSG-S we demonstrated enhanced engraftment compared to the NSG. In addition, we achieved faster readouts, observing engraftment in this strain as early as 6 weeks,

which had never been reported before for the CD34$^-$ SRCs. In the C-Kit^{Mutant} strain we observed a distinct engraftment kinetic beginning at 12 weeks post transplantation. The levels of reconstitution, however, steadily increased through the duration of the experiment reaching levels over 50%. While we observed human engraftment in all mice injected with CD34$^-$/CD38$^-$/CD93$^+$/CD117$^+$, this was not true for recipients transplanted with CD34$^-$/CD38$^-$/CD93$^+$/CD117$^-$, where as many as 10, 000 cells failed to engraft. With these experiments we made the first strides toward validating CD117 as positive selection marker for CD34$^-$ SRCs as well as the first insights in the benefits of the NSG-S and C-Kit^{Mutant} mouse models in revealing quiescent stem cell populations. We also further refined the gating strategy for the CD34$^-$ SRC, but dividing the CD117$^+$ compartment into CD117^{Hi} and CD117^{Int}. By injecting NSG, NSG-S, and C-Kit^{Mutant} mice with these two populations, we demonstrated that the cells driving *in vivo* reconstitution are restricted to the CD117^{Hi} fraction. Limiting dilution analysis estimated the SRC frequency in CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi} population at 1/52 cells.

We also conducted preliminary experiments that aimed to explore the potential presence of SRC populations within the CD34$^-$/CD38$^-$/CD93$^-$ compartment, where we could also detect expression CD117. We didn't observe any reconstitution in NSG-S mice transplanted with CD34$^-$/CD38$^-$/CD93$^-$/CD117$^{+/-}$, but these experiments need to be repeated with higher cell doses.

While the results reported in Chapter 3 of this thesis reaffirm the potency of the CD34$^-$ SRC, the experiments highlighted that any functional validations studies for potential HSC populations are highly dependent on the mouse models used. By testing CD34$^-$ SRCs in C-Kit^{Mutant} and NSG-S mouse models we achieved levels of reconstitution never observed before in NSG without the support of cytokine injections post-transplantation. Further to the tangible differences in engraftment kinetics and differentiation outputs in C-Kit^{Mutant} and NSG-S mice, the estimated frequencies of the CD34$^-$/CD38$^-$/CD93$^+$/CD117^{Hi} population differed significantly between strains: 1/52 in C-Kit^{Mutant} mice and 1/242 in NSG-S mice. We cannot discount the possibility that this discrepancy is due to technical issues as we observed engraftment in NSG-S mice transplanted with 50 cells, but only few mice injected with 100 cells. These experiments would need to be repeated to reaffirm if the estimated numbers of CD34$^-$ SRCs truly differ between strains. It is, however, entirely possible that some strains are better at revealing certain stem cell

populations compared with others. We have shown in this thesis that NSG-S and C-Kit^{Mutant} strains support CD34^{-} HSC engraftment far better than the classic NSG strain. The reason underlying remain speculative at this point. In the NSG-S, which requires irradiation conditioning like the NSG, we have seen increase in engraftment compared to the NSG and quicker engraftment. This is likely to be entirely driven by the human SCF, IL-3, and GM-CSF expressed in this mouse strain. In our scRNAseq data, we also explored the expression of a number of receptors involved in mediating cytokine signalling. In Table 6.1 we have summarised a list of interleukins and signalling molecules and their corresponding receptor complexes. The single cell RNAseq data suggested that a large number of receptors were not captured at all (coloured in red in Table 6.1), while some were captured at only very low levels (coloured in orange in Table 6.1). Expression of the primary subunits of the GM-CSF and IL-3 receptors were not expressed or expressed at low levels, respectively. This can suggest that the CD34^{-}/CD38^{-}/CD93^{+}/CD117^{Hi} lack the machinery to respond to IL-3 and GM-CSF stimulation. On the other hand, these cells do express CD117 (the SCF receptor) and CSF3R, which should enable them to respond to SCF and G-CSF. This is supported by the fact both of these cytokines were useful in increasing the engraftment of CD34^{-} SRCs in NSG (Anjos-Afonso et al., 2013). Of course, the fact we couldn't detect expression of other receptors doesn't necessarily mean they are not expressed. One of the caveats of scRNAseq data is that it can be informative about highly expressed genes and relative differences in gene expression between clustered populations, but cannot be used to reliably assess low-expressed genes. Quite a few receptors mediating interleukin signalling share subunits, and we found the genes for these shared subunits to be expressed. These genes are likely to be expressed at higher levels as their multi-functionality infers higher cellular demand for them. To reveal the expression patterns of the non-redundant subunits with higher confidence, other scRNAseq technologies that can achieve higher read depth can be used. Better understanding of the expression of these receptors can be useful in providing insight into signalling pathways that may aid the expansion the CD34^{-} HSC *in vitro. In vitro* work has not been the subject of this thesis, but is an important area of research with significant clinical potential that should not be ignored.

Signalling molecule	***Receptor complex***

IL-1 – TYPE 1	IL1R1, IL1RAP
IL-1 – TYPE 2	IL1R2, 1L1RAP
IL-2	IL2RA, IL2RB, IL2RG
IL-3	IL3RA, CSF2RB
IL-5	IL5RA, CSF2RB
IL-7	IL7RA, IL2RG
IL-9	IL9R
IL-15	IL15RA, IL2RB, IL2RG
IL-18	IL18R1, IL18RAP
IL-21	IL21R, IL2RG
GM-CSF	CSF2RA, CSF2RB
G-CSF	CSF3R
M-CSF	CSF1R

Table 6.1. Summary of common interleukins and chemokines, corresponding receptor complexes.

Text colours in Receptor Complex column correspond to expression status of receptor in scRNAseq data in Chapter 4. Red – not expressed; Orange – low expression; Black – expressed.

While our scRNAseq may not have been able to identify target pathways that could potentially help us study $CD34^-$ SRC *in vitro*, it did provide us with useful insights into the molecular makeup of this rare population. Surprisingly, we found a large number of cells expressed a robust HSC signature, suggesting that the HSC frequency we estimated for the population in our $C\text{-}Kit^{Mutant}$ LDA experiments may have actually grossly underestimated the number of HSCs within the $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ compartment. What this suggests is that only a fraction of the cells we transplant in mice actually engraft. Possibly, only a few $CD34^-/CD38^-/CD93^+/CD117^{Hi}$ cells have the capacity to home to the bone marrow. However, CXCR4, an important mediator haematopoietic cell bone marrow homing, was

expressed in the three clusters that seemed to contain the majority of HSC-like cells. Furthermore, it has been confirmed that $CD34^-$ SRCs do home to the murine marrow quickly after transplantation in NSG (Anjos-Afonso et al., 2013). If homing is not what impedes the engraftment of some of these cells, then it is likely that they fail to establish a niche in the murine bone marrow and expand. This is an interesting question to be addressed and can be potentially answered by clonal tracking using barcoding. These experiments conducted over serial transplants can reveal if some clones emerge only in secondary or even tertiary recipients, which may explain the lower *in vivo* HSC frequency than the scRNAseq data suggests. Barcoding of cells, however, usually involves transduction using a viral vector and successful integration of the barcodes requires a cycle of cell division. With $CD34^-$ HSCs being so difficult to culture, barcoding may be highly inefficient.

While our scRNAseq data was highly informative and established high level of similarity in gene expression between the $CD34^+$ and $CD34^-$ HSCs, it failed to address the differences between the two. To achieve this in the context of our own dataset, $CD34^+$ HSCs should have been submitted for scRNAseq together with the $CD34^-$ sample. This would have allowed us to sequence the transcripts in the two populations simultaneously using the same kits and sequencing platforms, producing data of comparable quality, which would have enabled us to generate differentially expressed gene lists. But the challenge of this experimental design would have been to decide which $CD34^+$ fraction to isolate. The HSC frequency in the $CD34^+/CD38^-$ fraction is too low and would have required high numbers of cells to be sequenced, while recent reports have cast into doubt the reliability of CD49f as a marker for the most potent $CD34^+$ HSCs (Belluschi et al., 2018; Notta et al., 2011). Another way of answering the question about the differences between $CD34^-$ and $CD34^+$ HSCs would be to use published CITEseq datasets, which provide simultaneously gene expression and cell surface maker expression data. After identifying the HSC populations based on their transcriptomic profile, CD34 cell surface marker expression can be mapped to distinguish the $CD34^-$ HSCs from the $CD34^+$ and enable differential gene expression analysis between the two. This also comes with challenges, of course. Most CITEseq datasets are done only on $CD34^+$ enriched populations, which would mean a very low representation of the $CD34^-$ HSCs, if any. Furthermore, the $CD34^-$ HSCs are extremely rare. In most of our experiments, we required several pools of cord blood in order to obtain sufficient numbers for *in vivo*

experiments and scRNAseq submission. Most scRNAseq datasets are generated from one patient or donor, decreasing the possibility that it would be able to capture a sufficient amount of $CD34^-$ HSCs for analysis.

While many questions remain to be addressed, in Chapters 3 and 4 of this thesis, we reported data that reaffirmed the existence of the $CD34^-$ HSC. Using functional and single cell transcriptomic approaches we have demonstrated that $CD34^-$ HSCs, despite challenges posed by their intrinsic nature, these cells can be studied with improved purification strategies and better mouse models. This rare and largely understudied stem cell population, similarly to findings in the $CD34^+$ compartment, may also hold heterogeneity of its own. However, for all the challenges the study of this population poses, it holds immense clinical potential.

One of the major goals of the project reported in Chapter 5 was to design a flow cytometry panel that brings together multiple HSC markers that have been reported independently over the years in mouse and human foetal haematopoiesis. By achieving this goal, we were able to conduct an analysis based on marker expression distribution, which enabled us to identify an HSC population present in both foetal liver and foetal bone marrow that was capable of *in vivo* reconstitution of $C\text{-}Kit^{Mutant}$ mice. We also identified a number of additional populations, whose self-renewal and differentiation potential remains to be tested. The data also enabled us to look into the population dynamics of different populations. The limitations posed by the cell surface markers we selected for our panel, disabled us from identifying cell populations that were more mature. Inclusion of some markers distinguishing different progenitors may have helped provide insight into the dynamics of these populations in early foetal haematopoiesis. Also, markers such as CXCR4 and some integrins implicated in the migration of foetal liver cells to the bone marrow, would have been also interesting to track over gestational stages. Still, our panel captured quite a few markers already and it is possible that the quality of the data may have diminished with further additions of antibodies.

What we also found fascinating during the analysis was the actively changing population dynamics taking place in early foetal haematopoiesis. We were observing a number of populations changing in size over time. By looking at the median expression of markers in specific clusters we could speculate about the potential identity of a population, but not the relationship between clusters. For example, one

question that was raised in the process of data analysis is if we can use the data to infer relationships between clusters at different time points in ancestor-descendant manner. Such methods have not been developed for high-parameter flow cytometry data. One reason for that is the limited nature of the data; even in a big panel, the number of measured variables may be insufficient to achieve this. But mathematical methods that could model this on the scale of flow cytometry data do exist. One such mathematical theory, Optimal Transport, has been used before and applied to scRNAseq data, establishing this form of mathematical modelling as a viable strategy to reconstructing temporally-related population relationships. This kind of analyses, however, require rates of cell division-related expansion of specific populations to be supplied, so that the algorithm can effectively distinguish between a cluster growing due to cell division or the influx of cells that belonged to different cluster in the preceding time point.

In this thesis, leveraging the power of high-parameter single cell analysis, we have presented evidence of the significant haematopoietic stem cell heterogeneity in prenatal and postnatal haematopoiesis. Through the use of *in vivo* repopulation assays we have validated populations of interest both in umbilical cord blood, foetal liver and foetal bone marrow. Much work remains to be done to establish the source of haematopoietic stem cell heterogeneity, but we have reported methodologies and findings that will hopefully inform further efforts into this area of research.

Chapter 7. Appendix

Marker	**Fluorochrome**	**Clone**	**Catalogue Number**	**Company**

CD10	FITC	HI10a	347503	BD
CD10	PE	HI10a	561707	BD
CD110	BV650	1.6.1	743579	BD
CD110	PE	1.6.1	562159	BD
CD117	BB700	104D2	742284	BD
CD117	PE	104D2	332785	BD
CD117	APC	104D2	17117842	eBioscience
CD11b	FITC	ICRF44	562793	BD
CD14	PE	M5E2	560995	BD
CD143	BV786	BB9	742900	BD
CD143	PE	BB9	557929	BD
CD144	BV711	55-7H1	743707	BD
CD144	PE	55-7H1	561714	BD
CD16	FITC	3G8	555406	BD
CD16	PE	3G8	345789	BD
CD19	FITC	SJ25C1	555412	BD
CD19	PE	SJ25C1	560961	BD
CD19	EF450	HIB19	48019942	eBioscience
CD2	FITC	RPA2.10	555326	BD
CD2	PE	RPA2.10	561760	BD
CD20	FITC	2H7	555622	BD
CD20	PE	2H7	561002	BD
CD201	BV421	RCR-252	743552	BD
CD201	PE	RCR-252	557950	BD
CD235A	FITC	HIR2	559943	BD
CD235A	PE	HIR2	555570	BD
CD24	FITC	ML5	555427	BD
CD24	PE	ML5	560991	BD
CD3	FITC	UCHT1	561806	BD
CD3	PE	UCHT1	561808	BD
CD3	APC	UCHT1	555335	BD
CD31	FITC	MEC13.3	555445	BD
CD31	PE	MEC13.3	345812	BD
CD33	PE	WM53	555450	BD
CD34	APCR700	8G12	659123	BD
CD34	PE	8G12	345802	BD
CD34	PerCP-Cy5.5	8G12	347222	BD
CD38	BUV737	HB7	564686	BD
CD38	PE	HB7	347687	BD
CD41	APC	HIP8	559777	BD
CD41	PE	HIP8	555467	BD
CD43	APCH7	IG10	655407	BD
CD43	PE	IG10	560199	BD

CD45	BUV805	Hi30	555482	BD
CD45	PE	Hi30	555483	BD
CD45	PEcy7	HI30	25045942	eBioscience
CD45RA	BUV395	HI100	740298	BD
CD45RA	PE	HI100	555489	BD
CD45RA	FITC	HI100	555488	BD
CD49F	PE	GOH3	561894	BD
CD49F	PECF594	GOH3	562474	BD
CD49f	APC-Cy7	GoH3	313628	Biolegend
CD56	FITC	B159	562794	BD
CD56	PE	B159	561650	BD
CD66B	FITC	G10F5	555724	BD
CD66B	PE	G10F5	561051	BD
CD7	BV605	M-T701	740392	BD
CD7	PE	M-T701	561934	BD
CD7	FITC	M-T701	555360	BD
CD90	PE	5E10	555596	BD
CD90	PECY7	5E10	561558	BD
CD90	BV605	5E10	747750	BD
CD93	PE	VIMD2	336107	Biologend
CD93	Biotin	VIMD2	336104	Biolegend
GPI-80	PE	3H9	D087-5	MBL International
Lineage Cocktail	FITC		340546	BD
mouse CD45	PERCPcy5.5	30-F11	45045182	eBioscience
Streptavidin	BUV496		564666	BD
Streptavidin	PE		554061	BD

Table 7.1. List of antibodies used

Reference List

Abe, T., Matsuoka, Y., Nagao, Y., Sonoda, Y., & Hanazono, Y. (2017). CD34-negative hematopoietic stem cells show distinct expression profiles of homing molecules that limit engraftment in mice and sheep. *International Journal of Hematology*, *106*(5), 631–637.

Adolfsson, J., Borge, O. J., Bryder, D., Theilgaard-Mönch, K., Åstrand-Grundström, I., Sitnicka, E., Sasaki, Y., & Jacobsen, S. E. W. (2001). Upregulation of Flt3 Expression within the Bone Marrow Lin−Sca1+c-kit+ Stem Cell Compartment Is Accompanied by Loss of Self-Renewal Capacity. *Immunity*, *15*(4), 659–669. https://doi.org/https://doi.org/10.1016/S1074-7613(01)00220-5

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L. A. M., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M., & Jacobsen, S. E. W. (2005). Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. *Cell*, *121*(2), 295–306. https://doi.org/https://doi.org/10.1016/j.cell.2005.02.013

Aihara, Y., Bühring, H.-J., Aihara, M., & Klein, J. (1986). An attempt to produce “pre-T” cell hybridomas and to identify their antigens. *European Journal of Immunology*, *16*(11), 1391–1399. https://doi.org/10.1002/eji.1830161113

Akashi, K., Traver, D., Miyamoto, T., & Weissman, I. L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, *404*(6774), 193–197. https://doi.org/10.1038/35004599

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nature Genetics*, *47*(12), 1402–1407. https://doi.org/10.1038/ng.3441

Andrews, R. G., Singer, J. W., & Bernstein, I. D. (1986). Monoclonal antibody 12-8 recognizes a 115-kd molecule present on both unipotent and multipotent hematopoietic colony-forming cells and their precursors. *Blood*, *67*(3), 842–845.

Andriole, G. L., Mulé, J. J., Hansen, C. T., Linehan, W. M., & Rosenberg, S. A. (1985). Evidence that lymphokine-activated killer cells and natural killer cells are distinct based on an analysis of congenitally immunodeficient mice. *The Journal of Immunology*, *135*(5), 2911–2913.

Anguita, E., Hughes, J., Heyworth, C., Blobel, G. A., Wood, W. G., & Higgs, D. R. (2004). Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *The EMBO Journal*, *23*(14), 2841–2852. https://doi.org/10.1038/sj.emboj.7600274

Anjos-Afonso, F., Currie, E., Palmer, H. G., Foster, K. E., Taussig, D. C., & Bonnet, D. (2013). CD34− Cells at the Apex of the Human Hematopoietic Stem Cell Hierarchy Have Distinctive Cellular and Molecular Signatures. *Cell Stem Cell*, *13*(2), 161–174. https://doi.org/10.1016/j.stem.2013.05.025

Baev, D. V, Peng, X.-H., Song, L., Barnhart, J. R., Crooks, G. M., Weinberg, K. I., & Metelitsa, L. S. (2004). Distinct homeostatic requirements of CD4+ and CD4- subsets of Valpha24-invariant natural killer T cells in humans. *Blood*, *104*(13), 4150–4156. https://doi.org/10.1182/blood-2004-04-1629

Bárcena, A., Muench, M. O., Galy, A. H. M., Cupp, J., Roncarolo, M. G., Phillips, J. H., & Spits, H. (1993). Phenotypic and Functional Analysis of T-Cell Precursors in the Human Fetal Liver and Thymus: CD7 Expression in the Early Stages of T- and Myeloid-Cell Development. *Blood*, *82*(11), 3401–3414. https://doi.org/https://doi.org/10.1182/blood.V82.11.3401.3401

Basch, R. S., & Berman, J. W. (1982). Thy-1 determinants are present on many murine hematopoietic

cells other than T cells. *European Journal of Immunology*, *12*(5), 359–364. https://doi.org/10.1002/eji.1830120502

Basu, S., Liang, H. P. H., Hernandez, I., Zogg, M., Fields, B., May, J., Ogoti, Y., Wyseure, T., Mosnier, L. O., Burns, R. T., Carlson, K., & Weiler, H. (2020). Role of thrombomodulin expression on hematopoietic stem cells. *Journal of Thrombosis and Haemostasis*, *18*(1), 123–135. https://doi.org/10.1111/jth.14663

Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A.-M., & Peault, B. (1992a). Isolation of a candidate human hematopoietic stem-cell population. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(7).

Baum, C. M., Weissman, I. L., Tsukamoto, A. S., Buckle, A.-M., & Peault, B. (1992b). Isolation of a candidate human hematopoietic stem-cell population. In *Proc. Nail. Acad. Sci. USA* (Vol. 89).

Beaudin, A. E., Boyer, S. W., Perez-Cunningham, J., Hernandez, G. E., Derderian, S. C., Jujjavarapu, C., Aaserude, E., MacKenzie, T., & Forsberg, E. C. (2016). A Transient Developmental Hematopoietic Stem Cell Gives Rise to Innate-like B and T Cells. *Cell Stem Cell*, *19*(6), 768–783. https://doi.org/10.1016/j.stem.2016.08.013

Beck, D., Thoms, J. A. I., Perera, D., Schütte, J., Unnikrishnan, A., Knezevic, K., Kinston, S. J., Wilson, N. K., O'Brien, T. A., Göttgens, B., Wong, P., & Pimanda, J. E. (2013). Genome-wide analysis of transcriptional regulators in human HSPCs reveals a densely interconnected network of coding and noncoding genes. *Blood*, *122*(14), e12–e22. https://doi.org/10.1182/blood-2013-03-490425

Becker, A. J., McCulloch, E. A., & Till, J. E. (1963). *Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells*.

Belluschi, S., Calderbank, E. F., Ciaurro, V., Pijuan-Sala, B., Santoro, A., Mende, N., Diamanti, E., Sham, K. Y. C., Wang, X., Lau, W. W. Y., Jawaid, W., Göttgens, B., & Laurenti, E. (2018). Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nature Communications*, *9*(1), 4100. https://doi.org/10.1038/s41467-018-06442-4

Bensinger, W. I., Buckner, C. D., Shannon-Dorcy, K., Rowley, S., Appelbaum, F. R., Benyunes, M., Clift, R., Martin, P., Demirer, T., & Storb, R. (1996). *Transplantation of allogeneic CD34+ peripheral blood stem cells in patients with advanced hematologic malignancy*.

Benveniste, P., Frelin, C., Janmohamed, S., Barbara, M., Herrington, R., Hyam, D., & Iscove, N. N. (2010). Intermediate-Term Hematopoietic Stem Cells with Extended but Time-Limited Reconstitution Potential. *Cell Stem Cell*, *6*(1), 48–58. https://doi.org/10.1016/j.stem.2009.11.014

Berenson, J., Andrews, R. G., Bensinger, W. I., Kalamasz, D., Knitter, G., Buckner, C. D., & Bernstein, I. D. (1988). Antigen CD34+ marrow cells engraft lethally irradiated baboons. *The Journal of Clinical Investigation*, *81*(3), 951–955. https://doi.org/10.1172/JCI113409

Berenson, J., Bensinger, W., Hill, R., Andrews, R., Garcia-Lopez, J., Kalamasz, D., Still, B., Spitzer, G., Buckner, C., & Bernstein, I. (1991). Engraftment after infusion of CD34+ Marrow Cells in patients with breast cancer or neuroblastoma. *Blood*, *77*, 1717–1722. https://doi.org/10.1182/blood.V77.8.1717.bloodjournal7781717

Bertrand, J. Y., Bastien Giroux, S., Golub, R., Le Klaine, M., Jalil, A., Boucontet, L., Godin, I., & Cumano, A. (2004). *Characterization of purified intraembryonic hematopoietic stem cells as a tool to define their site of origin*. www.pnas.orgcgidoi10.1073pnas.0402270102

Berzins, S. P., Cochrane, A. D., Pellicci, D. G., Smyth, M. J., & Godfrey, D. I. (2005). Limited correlation

between human thymus and blood NKT cell content revealed by an ontogeny study of paired tissue samples. *European Journal of Immunology*, *35*(5), 1399–1407. https://doi.org/10.1002/eji.200425958

Beverley, P. C. L., Linch, D., & Delia, D. (1980). Isolation of human haematopoietic progenitor cells using monoclonal antibodies. *Nature*, *287*(5780), 332–333. https://doi.org/10.1038/287332a0

Bhatia, M., Bonnet, D., Murdoch, B., Gan, O. I., & Dick, J. E. (1998). A newly discovered class of human hematopoietic cells with SCID-repopulating activity. *Nature Medicine*, *4*(9), 1038–1045. https://doi.org/10.1038/2023

Bhatia, M., Bonnet, D., Wu, D., Murdoch, B., Wrana, J., Gallacher, L., & Dick, J. E. (1999). Bone morphogenetic proteins regulate the developmental program of human hematopoietic stem cells. *The Journal of Experimental Medicine*, *189*(7), 1139–1148. http://www.ncbi.nlm.nih.gov/pubmed/10190905

Bhatia, M., Wang, J. C. Y., Kapp, U., Bonnet, D., & Dick, J. E. (1997). Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proceedings of the National Academy of Sciences*, *94*(10), 5320–5325.

Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M. P., Giannelli, S., Neduva, V., Dow, D. J., Schmidt, M., Von Kalle, C., Roncarolo, M. G., Ciceri, F., Vicard, P., Wit, E., … Aiuti, A. (2016). In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. *Cell Stem Cell*, *19*(1), 107–119. https://doi.org/10.1016/j.stem.2016.04.016

Billerbeck, E., Barry, W. T., Mu, K., Dorner, M., Rice, C. M., & Ploss, A. (2011). Development of human CD4⁺FoxP3⁺ regulatory T cells in human stem cell factor–, granulocyte-macrophage colony-stimulating factor–, and interleukin-3–expressing NOD-SCID IL2Rγ^{null} humanize. *Blood*, *117*(11), 3076 LP – 3086. http://www.bloodjournal.org/content/117/11/3076.abstract

Boos, M. D., Yokota, Y., Eberl, G., & Kee, B. L. (2007). Mature natural killer cell and lymphoid tissue–inducing cell development requires Id2-mediated suppression of E protein activity. *The Journal of Experimental Medicine*, *204*(5), 1119–1130.

Bowman, T. V, McCooey, A. J., Merchant, A. A., Ramos, C. A., Fonseca, P., Poindexter, A., Bradfute, S. B., Oliveira, D. M., Green, R., Zheng, Y., Jackson, K. A., Chambers, S. M., McKinney-Freeman, S. L., Norwood, K. G., Darlington, G., Gunaratne, P. H., Steffen, D., & Goodell, M. A. (2006). Differential mRNA Processing in Hematopoietic Stem Cells. *STEM CELLS*, *24*(3), 662–670. https://doi.org/https://doi.org/10.1634/stemcells.2005-0552

Bradley, T. R., & Metcalf, D. (1966). The growth of mouse bone marrow cells in vitro. *Australian Journal of Experimental Biology and Medical Science*, *44*(3), 287–300.

Brehm, M. A., Cuthbert, A., Yang, C., Miller, D. M., DiIorio, P., Laning, J., Burzenski, L., Gott, B., Foreman, O., Kavirayani, A., Herlihy, M., Rossini, A. A., Shultz, L. D., & Greiner, D. L. (2010). Parameters for establishing humanized mouse models to study human immunity: Analysis of human hematopoietic stem cell engraftment in three immunodeficient strains of mice bearing the IL2rγnull mutation. *Clinical Immunology*, *135*(1), 84–98. https://doi.org/10.1016/j.clim.2009.12.008

Brown, J., Greaves, M. F., & Molgaard, H. V. (1991). The gene encoding the stem cell antigen, CD34, is conserved in mouse and expressed in haemopoietic progenitor cell lines, brain, and embryonic fibroblasts. *International Immunology*, *3*(2), 175–184. https://doi.org/10.1093/intimm/3.2.175

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Höfer, T.,

& Rodewald, H.-R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, *518*(7540), 542–546. https://doi.org/10.1038/nature14242

Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D. B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., Wü nsche, P., Zeisberger, P., Brocks, D., Gu, L., Herrmann, C., Haas, S., Essers, M. A., Krijgsveld, J., & Trumpp, A. (2014). Cell Stem Cell Resource Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis. *Stem Cell*, *15*(4), 507–522. https://doi.org/10.1016/j.stem.2014.07.005

Calvanese, V., Nguyen, A. T., Bolan, T. J., Vavilina, A., Su, T., Lee, L. K., Wang, Y., Lay, F. D., Magnusson, M., Crooks, G. M., Kurdistani, S. K., & Mikkola, H. K. A. (2019). MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. *Nature*, *576*(7786), 281–286. https://doi.org/10.1038/s41586-019-1790-2

Calvanese, V., Prashad, S. L., Magnusson, M., & Mikkola, H. K. A. (2014). Analysis of Highly Self-Renewing GPI-80+ Human Fetal Hematopoietic Stem Cells Identifies Novel Regulators of Stemness. *Blood*, *124*(21). http://www.bloodjournal.org/content/124/21/4314?sso-checked=true

Cao, X., Wu, X., Frassica, D., Yu, B., Pang, L., Xian, L., Wan, M., Lei, W., Armour, M., Tryggestad, E., Wong, J., Wen, C. Y., Lu, W. W., & Frassica, F. J. (2011). Irradiation induces bone injury by damaging bone marrow microenvironment for stem cells. *Proceedings of the National Academy of Sciences*, *108*(4), 1609 LP – 1614. https://doi.org/10.1073/pnas.1015350108

Carrelha, J., Meng, Y., Kettyle, L. M., Luis, T. C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., Högstrand, K., Lord, A. M., Sanjuan-Pla, A., Woll, P. S., Nerlov, C., & Jacobsen, S. E. W. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*, *554*(7690), 106–111. https://doi.org/10.1038/nature25455

Christensen, J. L., & Weissman, I. L. (2001). Flk-2 is a marker in hematopoietic stem cell differentiation: A simple method to isolate long-term stem cells. *Proceedings of the National Academy of Sciences*, *98*(25), 14541 LP – 14546. https://doi.org/10.1073/pnas.261562798

Christensen, J. L., Wright, D. E., Wagers, A. J., & Weissman, I. L. (2004). Circulation and Chemotaxis of Fetal Hematopoietic Stem Cells. *PLoS Biology*, *2*(3), e75. https://doi.org/10.1371/journal.pbio.0020075

Civin, C. I., Strauss, L. C., Brovall, C., Fackler, M. J., Schwartz, J. F., & Shaper, J. H. (1984). Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *Journal of Immunology*, *133*(1).

Cochrane, S. W., Zhao, Y., Welner, R. S., & Sun, X.-H. (2009). Balance between Id and E proteins regulates myeloid-versus-lymphoid lineage decisions. *Blood, The Journal of the American Society of Hematology*, *113*(5), 1016–1026.

Coffman, R. L., & Weissman, I. L. (1981). A monoclonal antibody that recognizes B cells and B cell precursors in mice. *Journal of Experimental Medicine*, *153*(2), 269–279. https://doi.org/10.1084/jem.153.2.269

Corbel, C., & Salaün, J. (2002). αIIb integrin expression during development of the murine hemopoietic system. *Developmental Biology*, *243*(2), 301–311. https://doi.org/10.1006/dbio.2001.0553

Cortegano, I., Serrano, N., Ruiz, C., Rodríguez, M., Prado, C., Alía, M., Hidalgo, A., Cano, E., de Andrés, B., & Gaspar, M.-L. (2019). CD45 expression discriminates waves of embryonic megakaryocytes in the mouse. *Haematologica*, *104*(9), 1853–1865. https://doi.org/10.3324/haematol.2018.192559

Cortés, F., Debacker, C., Péault, B., & Labastie, M. C. (1999). Differential expression of KDR/VEGFR-2 and CD34 during mesoderm development of the early human embryo. *Mechanisms of Development*, *83*(1–2), 161–164. https://doi.org/10.1016/S0925-4773(99)00030-1

Cosgun, K. N., Rahmig, S., Mende, N., Reinke, S., Hauber, I., Schäfer, C., Petzold, A., Weisbach, H., Heidkamp, G., Purbojo, A., Cesnjevar, R., Platz, A., Bornhäuser, M., Schmitz, M., Dudziak, D., Hauber, J., Kirberg, J., & Waskow, C. (2014). Kit Regulates HSC Engraftment across the Human-Mouse Species Barrier. *Cell Stem Cell*, *15*(2), 227–238. https://doi.org/10.1016/J.STEM.2014.06.001

Covassin, L., Jangalwe, S., Jouvet, N., Laning, J., Burzenski, L., Shultz, L. D., & Brehm, M. A. (2013). Human immune system development and survival of non-obese diabetic (NOD)-scid IL2rγnull (NSG) mice engrafted with human thymus and autologous haematopoietic stem cells. *Clinical and Experimental Immunology*, *174*(3), 372–388. https://doi.org/10.1111/cei.12180

Craig, W., Kay, R., Cutler, R. L., & Lansdorp, P. M. (1993). Expression of Thy-1 on human hematopoietic progenitor cells. *The Journal of Experimental Medicine*, *177*(5), 1331–1342. https://doi.org/10.1084/jem.177.5.1331

Dagher, R. N., Hiatt, K., Traycoff, C., Srour, E. F., & Yoder, M. C. (1998). c-Kit and CD38 are expressed by long-term reconstituting hematopoietic cells present in the murine yolk sac. *Biology of Blood and Marrow Transplantation : Journal of the American Society for Blood and Marrow Transplantation*, *4*(2), 69–74. https://doi.org/10.1053/bbmt.1998.v4.pm9763109

Danet, G. H., Luongo, J. L., Butler, G., Lu, M. M., Tenner, A. J., Simon, M. C., & Bonnet, D. A. (2002). C1qRp defines a new human stem cell population with hematopoietic and hepatic potential. *Proceedings of the National Academy of Sciences*, *99*(16), 10441–10445. https://doi.org/10.1073/pnas.162104799

Dao, M. A., Arevalo, J., & Nolta, J. A. (2003). Reversibility of CD34 expression on human hematopoietic stem cells that retain the capacity for secondary reconstitution. *Blood*, *101*(1), 112–118. https://doi.org/10.1182/blood-2002-01-0025

Davi, F., Faili, A., Gritti, C., Blanc, C., Laurent, C., Sutton, L., Schmitt, C., & Merle-Béral, H. (1997). Early onset of immunoglobulin heavy chain gene rearrangements in normal human bone marrow CD34+ cells. *Blood*, *90*(10), 4014–4021. http://www.ncbi.nlm.nih.gov/pubmed/9354670

de Bruijn, M. F. T. R., Speck, N. A., Peeters, M. C. E., & Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *The EMBO Journal*, *19*(11), 2465–2474. https://doi.org/10.1093/emboj/19.11.2465

Dick, J., Kamel-Reid, S., Murdoch, B., & Doedens, M. (1991). Gene transfer into normal human hematopoietic cells using in vitro and in vivo assays. *Blood*, *78*(3), 624–634. https://doi.org/10.1182/blood.V78.3.624.624

Doulatov, S., Notta, F., Eppert, K., Nguyen, L. T., Ohashi, P. S., & Dick, J. E. (2010). Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nature Immunology*, *11*(7), 585–593. https://doi.org/10.1038/ni.1889

Doulatov, S., Notta, F., Laurenti, E., & Dick, J. E. (2012). Hematopoiesis: a human perspective. *Cell Stem Cell*, *10*(2), 120–136. https://doi.org/10.1016/j.stem.2012.01.006

Dykstra, B., Kent, D., Bowie, M., McCaffrey, L., Hamilton, M., Lyons, K., Lee, S. J., Brinkman, R., & Eaves, C. (2007). Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo. *Cell Stem Cell*, *1*(2), 218–229. https://doi.org/10.1016/j.stem.2007.05.015

Edvardsson, L., Dykes, J., & Olofsson, T. (2006). Isolation and characterization of human myeloid

progenitor populations--TpoR as discriminator between common myeloid and megakaryocyte/erythroid progenitors. *Experimental Hematology*, *34*(5), 599–609. https://doi.org/10.1016/j.exphem.2006.01.017

Ema, H., & Nakauchi, H. (2000). Expansion of hematopoietic stem cells in the developing liver of a mouse embryo. *Blood*, *95*(7), 2284–2288. https://doi.org/10.1182/blood.V95.7.2284

Emambokus, N. R., & Frampton, J. (2003). The glycoprotein IIb molecule is expressed on early murine hematopoietic progenitors and regulates their numbers in sites of hematopoiesis. *Immunity*, *19*(1), 33–45.

Eppert, K., Takenaka, K., Lechman, E. R., Waldron, L., Nilsson, B., van Galen, P., Metzeler, K. H., Poeppl, A., Ling, V., Beyene, J., Canty, A. J., Danska, J. S., Bohlander, S. K., Buske, C., Minden, M. D., Golub, T. R., Jurisica, I., Ebert, B. L., & Dick, J. E. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. *Nature Medicine*, *17*(9), 1086–1093. https://doi.org/10.1038/nm.2415

Ezine, S., Weissman, I. L., & Rouse, R. V. (1984). Bone marrow cells give rise to distinct cell clones within the thymus. *Nature*, *309*(5969), 629–631. https://doi.org/10.1038/309629a0

Ferkowicz, M. J., Starr, M., Xie, X., Li, W., Johnson, S. A., Shelley, W. C., Morrison, P. R., & Yoder, M. C. (2003). CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development (Cambridge, England)*, *130*(18), 4393–4403. https://doi.org/10.1242/dev.00632

Ficara, F., Crisafulli, L., Lin, C., Iwasaki, M., Smith, K. S., Zammataro, L., & Cleary, M. L. (2013). Pbx1 restrains myeloid maturation while preserving lymphoid potential in hematopoietic progenitors. *Journal of Cell Science*, *126*(Pt 14), 3181–3191. https://doi.org/10.1242/jcs.125435

Forsberg, E. C., Serwold, T., Kogan, S., Weissman, I. L., & Passegué, E. (2006). New Evidence Supporting Megakaryocyte-Erythrocyte Potential of Flk2/Flt3+ Multipotent Hematopoietic Progenitors. *Cell*, *126*(2), 415–426. https://doi.org/https://doi.org/10.1016/j.cell.2006.06.037

Fukuda, T. (1973). Fetal hemopoiesis. I. Electron microscopic studies on human yolk sac hemopoiesis. *Virchows Archiv. B, Cell Pathology*, *14*(3), 197–213.

Galy, A., Travis, M., Cen, D., & Chen, B. (1995). Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. *Immunity*, *3*(4), 459–473. https://doi.org/https://doi.org/10.1016/1074-7613(95)90175-2

Georgantas, R. W., Tanadve, V., Malehorn, M., Heimfeld, S., Chen, C., Carr, L., Martinez-Murillo, F., Riggins, G., Kowalski, J., & Civin, C. I. (2004). Microarray and Serial Analysis of Gene Expression Analyses Identify Known and Novel Transcripts Overexpressed in Hematopoietic Stem Cells. *Cancer Research*, *64*(13), 4434 LP – 4441. https://doi.org/10.1158/0008-5472.CAN-03-3247

Godin, I., Dieterlen-Lièvre, F., & Cumano, A. (1995). Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(3), 773–777. https://doi.org/10.1073/pnas.92.3.773

Godin, I., Garcia-Porrero, J. A., Coutinho, A., Dieterlen-Lièvre, F., & Marcos, M. A. R. (1993). Para-aortic splanchnopleura from early mouse embryos contains B1a cell progenitors. *Nature*, *364*(6432), 67–70. https://doi.org/10.1038/364067a0

Goodell, M. A., Brose, K., Paradis, G., Conner, A. S., & Mulligan, R. C. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *Journal of Experimental*

Medicine, *183*(4), 1797–1806. https://doi.org/10.1084/jem.183.4.1797

Goodell, M. A., Rosenzweig, M., Kim, H., Marks, D. F., DeMaria, M., Paradis, G., Grupp, S. A., Sieff, C. A., Mulligan, R. C., & Johnson, R. P. (1997a). Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. *Nature Medicine*, *3*(12), 1337–1345. http://www.ncbi.nlm.nih.gov/pubmed/9396603

Goodell, M. A., Rosenzweig, M., Kim, H., Marks, D. F., DeMaria, M., Paradis, G., Grupp, S. A., Sieff, C. A., Mulligan, R. C., & Johnson, R. P. (1997b). Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species. *Nature Medicine*, *3*(12), 1337–1345. https://doi.org/10.1038/nm1297-1337

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., & van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, *525*(7568), 251–255. https://doi.org/10.1038/nature14966

Grün, D., Muraro, M. J., Boisset, J. C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., van Es, J., Jansen, E., Clevers, H., de Koning, E. J. P., & van Oudenaarden, A. (2016). De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell*, *19*(2), 266–277. https://doi.org/10.1016/j.stem.2016.05.010

Guenechea, G., Gan, O. I., Dorrell, C., & Dick, J. E. (2001). Distinct classes of human stem cells that differ in proliferative and self-renewal potential. *Nature Immunology*, *2*(1), 75–82. https://doi.org/10.1038/83199

Gunji, Y., Nakamura, M., Osawa, H., Nagayoshi, K., Nakauchi, H., Miura, Y., Yanagisawa, M., & Suda, T. (1993). Human primitive hematopoietic progenitor cells are more enriched in KITlow cells than in KIThigh cells. *Blood*, *82*(11), 3283–3289.

Guo, G., Luc, S., Marco, E., Lin, T. W., Peng, C., Kerenyi, M. A., Beyaz, S., Kim, W., Xu, J., Das, P. P., Neff, T., Zou, K., Yuan, G. C., & Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, *13*(4), 492–505. https://doi.org/10.1016/j.stem.2013.07.017

Haddad, R., Guimiot, F., Six, E., Jourquin, F., Setterblad, N., Kahn, E., Yagello, M., Schiffer, C., Andre-Schmutz, I., Cavazzana-Calvo, M., Gluckman, J. C., Delezoide, A.-L., Pflumio, F., & Canque, B. (2006). Dynamics of Thymus-Colonizing Cells during Human Development. *Immunity*, *24*(2), 217–230. https://doi.org/https://doi.org/10.1016/j.immuni.2006.01.008

Haghverdi, L., Buettner, F., & Theis, F. J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, *31*(18), 2989–2998. https://doi.org/10.1093/bioinformatics/btv325

Hamlett, I., Draper, J., Strouboulis, J., Iborra, F., Porcher, C., & Vyas, P. (2008). Characterization of megakaryocyte GATA1-interacting proteins: The corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. *Blood*, *112*(7), 2738–2749. https://doi.org/10.1182/blood-2008-03-146605

Hao, Q. L., Shah, A. J., Thiemann, F. T., Smogorzewska, E. M., & Crooks, G. M. (1995). A functional comparison of CD34 + CD38- cells in cord blood and bone marrow. *Blood*, *86*(10), 3745–3753. https://doi.org/10.1182/blood.V86.10.3745.bloodjournal86103745

Hao, Q. L., Zhu, J., Price, M. A., Payne, K. J., Barsky, L. W., & Crooks, G. M. (2001). Identification of a novel, human multilymphoid progenitor in cord blood. *Blood*, *97*(12), 3683–3690. https://doi.org/10.1182/blood.V97.12.3683

Harly, C., Kenney, D., Wang, Y., Ding, Y., Zhao, Y., Awasthi, P., & Bhandoola, A. (2020). A Shared Regulatory Element Controls the Initiation of Tcf7 Expression During Early T Cell and Innate Lymphoid Cell Developments. *Frontiers in Immunology*, *11*, 470. https://doi.org/10.3389/fimmu.2020.00470

Harrison, D. E., Zhong, R. K., Jordan, C. T., Lemischka, I. R., & Astle, C. M. (1997). Relative to adult marrow, fetal liver repopulates nearly five times more effectively long-term than short-term. *Experimental Hematology*, *25*(4), 293–297.

Hassan, H. T., Zeller, W., Stockschläder, M., Krüger, W., Hoffknecht, M. M., & Zander, A. R. (1996). Comparison between Bone Marrow and G-CSF-Mobilized Peripheral Blood Allografts Undergoing Clinical Scale CD34+ Cell Selection. *STEM CELLS*, *14*(4), 419–429. https://doi.org/10.1002/stem.140419

Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L., & Salomonis, N. (2018). The Human Cell Atlas bone marrow single-cell interactive web portal. *Experimental Hematology*, *68*, 51–61. https://doi.org/10.1016/j.exphem.2018.09.004

Haynes, B. F., Martin, M. E., Kay, H. H., & Kurtzberg, J. (1988). Early events in human T cell ontogeny. Phenotypic characterization and immunohistologic localization of T cell precursors in early human fetal tissues. *The Journal of Experimental Medicine*, *168*(3), 1061–1080.

Higuchi, Y., Zeng, H., & Ogawa, M. (2003). CD38 expression by hematopoietic stem cells of newborn and juvenile mice. *Leukemia*, *17*(1), 171–174. https://doi.org/10.1038/sj.leu.2402785

Hill, B., Rozler, E., Travis, M., Chen, S., Zannetino, A., Simmons, P., Galy, A., Chen, B., & Hoffman, R. (1996). High-level expression of a novel epitope of CD59 identifies a subset of CD34+ bone marrow cells highly enriched for pluripotent stem cells. *Experimental Hematology*, *24*(8), 936–943.

Hoffmann, A., & Spengler, D. (2019). Chromatin Remodeling Complex NuRD in Neurodevelopment and Neurodevelopmental Disorders. *Frontiers in Genetics*, *10*, 682. https://doi.org/10.3389/fgene.2019.00682

Hokland, P., Hokland, M., Daley, J., & Ritzi, J. (1987). Identification and cloning of a prethymic precursor T lymphocyte from a population of common acute lymphoblastic leukemia antigen (CALLA)-positive fetal bone marrow cells. *Journal of Experimental Medicine*, *165*(6), 1749–1754. https://doi.org/10.1084/jem.165.6.1749

Holmberg Olausson, K., Maire, C. L., Haidar, S., Ling, J., Learner, E., Nistér, M., & Ligon, K. L. (2014). Prominin-1 (CD133) Defines Both Stem and Non-Stem Cell Populations in CNS Development and Gliomas. *PLoS ONE*, *9*(9), e106694. https://doi.org/10.1371/journal.pone.0106694

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441. https://doi.org/10.1037/h0071325

Hu, Y., & Smyth, G. K. (2009). ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of Immunological Methods*, *347*(1–2), 70–78. https://doi.org/10.1016/J.JIM.2009.06.008

Huang, Z., Dore, L. C., Li, Z., Orkin, S. H., Feng, G., Lin, S., & Crispino, J. D. (2009). GATA-2 Reinforces Megakaryocyte Development in the Absence of GATA-1. *Molecular and Cellular Biology*, *29*(18), 5168–5180. https://doi.org/10.1128/mcb.00482-09

Hughes, P. F. D., Eaves, C. J., Hogge, D. E., & Humphries, R. K. (1989). High-Efficiency Gene Transfer to Human Hematopoietic Cells Maintained in Long-Term Marrow Culture. In *Blood* (Vol. 74, Issue 6). www.bloodjournal.org

Ikuta, K., & Weissman, I. L. (1992). Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. *Proceedings of the National Academy of Sciences*, *89*(4), 1502 LP – 1506. https://doi.org/10.1073/pnas.89.4.1502

Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., & Teichmann, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biology*, *17*, 29. https://doi.org/10.1186/s13059-016-0888-1

Ingersoll, M. A., Spanbroek, R., Lottaz, C., Gautier, E. L., Frankenberger, M., Hoffmann, R., Lang, R., Haniffa, M., Collin, M., Tacke, F., Habenicht, A. J. R., Ziegler-Heitbrock, L., & Randolph, G. J. (2010). Comparison of gene expression profiles between human and mouse monocyte subsets. *Blood*, *115*(3), e10-9. https://doi.org/10.1182/blood-2009-07-235028

Iscove, N. (1990). Searching for stem cells. In *Nature* (Vol. 347, Issue 6289, pp. 126–127). Nature. https://doi.org/10.1038/347126a0

Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A., & Lemischka, I. R. (2002). A Stem Cell Molecular Signature. *Science*, *298*(5593), 601 LP – 604. https://doi.org/10.1126/science.1073823

Ivanovs, A., Rybtsov, S., Anderson, R. A., & Medvinsky, A. (2014). CD43 but Not CD41 Marks the First Hematopoietic Stem Cells in the Human Embryo. *Blood*, *124*(21), 4330. https://doi.org/10.1182/blood.V124.21.4330.4330

Ivanovs, A., Rybtsov, S., Anderson, R. A., Turner, M. L., & Medvinsky, A. (2014). Identification of the niche and phenotype of the first human hematopoietic stem cells. *Stem Cell Reports*, *2*(4), 449–456. https://doi.org/10.1016/j.stemcr.2014.02.004

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R. A., Turner, M. L., & Medvinsky, A. (2011). Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *The Journal of Experimental Medicine*, *208*(12), 2417–2427. https://doi.org/10.1084/jem.20111688

Iwasaki, H., Arai, F., Kubota, Y., Dahl, M., & Suda, T. (2010). Endothelial protein C receptor-expressing hematopoietic stem cells reside in the perisinusoidal niche in fetal liver. *Blood*, *116*(4), 544–553. https://doi.org/10.1182/blood-2009-08-240903

Iwasaki, M., Liedtke, M., Gentles, A. J., & Cleary, M. L. (2015). CD93 Marks a Non-Quiescent Human Leukemia Stem Cell Population and Is Required for Development of MLL-Rearranged Acute Myeloid Leukemia. *Cell Stem Cell*, *17*(4), 412–421. https://doi.org/10.1016/j.stem.2015.08.008

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V, Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz, G., … Ebert, B. L. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *The New England Journal of Medicine*, *371*(26), 2488–2498. https://doi.org/10.1056/NEJMoa1408617

Jangalwe, S., Shultz, L. D., Mathew, A., & Brehm, M. A. (2016). Improved B cell development in humanized NOD-scid IL2Rγ(null) mice transgenically expressing human stem cell factor, granulocyte-macrophage colony-stimulating factor and interleukin-3. *Immunity, Inflammation and Disease*, *4*(4), 427–440. https://doi.org/10.1002/iid3.124

Jankovic, V., Ciarrocchi, A., Boccuni, P., DeBlasio, T., Benezra, R., & Nimer, S. D. (2007). Id1 restrains myeloid commitment, maintaining the self-renewal capacity of hematopoietic stem cells. *Proceedings of the National Academy of Sciences*, *104*(4), 1260–1265.

Jokubaitis, V. J., Sinka, L., Driessen, R., Whitty, G., Haylock, D. N., Bertoncello, I., Smith, I., Péault, B., Tavian, M., & Simmons, P. J. (2008). Angiotensin-converting enzyme (CD143) marks hematopoietic stem cells in human embryonic, fetal, and adult hematopoietic tissues. *Blood*, *111*(8).

Julien, E., El Omar, R., & Tavian, M. (2016). Origin of the hematopoietic system in the human embryo. *FEBS Letters*, *590*(22), 3987–4001. https://doi.org/https://doi.org/10.1002/1873-3468.12389

Kamel-Reid, S., & Dick, J. E. (1988). Engraftment of immune-deficient mice with human hematopoietic stem cells. *Science*, *242*(4886), 1706–1709. https://doi.org/10.1126/science.2904703

Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., Usukhbayar, B., Calvo, J., Taylor, S., Goardon, N., Six, E., Pflumio, F., Porcher, C., Majeti, R., Göttgens, B., & Vyas, P. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells article. *Nature Immunology*, *19*(1), 85–97. https://doi.org/10.1038/s41590-017-0001-2

Kawabata, K., Ujikawa, M., Egawa, T., Kawamoto, H., Tachibana, K., Iizasa, H., Katsura, Y., Kishimoto, T., & Nagasawa, T. (1999). A cell-autonomous requirement for CXCR4 in long-term lymphoid and myeloid reconstitution. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(10), 5663–5667. https://doi.org/10.1073/pnas.96.10.5663

Khan, J. A., Mendelson, A., Kunisaki, Y., Birbrair, A., Kou, Y., Arnal-Estapé, A., Pinho, S., Ciero, P., Nakahara, F., Ma'ayan, A., Bergman, A., Merad, M., & Frenette, P. S. (2016). Fetal liver hematopoietic stem cell niches associate with portal vessels. *Science*, *351*(6269), 176–180. https://doi.org/10.1126/science.aad0084

Kim, I., Yilmaz, O. H., & Morrison, S. J. (2005). CD144 (VE-cadherin) is transiently expressed by fetal liver hematopoietic stem cells. *Blood*, *106*(3), 903–905. https://doi.org/10.1182/blood-2004-12-4960

Kobayashi, M., Laver, J. H., Kato, T., Miyazaki, H., & Ogawa, M. (1996). Thrombopoietin supports proliferation of human primitive hematopoietic cells in synergy with steel factor and/or interleukin-3. *Blood*, *88*(2), 429–436. https://doi.org/10.1182/blood.v88.2.429.bloodjournal882429

Köberle, M., Kaesler, S., Kempf, W., Wölbing, F., & Biedermann, T. (2012). Tetraspanins in Mast Cells. *Frontiers in Immunology*, *3*, 106. https://doi.org/10.3389/fimmu.2012.00106

Kollet, O., Peled, A., Byk, T., Ben-Hur, H., Greiner, D., Shultz, L., & Lapidot, T. (2000). Beta2 microglobulin-deficient (B2m(null)) NOD/SCID mice are excellent recipients for studying human stem cell function. *Blood*, *95*, 3102–3105. https://doi.org/10.1182/blood.V95.10.3102.010k14_3102_3105

Komeno, Y., Yan, M., Matsuura, S., Lam, K., Lo, M.-C., Huang, Y.-J., Tenen, D. G., Downing, J. R., & Zhang, D.-E. (2014). Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood*, *123*(24), 3760–3769. https://doi.org/10.1182/blood-2013-08-521252

Kondo, M., Wagers, A. J., Manz, M. G., Prohaska, S. S., Scherer, D. C., Beilhack, G. F., Shizuru, J. A., & Weissman, I. L. (2003). BIOLOGY OF HEMATOPOIETIC STEM CELLS AND PROGENITORS : Implications for Clinical Application . *Annual Review of Immunology*, *21*(1), 759–806. https://doi.org/10.1146/annurev.immunol.21.120601.141007

Kondo, M., Weissman, I. L., & Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, *91*(5), 661–672. https://doi.org/10.1016/S0092-8674(00)80453-5

Krause, D. S., Fackler, M., Civin, C., & May, W. (1996). CD34: structure, biology, and clinical utility [see comments]. *Blood*, *87*(1), 1–13. https://doi.org/10.1182/blood.V87.1.1.1

Krause, D. S., Ito, T., Fackler, M. J., Smith, O. M., Collector, M. I., Sharkis, S. J., & May, W. S. (1994). Characterization of murine CD34, a marker for hematopoietic progenitor and stem cells. *Blood*,

84(3), 691–701.

Krauter, J., Hartl, M., Hambach, L., Kohlenberg, A., Gunsilius, E., Ganser, A., & Heil, G. (2001). Receptor-mediated endocytosis of CD34 on hematopoietic cells after stimulation with the monoclonal antibody anti-HPCA-1. *Journal of Hematotherapy & Stem Cell Research*, *10*(6), 863–871. https://doi.org/10.1089/152581601317210953

Kunisaki, Y., Bruns, I., Scheiermann, C., Ahmed, J., Pinho, S., Zhang, D., Mizoguchi, T., Wei, Q., Lucas, D., Ito, K., Mar, J. C., Bergman, A., & Frenette, P. S. (2013). Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature*, *502*(7473), 637–643. https://doi.org/10.1038/nature12612

Labastie, M.-C., Cortés, F., Roméo, P.-H., Dulac, C., & Péault, B. (1998). Molecular Identity of Hematopoietic Precursor Cells Emerging in the Human Embryo. *Blood*, *92*(10), 3624–3635. https://doi.org/10.1182/blood.V92.10.3624

Lai, A. Y., & Kondo, M. (2006). Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *Journal of Experimental Medicine*, *203*(8), 1867–1873. https://doi.org/10.1084/jem.20060697

Lansdorp, P. M., Sutherland, H. J., & Eaves, C. J. (1990). Selective expression of CD45 isoforms on functional subpopulations of CD34+ hemopoietic cells from human bone marrow. *Journal of Experimental Medicine*, *172*(1).

Larochelle, Andre, Savona, M., Wiggins, M., Anderson, S., Ichwan, B., Keyvanfar, K., Morrison, S. J., & Dunbar, C. E. (2011). Human and rhesus macaque hematopoietic stem cells cannot be purified based only on SLAM family markers. *Blood*, *117*(5), 1550–1554. https://doi.org/10.1182/blood-2009-03-212803

Larochelle, André, Vormoor, J., Hanenberg, H., Wang, J. C. Y., Bhatia, M., Lapidot, T., Moritz, T., Murdoch, B., Xiao, X. L., Kato, I., WIlliams, D. A., & Dick, J. E. (1996). Identification of primitive human hematopoietic cells capable of repopulating NOD/SCID mouse bone marrow: Implications for gene therapy. *Nature Medicine*, *2*(12), 1329–1337. https://doi.org/10.1038/nm1296-1329

Laurenti, E., Doulatov, S., Zandi, S., Plumb, I., Chen, J., April, C., Fan, J.-B., & Dick, J. E. (2013). The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nature Immunology*, *14*(7), 756–763. https://doi.org/10.1038/ni.2615

Lawrence, H. J., Christensen, J., Fong, S., Hu, Y.-L., Weissman, I., Sauvageau, G., Humphries, R. K., & Largman, C. (2005). Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. *Blood*, *106*(12), 3988–3994. https://doi.org/10.1182/blood-2005-05-2003

Lee-Six, H., Obro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J. P., Martincorena, I., Anderson, E., O'Neill, L., Stratton, M. R., Laurenti, E., Green, A. R., Kent, D. G., & Campbell, P. J. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*, *561*(7724), 473–478. https://doi.org/10.1038/s41586-018-0497-0

Lee, J., Zhou, Y. J., Ma, W., Zhang, W., Aljoufi, A., Luh, T., Lucero, K., Liang, D., Thomsen, M., Bhagat, G., Shen, Y., & Liu, K. (2017). Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors. *Nature Immunology*, *18*(8), 877–888. https://doi.org/10.1038/ni.3789

Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir, E. D., Tadmor, M. D., Litvin, O., Fienberg, H. G., Jager, A., Zunder, E. R., Finck, R., Gedman, A. L., Radtke, I., Downing, J. R., Pe'er, D., &

Nolan, G. P. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*, *162*(1), 184–197. https://doi.org/10.1016/j.cell.2015.05.047

Li, D., Xue, W., Li, M., Dong, M., Wang, J., Wang, X., Li, X., Chen, K., Zhang, W., Wu, S., Zhang, Y., Gao, L., Chen, Y., Chen, J., Zhou, B. O., Zhou, Y., Yao, X., Li, L., Wu, D., & Pan, W. (2018). VCAM-1+ macrophages guide the homing of HSPCs to a vascular niche. *Nature*, *564*(7734), 119–124. https://doi.org/10.1038/s41586-018-0709-7

Li, Y., Liu, B., Lukin, K., Finkelman, F., Hagman, J., Roers, A., & Huang, H. (2016). GATA2 is critical for mast cell differentiation and maintenance. *The Journal of Immunology*, *196*(1 Supplement), 202.4 LP-202.4. http://www.jimmunol.org/content/196/1_Supplement/202.4.abstract

Link, H., Arseniev, L., Bahre, O., Kadar, J. G., Diedrich, H., & Poliwoda, H. (1996). Transplantation of allogeneic CD34+ blood cells. *Blood*, *87*(11), 4903–4909. https://doi.org/10.1182/blood.V87.11.4903.bloodjournal87114903

Love, P. E., Warzecha, C., & Li, L. (2014). Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends in Genetics : TIG*, *30*(1), 1–9. https://doi.org/10.1016/j.tig.2013.10.001

Luetke-Eversloh, M., Killig, M., & Romagnani, C. (2013). Signatures of human NK cell development and terminal differentiation. *Frontiers in Immunology*, *4*, 499. https://doi.org/10.3389/fimmu.2013.00499

Maillard, L., Sanfilippo, S., Domenech, C., Kasmi, N., Petit, L., Jacques, S., Delezoide, A.-L., Guimiot, F., Eladak, S., Moison, D., Nicolas, N., Rouiller-Fabre, V., Pozzi-Godin, S., Mennesson, B., Brival, M.-L., Letourneur, F., Jaffredo, T., Chomienne, C., & Souyri, M. (2020). CD117(hi) expression identifies a human fetal hematopoietic stem cell population with high proliferation and self-renewal potential. In *Haematologica* (Vol. 105, Issue 2, pp. e43–e47). https://doi.org/10.3324/haematol.2018.207811

Majeti, R., Park, C. Y., & Weissman, I. L. (2007). Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. *Cell Stem Cell*, *1*(6). https://doi.org/10.1016/j.stem.2007.10.001

Majeti, Ravindra, Park, C. Y., & Weissman, I. L. (2007). Identification of a Hierarchy of Multipotent Hematopoietic Progenitors in Human Cord Blood. *Cell Stem Cell*, *1*(6), 635–645. https://doi.org/http://dx.doi.org/10.1016/j.stem.2007.10.001

Manz, M. G., Miyamoto, T., Akashi, K., & Weissman, I. L. (2002). Prospective isolation of human clonogenic common myeloid progenitors. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(18), 11872–11877. https://doi.org/10.1073/pnas.172384399

Matsuoka, Y., Sumide, K., Kawamura, H., Nakatsuka, R., Fujioka, T., & Sonoda, Y. (2016). GPI-80 expression highly purifies human cord blood–derived primitive CD34-negative hematopoietic stem cells. *Blood*, *128*(18). http://www.bloodjournal.org/content/128/18/2258?sso-checked=true

Matsuoka, Y., Takahashi, M., Sumide, K., Kawamura, H., Nakatsuka, R., Fujioka, T., & Sonoda, Y. (2017). CD34 Antigen and the MPL Receptor Expression Defines a Novel Class of Human Cord Blood-Derived Primitive Hematopoietic Stem Cells. *Cell Transplantation*, *26*(6), 1043–1058. https://doi.org/10.3727/096368916X694201

Mayani, H., Dragowska, W., & Lansdorp, P. (1993). Characterization of functionally distinct subpopulations of CD34+ cord blood cells in serum-free long-term cultures supplemented with hematopoietic cytokines. *Blood*, *82*, 2664–2672. https://doi.org/10.1182/blood.V82.9.2664.bloodjournal8292664

Mazurier, F., Gan, O. I., McKenzie, J. L., Doedens, M., & Dick, J. E. (2004). Lentivector-mediated clonal tracking reveals intrinsic heterogeneity in the human hematopoietic stem cell compartment and culture-induced stem cell impairment. *Blood*, *103*(2), 545–552. https://doi.org/10.1182/blood-2003-05-1558

McCune, J. M., Namikawa, R., Kaneshima, H., Shultz, L. D., Lieberman, M., & Weissman, I. L. (1988). The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function. *Science*, *241*(4873), 1632 LP – 1639. http://science.sciencemag.org/content/241/4873/1632.abstract

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. http://arxiv.org/abs/1802.03426

McIntosh, B. E., Brown, M. E., Duffin, B. M., Maufort, J. P., Vereide, D. T., Slukvin, I. I., & Thomson, J. A. (2018). Nonirradiated NOD,B6.SCID Il2rγ−/− KitW41/W41 (NBSGW) Mice Support Multilineage Engraftment of Human Hematopoietic Cells. *Stem Cell Reports*, *4*(2), 171–180. https://doi.org/10.1016/j.stemcr.2014.12.005

McKearn, J. P., McCubrey, J., & Fagg, B. (1985). Enrichment of hematopoietic precursor cells and cloning of multipotential B-lymphocyte precursors. *Proceedings of the National Academy of Sciences*, *82*(21), 7414 LP – 7418. https://doi.org/10.1073/pnas.82.21.7414

McKenzie, J. L., Gan, O. I., Doedens, M., & Dick, J. E. (2004). Development of a Novel NOD/SCID Transplant System That Provides Enhanced Detection of Rapid-SRC and Insight into Their Self-Renewal and Mobilization. *Blood*, *104*(11), 249. https://doi.org/10.1182/blood.V104.11.249.249

McKenzie, J. L., Gan, O. I., Doedens, M., Wang, J. C. Y., & Dick, J. E. (2006). Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment. *Nature Immunology*, *7*(11), 1225–1233. https://doi.org/10.1038/ni1393

Medvinsky, A., & Dzierzak, E. (1996). Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, *86*(6), 897–906. https://doi.org/10.1016/S0092-8674(00)80165-8

Medvinsky, A., Samoylina, N. L., Müller, A. M., & Dzierzak, E. A. (1993). An early pre-liver intraembryonic source of CFU-S in the developing mouse. *Nature*, *364*(6432), 64–67. https://doi.org/10.1038/364064a0

Mikkola, H. K. A., Fujiwara, Y., Schlaeger, T. M., Traver, D., & Orkin, S. H. (2003). Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. *Blood*, *101*(2), 508–516. https://doi.org/10.1182/blood-2002-06-1699

Mitjavila-Garcia, M. T., Cailleret, M., Godin, I., Nogueira, M. M., Cohen-Solal, K., Schiavon, V., Lecluse, Y., Le Pesteur, F., Lagrue, A. H., & Vainchenker, W. (2002). Expression of CD41 on hematopoietic progenitors derived from embryonic hematopoietic cells. *Development (Cambridge, England)*, *129*(8), 2003–2013.

Miyamoto, T., & Akashi, K. (2005). Lineage Promiscuous Expression of Transcription Factors in Normal Hematopoiesis. *International Journal of Hematology*, *81*(5), 361–367. https://doi.org/10.1532/IJH97.05003

Miyawaki, K., Iwasaki, H., Jiromaru, T., Kusumoto, H., Yurino, A., Sugio, T., Uehara, Y., Odawara, J., Daitoku, S., Kunisaki, Y., Mori, Y., Arinobu, Y., Tsuzuki, H., Kikushige, Y., Iino, T., Kato, K., Takenaka, K., Miyamoto, T., Maeda, T., & Akashi, K. (2017). Identification of unipotent megakaryocyte progenitors in human hematopoiesis. *Blood*, *129*(25), 3332–3343. https://doi.org/10.1182/blood-2016-09-741611

Moignard, V., MacAulay, I. C., Swiers, G., Buettner, F., Schütte, J., Calero-Nieto, F. J., Kinston, S., Joshi, A., Hannah, R., Theis, F. J., Jacobsen, S. E., De Bruijn, M. F., & Göttgens, B. (2013). Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nature Cell Biology*, *15*(4), 363–372.

https://doi.org/10.1038/ncb2709

Montecino-Rodriguez, E., Leathers, H., & Dorshkind, K. (2006). Identification of a B-1 B cell-specified progenitor. *Nature Immunology*, *7*(3), 293–301. https://doi.org/10.1038/ni1301

Morel, F., Galy, A., Chen, B., & Szilvassy, S. J. (1998). Equal distribution of competitive long-term repopulating stem cells in the CD34+ and CD34- fractions of Thy-1lowLin-/lowSca-1+ bone marrow cells. *Experimental Hematology*, *26*(5), 440–448.

Moritz, T., Patel, V. P., & Williams, D. A. (1994). Bone marrow extracellular matrix molecules improve gene transfer into human hematopoietic cells via retroviral vectors. *Journal of Clinical Investigation*, *93*(4), 1451–1457. https://doi.org/10.1172/JCI117122

Morrison, S. J., Hemmati, H. D., Wandycz, A. M., & Weissman, I. L. (1995). The purification and characterization of fetal liver hematopoietic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(22), 10302–10306. https://doi.org/10.1073/pnas.92.22.10302

Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E., & Weissman, I. L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. *Development*, *124*(10).

Morrison, S. J., & Weissman, I. L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*, *1*(8), 661–673. https://doi.org/10.1016/1074-7613(94)90037-X

Mouthon, M. (1993). Expression of tal-1 and GATA-binding proteins during human hematopoiesis. *Blood*, *81*, 647–655.

Mozazarello, P., Calligaro, A. L., & Calligaro, A. (2001). Giulio Bizzozero: a pioneer of cell biology. *Molecular Cell Biology*, *2*, 776–780.

Muller-Sieburg, C. E., Whitlock, C. A., & Weissman, I. L. (1986). Isolation of two early B lymphocyte progenitors from mouse marrow: A committed Pre-Pre-B cell and a clonogenic Thy-1lo hematopoietic stem cell. *Cell*, *44*(4), 653–662. https://doi.org/10.1016/0092-8674(86)90274-6

Murray, L., Chen, B., Galy, A., Chen, S., Tushinski, R., Uchida, N., Negrin, R., Tricot, G., Jagannath, S., & Vesole, D. (1995). Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood. *Blood*, *85*(2), 368–378.

Murray, L., DiGiusto, D., Chen, B., Chen, S., Combs, J., Conti, A., Galy, A., Negrin, R., Tricot, G., & Tsukamoto, A. (1994). Analysis of human hematopoietic stem cell populations. *Blood Cells*, *20*(2–3), 364–370.

Murray, P. (1932). The development in vitro of the blood of the early chick embryo. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *111*(773), 497–521. https://doi.org/10.1098/rspb.1932.0070

Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H., & Kishimoto, T. (1996). Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature*, *382*(6592), 635–638.

Nicolini, F. E., Cashman, J. D., Hogge, D. E., Humphries, R. K., & Eaves, C. J. (2004). 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*, *18*(2), 341.

Nie, Y., Han, Y.-C., & Zou, Y.-R. (2008). CXCR4 is required for the quiescence of primitive hematopoietic cells. *Journal of Experimental Medicine*, *205*(4), 777–783.

North, T. E., De Bruijn, M. F. T. R., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E., &

Speck, N. A. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity*, *16*(5), 661–672. https://doi.org/10.1016/S1074-7613(02)00296-0

Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., & Dick, J. E. (2011). Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. *Science*, *333*(6039). http://science.sciencemag.org/content/333/6039/218.full

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y., & Dick, J. E. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, *351*(6269). http://science.sciencemag.org/content/351/6269/aab2116.abstract

Novelli, E. M., Ramirez, M., & Civin, C. I. (1998). Biology of CD34+CD38- cells in lymphohematopoiesis. *Leukemia & Lymphoma*, *31*(3–4), 285–293. https://doi.org/10.3109/10428199809059221

Oberlin, E., Fleury, M., Clay, D., Petit-Cocault, L., Candelier, J.-J., Mennesson, B., Jaffredo, T., & Souyri, M. (2010). VE-cadherin expression allows identification of a new class of hematopoietic stem cells within human embryonic liver. *Blood*, *116*(22), 4444–4455. https://doi.org/10.1182/blood-2010-03-272625

Okada, S., Nakauchi, H., Nagayoshi, K., Nishikawa, S., Miura, Y., & Suda, T. (1992). In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells. *Blood*, *80*(12), 3044–3050.

Okuno, Y., Iwasaki, H., Huettner, C. S., Radomska, H. S., Gonzalez, D. A., Tenen, D. G., & Akashi, K. (2002). Differential regulation of the human and murine CD34 genes in hematopoietic stem cells. *Proceedings of the National Academy of Sciences*, *99*(9), 6246 LP – 6251. https://doi.org/10.1073/pnas.092027799

Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H., & Grimes, H. L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature*, *537*(7622), 698–702. https://doi.org/10.1038/nature19348

Orchansky, P., Rubinstein, M., & Fischer, D. G. (1986). The interferon-gamma receptor in human monocytes is different from the one in nonhematopoietic cells. *Journal of Immunology (Baltimore, Md. : 1950)*, *136*(1), 169–173.

Osawa, M., Hanada, K. I., Hamada, H., & Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34- low/negative hematopoietic stem cell. *Science*, *273*(5272), 242–245. https://doi.org/10.1126/science.273.5272.242

Otis, E. M., & Brent, R. (1954). Equivalent ages in mouse and human embryos - Embryology. *The Anatomical Record*, *120*(1), 33–63. https://embryology.med.unsw.edu.au/embryology/index.php/Paper_-_Equivalent_ages_in_mouse_and_human_embryos

Pang, W. W., Price, E. A., Sahoo, D., Beerman, I., Maloney, W. J., Rossi, D. J., Schrier, S. L., & Weissman, I. L. (2011). Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proceedings of the National Academy of Sciences*, *108*(50), 20012–20017.

Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C., & Weissman, I. L. (2005). Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *Journal of Experimental Medicine*, *202*(11), 1599–1611. https://doi.org/10.1084/jem.20050967

Pellin, D., Loperfido, M., Baricordi, C., Wolock, S. L., Montepeloso, A., Weinberg, O. K., Biffi, A., Klein, A. M., & Biasco, L. (2019). A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications*, *10*(1), 2395. https://doi.org/10.1038/s41467-019-10291-0

Petit Cocault, L., Fleury, M., Clay, D., Larghero, J., Vanneaux, V., & Souyri, M. (2016). Monoclonal antibody 1.6.1 against human MPL receptor allows HSC enrichment of CB and BM CD34+CD38− populations. *Experimental Hematology*, *44*(4), 297-302.e1. https://doi.org/10.1016/J.EXPHEM.2015.12.004

Petrenko, O., Beavis, A., Klaine, M., Kittappa, R., Godin, I., & Lemischka, I. R. (1999). The Molecular Characterization of the Fetal Stem Cell Marker AA4. *Immunity*, *10*(6), 691–700. https://doi.org/10.1016/S1074-7613(00)80068-0

Pietras, E. M., Reynaud, D., Kang, Y. A., Carlin, D., Calero-Nieto, F. J., Leavitt, A. D., Stuart, J. A., Göttgens, B., & Passegué, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell*, *17*(1), 35–46. https://doi.org/10.1016/j.stem.2015.05.003

Pluznik, D. H., & Sachs, L. (1965). The cloning of normal “mast” cells in tissue culture. *Journal of Cellular and Comparative Physiology*, *66*(3), 319–324.

Pop, R., Shearstone, J. R., Shen, Q., Liu, Y., Hallstrom, K., Koulnis, M., Gribnau, J., & Socolovsky, M. (2010). A key commitment step in erythropoiesis is synchronized with the cell cycle clock through mutual inhibition between PU.1 and S-phase progression. *PLoS Biology*, *8*(9). https://doi.org/10.1371/journal.pbio.1000484

Popescu, D. M., Botting, R. A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E. F., Polanski, K., Goh, I., Efremova, M., Acres, M., Maunder, D., Vegh, P., Gitton, Y., Park, J. E., Vento-Tormo, R., Miao, Z., Dixon, D., Rowell, R., … Haniffa, M. (2019). Decoding human fetal liver haematopoiesis. *Nature*, *574*(7778), 365–371. https://doi.org/10.1038/s41586-019-1652-y

Prashad, S. L., Calvanese, V., Yao, C. Y., Kaiser, J., Wang, Y., Sasidharan, R., Crooks, G., Magnusson, M., & Mikkola, H. K. A. (2015). GPI-80 Defines Self-Renewal Ability in Hematopoietic Stem Cells during Human Development. In *Cell Stem Cell* (Vol. 16, Issue 1). https://doi.org/10.1016/j.stem.2014.10.020

Prussin, C., & Metcalfe, D. D. (2006). 5. IgE, mast cells, basophils, and eosinophils. *Journal of Allergy and Clinical Immunology*, *117*(2), S450–S456. https://doi.org/10.1016/j.jaci.2005.11.016

Psaila, B., Barkas, N., Iskander, D., Roy, A., Anderson, S., Ashley, N., Caputo, V. S., Lichtenberg, J., Loaiza, S., Bodine, D. M., Karadimitris, A., Mead, A. J., & Roberts, I. (2016). Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biology*, *17*(1), 83. https://doi.org/10.1186/s13059-016-0939-7

Qian, H., Georges-Labouesse, E., Nyström, A., Domogatskaya, A., Tryggvason, K., Jacobsen, S. E. W., & Ekblom, M. (2007). Distinct roles of integrins alpha6 and alpha4 in homing of fetal liver hematopoietic stem and progenitor cells. *Blood*, *110*(7), 2399–2407. https://doi.org/10.1182/blood-2006-10-051276

Randall, T. D., Lund, F. E., Howard, M. C., & Weissman, I. L. (1996). Expression of murine CD38 defines a population of long-term reconstituting hematopoietic stem cells. *Blood*, *87*(10), 4057–4067.

Rebel, V. I., Miller, C. L., Eaves, C. J., & Lansdorp, P. M. (1996). The repopulation potential of fetal liver hematopoietic stem cells in mice exceeds that of their adult bone marrow counterparts. *Blood*, *87*(8),

3500–3507. https://doi.org/10.1182/blood.v87.8.3500.bloodjournal8783500

Reckzeh, K., Kizilkaya, H., Helbo, A. S., Alrich, M. E., Deslauriers, A. G., Grover, A., Rapin, N., Asmar, F., Grønbæk, K., Porse, B., Borregaard, N., Vestweber, D., Nerlov, C., & Theilgaard-Mönch, K. (2018). Human adult HSCs can be discriminated from lineage-committed HPCs by the expression of endomucin. *Blood Advances*, *2*(13), 1628–1632. https://doi.org/10.1182/bloodadvances.2018015743

Reif, A. E., & Allen, J. M. (1966). The Antigenic Stability of 3 AKR Leukemias on Isotransplantation and the Serologic Detection of Thymus-derived Leukemia Cells. *Cancer Research*, *26*(1), 123–130.

Robin, C., Ottersbach, K., Boisset, J.-C., Oziemlak, A., & Dzierzak, E. (2011). CD41 is developmentally regulated and differentially expressed on mouse hematopoietic stem cells. *Blood*, *117*(19), 5088–5091. https://doi.org/10.1182/blood-2011-01-329516

Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., & Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature*, *553*(7687), 212–216. https://doi.org/10.1038/nature25168

Ropert, C. (2019). How toll-like receptors reveal monocyte plasticity: the cutting edge of antiinflammatory therapy. *Cellular and Molecular Life Sciences*, *76*(4), 745–755. https://doi.org/10.1007/s00018-018-2959-9

Rossi, D. J., Bryder, D., Zahn, J. M., Ahlenius, H., Sonu, R., Wagers, A. J., & Weissman, I. L. (2005). Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proceedings of the National Academy of Sciences*, *102*(26), 9194–9199.

Rossi, M., Yokota, T., Medina, K. L., Garrett, K. P., Comp, P. C., Schipul, A. H., & Kincade, P. W. (2003). B lymphopoiesis is active throughout human life, but there are developmental age-related changes. *Blood*, *101*(2), 576–584. https://doi.org/10.1182/blood-2002-03-0896

Rybtsov, S., Batsivari, A., Bilotkach, K., Paruzina, D., Senserrich, J., Nerushev, O., & Medvinsky, A. (2014). Tracing the Origin of the HSC Hierarchy Reveals an SCF-Dependent, IL-3-Independent CD43− Embryonic Precursor. *Stem Cell Reports*, *3*(3), 489–501. https://doi.org/10.1016/j.stemcr.2014.07.009

Rylski, M., Welch, J. J., Chen, Y.-Y., Letting, D. L., Diehl, J. A., Chodosh, L. A., Blobel, G. A., & Weiss, M. J. (2003). GATA-1-mediated proliferation arrest during erythroid maturation. *Molecular and Cellular Biology*, *23*(14), 5031–5042. https://doi.org/10.1128/mcb.23.14.5031-5042.2003

Sanjuan-Pla, A., Macaulay, I. C., Jensen, C. T., Woll, P. S., Luis, T. C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Jones, T. B., Chowdhury, O., Stenson, L., Lutteropp, M., Green, J. C. A., Facchini, R., Boukarabila, H., Grover, A., Gambardella, A., Thongjuea, S., … Jacobsen, S. E. W. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, *502*(7470), 232–236. https://doi.org/10.1038/nature12495

Sato, T., Laver, J. H., & Ogawa, M. (1999). Reversible expression of CD34 by murine hematopoietic stem cells. *Blood*, *94*(8), 2548–2554.

Sauvageau, G., Lansdorp, P. M., Eaves, C. J., Hogge, D. E., Dragowska, W. H., Reid, D. S., Largman, C., Lawrence, H. J., & Humphries, R. K. (1994). Differential expression of homeobox genes in functionally distinct CD34+ subpopulations of human bone marrow cells. *Proceedings of the National Academy of Sciences*, *91*(25), 12223 LP – 12227. https://doi.org/10.1073/pnas.91.25.12223

Severe, N., Karabacak, N. M., Gustafsson, K., Baryawno, N., Courties, G., Kfoury, Y., Kokkaliaris, K. D., Rhee, C., Lee, D., Scadden, E. W., Garcia-Robledo, J. E., Brouse, T., Nahrendorf, M., Toner, M., & Scadden, D. T. (2019). Stress-Induced Changes in Bone Marrow Stromal Cell Populations Revealed

through Single-Cell Protein Expression Mapping. *Cell Stem Cell*, *25*(4), 570-583.e7. https://doi.org/https://doi.org/10.1016/j.stem.2019.06.003

Shen, F. W., Tung, J. S., & Boyse, E. A. (1986). Further definition of the Ly-5 system. *Immunogenetics*, *24*(3), 146–149. https://doi.org/10.1007/BF00364741

Shen, W. F., Chang, C. P., Rozenfeld, S., Sauvageau, G., Humphries, R. K., Lu, M., Lawrence, H. J., Cleary, M. L., & Largman, C. (1996). Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. *Nucleic Acids Research*, *24*(5), 898–906. https://doi.org/10.1093/nar/24.5.898

Shen, W. F., Montgomery, J. C., Rozenfeld, S., Moskow, J. J., Lawrence, H. J., Buchberg, A. M., & Largman, C. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Molecular and Cellular Biology*, *17*(11), 6448–6458. https://doi.org/10.1128/mcb.17.11.6448

Shin, J. Y., Hu, W., Naramura, M., & Park, C. Y. (2014). High c-Kit expression identifies hematopoietic stem cells with impaired self-renewal and megakaryocytic bias. *The Journal of Experimental Medicine*, *211*(2), 217 LP – 231. http://jem.rupress.org/content/211/2/217.abstract

Shultz, L. D., Schweitzer, P. A., Christianson, S. W., Gott, B., Schweitzer, I. B., Tennent, B., McKenna, S., Mobraaten, L., Rajan, T. V, & Greiner, D. L. (1995). Multiple defects in innate and adaptive immunologic function in NOD/LtSz-scid mice. *Journal of Immunology (Baltimore, Md. : 1950)*, *154*(1), 180–191. http://www.ncbi.nlm.nih.gov/pubmed/7995938

Simsek, T., Kocabas, F., Zheng, J., Deberardinis, R. J., Mahmoud, A. I., Olson, E. N., Schneider, J. W., Zhang, C. C., & Sadek, H. A. (2010). The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell*, *7*(3), 380–390. https://doi.org/10.1016/j.stem.2010.07.011

Sinka, L., Biasch, K., Khazaal, I., Péault, B., & Tavian, M. (2012). Angiotensin-converting enzyme (CD143) specifies emerging lympho-hematopoietic progenitors in the human embryo. *Blood*, *119*(16).

Sintes, J., Romero, X., Marin, P., Terhorst, C., & Engel, P. (2008). Differential expression of CD150 (SLAM) family receptors by human hematopoietic stem and progenitor cells. *Experimental Hematology*, *36*(9), 1199–1204. https://doi.org/10.1016/j.exphem.2008.03.015

Six, E. M., Bonhomme, D., Monteiro, M., Beldjord, K., Jurkowska, M., Cordier-Garcia, C., Garrigue, A., Dal Cortivo, L., Rocha, B., Fischer, A., Cavazzana-Calvo, M., & André-Schmutz, I. (2007). A human postnatal lymphoid progenitor capable of circulating and seeding the thymus. *Journal of Experimental Medicine*, *204*(13), 3085–3093. https://doi.org/10.1084/jem.20071003

Snippert, H. J., van Es, J. H., van den Born, M., Begthel, H., Stange, D. E., Barker, N., & Clevers, H. (2009). Prominin-1/CD133 Marks Stem Cells and Early Progenitors in Mouse Small Intestine. *Gastroenterology*, *136*(7). https://doi.org/10.1053/j.gastro.2009.03.002

Solar, G. P., Kerr, W. G., Zeigler, F. C., Hess, D., Donahue, C., De Sauvage, F. J., & Eaton, D. L. (1998). Role of c-mpl in early hematopoiesis. In *Blood* (Vol. 92, Issue 1, pp. 4–10). https://doi.org/10.1182/blood.v92.1.4.413k38_4_10

Sommerkamp, P., Renders, S., Ladel, L., Hotz-Wagenblatt, A., Schönberger, K., Zeisberger, P., Przybylla, A., Sohn, M., Zhou, Y., Klibanski, A., Cabezas-Wallscheid, N., & Trumpp, A. (2019). The long non-coding RNA Meg3 is dispensable for hematopoietic stem cells. *Scientific Reports*, *9*(1), 2110. https://doi.org/10.1038/s41598-019-38605-8

Spangrude, G J, Brooks, D. M., & Tumas, D. B. (1995). 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*, *85*(4), 1006–1016. https://doi.org/10.1182/blood.V85.4.1006.bloodjournal8541006

Spangrude, Gerald J., Heimfeld, S., & Weissman, I. L. (1988). Purification and characterization of mouse hematopoietic stem cells. *Science*, *241*(4861), 58–62. https://doi.org/10.1126/science.2898810

Spits, H., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J. P., Eberl, G., Koyasu, S., Locksley, R. M., McKenzie, A. N. J., Mebius, R. E., Powrie, F., & Vivier, E. (2013). Innate lymphoid cells — a proposal for uniform nomenclature. *Nature Reviews Immunology*, *13*(2), 145–149. https://doi.org/10.1038/nri3365

Subramaniam, A., Talkhoncheh, M. S., Magnusson, M., & Larsson, J. (2019). Endothelial protein C receptor (EPCR) expression marks human fetal liver hematopoietic stem cells. *Haematologica*, *104*(2), e47–e50. https://doi.org/10.3324/haematol.2018.198515

Sudo, K., Ema, H., Morita, Y., & Nakauchi, H. (2000). Age-associated characteristics of murine hematopoietic stem cells. *The Journal of Experimental Medicine*, *192*(9), 1273–1280.

Sugiyama, T., Kohara, H., Noda, M., & Nagasawa, T. (2006). Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity*, *25*(6), 977–988.

Sugrue, T., Lowndes, N. F., & Ceredig, R. (2014). Hypoxia Enhances the Radioresistance of Mouse Mesenchymal Stromal Cells. *STEM CELLS*, *32*(8), 2188–2200. https://doi.org/10.1002/stem.1683

Sumide, K., Matsuoka, Y., Kawamura, H., Nakatsuka, R., Fujioka, T., Asano, H., Takihara, Y., & Sonoda, Y. (2018). A revised road map for the commitment of human cord blood CD34-negative hematopoietic stem cells. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04441-z

Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J. J., Klein, A., Hofmann, O., & Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. *Nature*, *514*(7522). https://doi.org/10.1038/nature13824

Sutherland, H., Eaves, C., Eaves, A., Dragowska, W., & Lansdorp, P. (1989). Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. *Blood*, *74*(5), 1563–1570. https://doi.org/10.1182/blood.V74.5.1563.1563

Tajima, F., Sato, T., Laver, J. H., & Ogawa, M. (2000). CD34 expression by murine hematopoietic stem cells mobilized by granulocyte colony-stimulating factor. *Blood*, *96*(5), 1989–1993. https://doi.org/10.1182/blood.V96.5.1989

Takahashi, K., Yamamura, F., & Naito, M. (1989). Differentiation, maturation, and proliferation of macrophages in the mouse yolk sac: A light-microscopic, enzyme-cytochemical, immunohistochemical, and ultrastructural study. *Journal of Leukocyte Biology*, *45*(2), 87–96. https://doi.org/10.1002/jlb.45.2.87

Takahashi, M., Matsuoka, Y., Sumide, K., Nakatsuka, R., Fujioka, T., Kohno, H., Sasaki, Y., Matsui, K., Asano, H., & Kaneko, K. (2014). CD133 is a positive marker for a distinct class of primitive human cord blood-derived CD34-negative hematopoietic stem cells. *Leukemia*, *28*(6), 1308.

Takashina, T. (1987). Haemopoiesis in the human yolk sac. *Journal of Anatomy*, *151*, 125–135.

Takubo, K., Nagamatsu, G., Kobayashi, C. I., Nakamura-Ishizu, A., Kobayashi, H., Ikeda, E., Goda, N., Rahimi, Y., Johnson, R. S., Soga, T., Hirao, A., Suematsu, M., & Suda, T. (2013). Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. *Cell Stem Cell*, *12*(1), 49–61. https://doi.org/10.1016/j.stem.2012.10.011

Tavian, M., Coulombel, L., Luton, D., Clemente, H. S., Dieterlen-Lievre, F., & Peault, B. (1996). Aorta-associated CD34+ hematopoietic cells in the early human embryo. *Blood*, *87*(1), 67–72. https://doi.org/10.1182/blood.V87.1.67.67

Tavian, M., Hallais, M. F., & Peault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development*, *126*(4).

Tavian, M., Hallais, M. F., & Péault, B. (1999). Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development*, *126*(4), 793–803.

Tavian, M., Robin, C., Coulombel, L., & Péault, B. (2001). The human embryo, but not its yolk sac, generates lympho-myeloid stem cells: Mapping multipotent hematopoietic cell fate in intraembryonic mesoderm. *Immunity*, *15*(3), 487–495. https://doi.org/10.1016/S1074-7613(01)00193-5

Thomas, E. D., Lochte Jr., H. L., Cannon, J. H., Sahler, O. D., & Ferrebee, J. W. (1959). SUPRALETHAL WHOLE BODY IRRADIATION AND ISOLOGOUS MARROW TRANSPLANTATION IN MAN. *The Journal of Clinical Investigation*, *38*(10), 1709–1716. https://doi.org/10.1172/JCI103949

Thomas, E. D., Lochte Jr, H. L., Lu, W. C., & Ferrebee, J. W. (1957). Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. *New England Journal of Medicine*, *257*(11), 491–496.

Till, J. E., & McCulloch, E. A. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation Research*, *14*.

Till, J. E., McCulloch, E. A., & Siminovitch, L. (1964). A STOCHASTIC MODEL OF STEM CELL PROLIFERATION, BASED ON THE GROWTH OF SPLEEN COLONY-FORMING CELLS. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(1), 29–36. https://doi.org/10.1073/pnas.51.1.29

Tindle, R. W., Nichols, R. A. B., Chan, L., Campana, D., Catovsky, D., & Birnie, G. D. (1985). A novel monoclonal antibody BI-3C5 recognises myeloblasts and non-B non-T lymphoblasts in acute leukaemias and CGL blast crises, and reacts with immature cells in normal bone marrow. *Leukemia Research*, *9*(1), 1–9. https://doi.org/https://doi.org/10.1016/0145-2126(85)90016-5

Tipping, A. J., Pina, C., Castor, A., Hong, D., Rodrigues, N. P., Lazzari, L., May, G. E., Jacobsen, S. E. W., & Enver, T. (2009). High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle. *Blood*, *113*(12), 2661–2672. https://doi.org/10.1182/blood-2008-06-161117

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., … Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, *352*(6282), 189 LP – 196. https://doi.org/10.1126/science.aad0501

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*(4), 381–386. https://doi.org/10.1038/nbt.2859

Tschopp, J., Jenne, D. E., Hertig, S., Preissner, K. T., Morgenstern, H., Sapino, A. P., & French, L. (1993). Human megakaryocytes express clusterin and package it without apolipoprotein A-1 into alpha-granules. *Blood*, *82*(1), 118–125.

Umemoto, T., Hashimoto, M., Matsumura, T., Nakamura-Ishizu, A., & Suda, T. (2018). Ca(2+)-

mitochondria axis drives cell division in hematopoietic stem cells. *The Journal of Experimental Medicine*, *215*(8), 2097–2113. https://doi.org/10.1084/jem.20180421

van Galen, P., Kreso, A., Wienholds, E., Laurenti, E., Eppert, K., Lechman, E. R., Mbong, N., Hermans, K., Dobson, S., April, C., Fan, J.-B., & Dick, J. E. (2014). Reduced Lymphoid Lineage Priming Promotes Human Hematopoietic Stem Cell Expansion. *Cell Stem Cell*, *14*(1), 94–106. https://doi.org/https://doi.org/10.1016/j.stem.2013.11.021

Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. In *Journal of Machine Learning Research* (Vol. 9).

Vannini, N., Girotra, M., Naveiras, O., Nikitin, G., Campos, V., Giger, S., Roch, A., Auwerx, J., & Lutolf, M. P. (2016). Specification of haematopoietic stem cell fate via modulation of mitochondrial activity. *Nature Communications*, *7*, 13125. https://doi.org/10.1038/ncomms13125

Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., Boch, T., Hofmann, W. K., Ho, A. D., Huber, W., Trumpp, A., Essers, M. A. G., & Steinmetz, L. M. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. *Nature Cell Biology*, *19*(4), 271–281. https://doi.org/10.1038/ncb3493

Visserc, J. W. M., Bauman, J. G. J., Mulder, A. H., Euason, J. F., & De Leeuw, A. M. (1984). Isolation of murine pluripotent hemopoietic stem cells. *Journal of Experimental Medicine*, *159*(6), 1576–1590. https://doi.org/10.1084/jem.159.6.1576

Vitali, C., Bassani, C., Chiodoni, C., Fellini, E., Guarnotta, C., Miotti, S., Sangaletti, S., Fuligni, F., De Cecco, L., Piccaluga, P. P., Colombo, M. P., & Tripodo, C. (2015). SOCS2 Controls Proliferation and Stemness of Hematopoietic Cells under Stress Conditions and Its Deregulation Marks Unfavorable Acute Leukemias. *Cancer Research*, *75*(11), 2387–2399. https://doi.org/10.1158/0008-5472.CAN-14-3625

Vodyanik, M. A., Thomson, J. A., & Slukvin, I. I. (n.d.). *STEM CELLS IN HEMATOLOGY Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures*. https://doi.org/10.1182/blood-2006-02-003327

Vodyanik, M. A., Thomson, J. A., & Slukvin, I. I. (2006). Leukosialin (CD43) defines hematopoietic progenitors in human embryonic stem cell differentiation cultures. *Blood*, *108*(6), 2095–2105. https://doi.org/10.1182/blood-2006-02-003327

Vogelstein, B., Kinzler, K., Weissman, I. L., & Quake, S. R. (2006). Digital PCR. *PNAS*, *96*(16), 9236–9241. https://doi.org/10.1073/pnas.96.16.9236

Wang, H., He, J., Xu, C., Chen, X., Yang, H., Shi, S., Liu, C., Zeng, Y., Wu, D., Bai, Z., Wang, M., Wen, Y., Su, P., Xia, M., Huang, B., Ma, C., Bian, L., Lan, Y., Cheng, T., … Zhou, J. (2020). Decoding Human Megakaryocyte Development. *Cell Stem Cell*. https://doi.org/https://doi.org/10.1016/j.stem.2020.11.006

Warren, L., Bryder, D., Weissman, I. L., & Quake, S. R. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(47), 17807–17812. https://doi.org/10.1073/pnas.0608512103

Weiss, M. J., & Orkin, S. H. (1995). GATA transcription factors: key regulators of hematopoiesis. *Experimental Hematology*, *23*(2), 99–107.

Welch, J. J., Watts, J. A., Vakoc, C. R., Yao, Y., Wang, H., Hardison, R. C., Blobel, G. A., Chodosh, L. A., & Weiss, M. J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*, *104*(10), 3136–3147. https://doi.org/10.1182/blood-2004-04-1603

Whitlock, C. A., & Witte, O. N. (1982). Long-term culture of B lymphocytes and their precursors from murine bone marrow. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(11 I), 3608–3612. https://doi.org/10.1073/pnas.79.11.3608

Whitlock, Cheryl A., Tidmarsh, G. F., Muller-Sieburg, C., & Weissman, I. L. (1987). Bone marrow stromal cell lines with lymphopoietic activity express high levels of a pre-B neoplasia-associated molecule. *Cell*, *48*(6), 1009–1021. https://doi.org/10.1016/0092-8674(87)90709-4

Willinger, T., Rongvaux, A., Strowig, T., Manz, M. G., & Flavell, R. A. (2011). Improving human hemato-lymphoid-system mice by cytokine knock-in gene replacement. *Trends in Immunology*, *32*(7), 321–327.

Wilson, N. K., Kent, D. G., Buettner, F., Shehata, M., Macaulay, I. C., Calero-Nieto, F. J., Sánchez Castillo, M., Oedekoven, C. A., Diamanti, E., Schulte, R., Ponting, C. P., Voet, T., Caldas, C., Stingl, J., Green, A. R., Theis, F. J., & Göttgens, B. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell*, *16*(6), 712–724. https://doi.org/10.1016/j.stem.2015.04.004

Wong, J., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W. H., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., Thoeng, A., Khoo, T.-L., Bailey, C. G., Holst, J., & Rasko, J. E. J. (2013). Orchestrated Intron Retention Regulates Normal Granulocyte Differentiation. *Cell*, *154*(3), 583–595. https://doi.org/10.1016/j.cell.2013.06.052

Wong, P., Chung, S. W., Reicheld, S. M., & Chui, D. H. K. (1986). Hemoglobin switching during murine embryonic development: Evidence for two populations of embryonic erythropoietic progenitor cells. *Blood*, *67*(3), 716–721. https://doi.org/10.1182/blood.v67.3.716.bloodjournal673716

Worton, R. G., McCulloch, E. A., & Till, J. E. (1969). Physical separation of hemopoietic stem cells differing in their capacity for self-renewal. *The Journal of Experimental Medicine*, *130*(1), 91–103.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H., & Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*, *154*(5), 1112–1126. https://doi.org/10.1016/j.cell.2013.08.007

Yamane, T., Hosen, N., Yamazaki, H., & Weissman, I. L. (2009). Expression of AA4.1 marks lymphohematopoietic progenitors in early mouse development. *Proceedings of the National Academy of Sciences*, *106*(22), 8953 LP – 8958. https://doi.org/10.1073/pnas.0904090106

Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Månsson, R., Sigvardsson, M., & Jacobsen, S. E. W. (2005). Identification of Lin-Sca1+kit+CD34 +Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood*, *105*(7), 2717–2723. https://doi.org/10.1182/blood-2004-06-2159

Yoder, M. C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D. M., & Orlic, D. (1997). Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity*, *7*(3), 335–344. https://doi.org/10.1016/s1074-7613(00)80355-6

Yoshitake, H., Takeda, Y., Nitto, T., Sendo, F., & Araki, Y. (2003). GPI-80, a β2 integrin associated glycosylphosphatidylinositol-anchored protein, concentrates on pseudopodia without association with β2 integrin during neutrophil migration. *Immunobiology*, *208*(4), 391–399. https://doi.org/https://doi.org/10.1078/0171-2985-00281

Zanjani, E. D., Almeida-Porada, G., Livingston, A. G., Flake, A. W., & Ogawa, M. (1998). Human bone marrow CD34- cells engraft in vivo and undergo multilineage expression that includes giving rise to CD34+ cells. *Experimental Hematology*, *26*(4), 353–360.

http://www.ncbi.nlm.nih.gov/pubmed/9546319

Zheng, S., Papalexi, E., Butler, A., Stephenson, W., & Satija, R. (2018). Molecular transitions in early progenitors during human cord blood hematopoiesis. *Molecular Systems Biology*, *14*(3), e8041. https://doi.org/https://doi.org/10.15252/msb.20178041

Zhou, B. O., Ding, L., & Morrison, S. J. (2015). Hematopoietic stem and progenitor cells regulate the regeneration of their niche by secreting Angiopoietin-1. *ELife*, *4*, e05521. https://doi.org/10.7554/eLife.05521

Ziegler-Heitbrock, L., & Hofer, T. P. J. (2013). Toward a refined definition of monocyte subsets. *Frontiers in Immunology*, *4*, 23.

Zjablovskaja, P., Kardosova, M., Danek, P., Angelisova, P., Benoukraf, T., Wurm, A. A., Kalina, T., Sian, S., Balastik, M., Delwel, R., Brdicka, T., Tenen, D. G., Behre, G., Fiore, F., Malissen, B., Horejsi, V., & Alberich-Jorda, M. (2017). EVI2B is a C/EBPα target gene required for granulocytic differentiation and functionality of hematopoietic progenitors. *Cell Death & Differentiation*, *24*(4), 705–716. https://doi.org/10.1038/cdd.2017.6